

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. Agency Use Only (Leave blank)		2. Report Date 30 Apr 02	3. Report Type and Period Covered (i.e., annual 1 Jun 00 - 31 May 01) Final 1 Dec 00 - 30 April 02	
4. Title and Subtitle Northeastern University Demining MURI Fifth Year Progress and Final Report: An Integrated approach to the detection, localization, and classification of mines			5. Award Number DAAG55-97-1-0013	
6. Author(s) Carey Rappaport, Eric Miller, Charles DiMarzio, Stephen McKnight				
7. Performing Organization Name (Include Name, City, State, Zip Code and Email for Principal Investigator) Northeastern University 302 Stearns Building Boston, MA 02115 E-Mail: Rappaport@neu.edu			8. Performing Organization Report Number	
9. Sponsoring/Monitoring Agency Name and Address U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. Sponsoring/Monitoring Agency Report Number 36275,1-EV-MUR	
11. Supplementary Notes (i.e., report contains color photos, report contains appendix in non-print form, etc.) The views, opinions and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.				
12a. Distribution/Availability Statement (check one) Approved for public release; distribution unlimited			12b. Distribution Code	
13. Abstract (Maximum 200 Words) (abstract should contain no proprietary or confidential information) The best mine detectors, used both in US Army countermine operations and by humanitarian demining efforts that can afford them, are based on fifty-year-old technology comparable to metal detectors used by beachcombers. There is tremendous potential for improving the reliability of mine detection using state-of-the-art sensors, computer modeling, and signal processing, combined with the phenomenal recent advances in computer power. The overall goal of humanitarian demining is to use these tools to help people around the world clear hazardous land at a cost that even developing countries can afford.				
14. Subject Terms (keywords previously assigned to proposal abstract or terms which apply to this award)			15. Number of Pages 2	
			16. Price Code	
17. Security Classification of Report Unclassified	18. Security Classification of this Page Unclassified	19. Security Classification of Abstract Unclassified	20. Limitation of Abstract Unlimited	

Northeastern University Demining MURI

PATENTS

1. DiMarzio, Charles A., Stephen W. McKnight, and Scott C. Lindberg, *Optical--Pulse--Induced Acoustic Mine Detection*, U.S. Patent 6,069,843. Awarded on 30 May 2000.
2. DiMarzio, Charles A., Carey M. Rappaport, Gerhard O. Sauermann, and Taner Oktar, *Multiple Wavelength Microwave Enhanced Infrared Thermography*, submitted March, 2000.
3. Rappaport, Carey L., and Charles A. DiMarzio, *Microwave Enhanced Infrared Thermography*, Provisional Application 60,057,253, 29 Aug 1997. Full application submitted 27 August 1998.

Wavelet Domain Image Restoration Using Edge Preserving Prior Models

Murat Belge and Eric L. Miller
235 Forsyth Building
Northeastern University
Boston, MA 02115

Abstract

In this paper we consider a wavelet-based edge preserving regularization scheme for use in linear image restoration problems. Our efforts build on a collection of mathematical results indicating that wavelets are especially useful for representing functions that contain discontinuities (i.e. edges in two dimensions or jumps in 1D). We interpret the resulting theory in a statistical signal processing framework and obtain a highly flexible framework for adapting the degree of regularization to the local structure of the underlying image. In particular, we are able to adapt quite easily to scale-varying and orientation-varying features in the image while all the time retaining the edge preservation properties of the regularizer. We demonstrate an efficient algorithm for obtaining the reconstructions from observed data and for choosing the multiple regularization parameters governing the priors.

1 Introduction

In this work, we consider the problem of recovering an unknown image \mathbf{x} from measurements \mathbf{y} where the two are related via

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

with \mathbf{H} a known matrix defining the linear degradation, \mathbf{x} and \mathbf{y} vectors representing lexicographically ordered pixels in the original and degraded images and \mathbf{n} representing additive white Gaussian noise (i.e. $\mathbf{n} \sim N(0, \sigma^2 \mathbf{I})$). The above problem is typically ill-posed in the sense that the operator \mathbf{H} may not be invertible or it may be ill-conditioned. In such a case, the original problem is replaced by a well conditioned one whose solution approximates that of the original. This is usually achieved by incorporating some *a priori information* about the original image.

In the stochastic interpretation of image restoration, the prior information on the image \mathbf{x} can be expressed in the form of an *a priori* probability distribution $P(\mathbf{x}|\boldsymbol{\theta})$ which is assumed to be dependent on a set of parameters $\boldsymbol{\theta}$. Bayes rule allows us to combine the prior information with the information contained

in the data to obtain the *a posteriori* distribution

$$P(\mathbf{x}|\mathbf{y}, \sigma, \boldsymbol{\theta}) = \frac{P(\mathbf{y}|\mathbf{x}, \sigma)P(\mathbf{x}|\boldsymbol{\theta})}{P(\mathbf{y})} \\ \propto \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 - \Phi(\mathbf{x}, \boldsymbol{\theta}) \right\} \quad (2)$$

where the prior distribution on \mathbf{x} is assumed to be proportional to $\exp\{-\Phi(\mathbf{x}, \boldsymbol{\theta})\}$. Conventional prior distributions are typically expressed as, $\exp\{-\frac{1}{2\sigma^2} \|\mathbf{L}\mathbf{x}\|_2^2\}$, which is equivalent to the assumption that \mathbf{x} is coming from a Gaussian distribution with covariance matrix $(\mathbf{L}^T \mathbf{L})^{-1}$. The restored image \mathbf{x}^* is then found by maximizing the posterior probability with respect to \mathbf{x} , i.e

$$\mathbf{x}^*(\boldsymbol{\theta}, \sigma) = \arg \min_{\mathbf{x}} \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \Phi(\mathbf{x}, \boldsymbol{\theta}). \quad (3)$$

The quality of the restoration depends on both the form of the prior (i.e. form of $\Phi(\cdot)$ in (3)) and the choice of hyperparameters $(\boldsymbol{\theta}, \sigma)$.

In this paper, we consider a Bayesian image restoration algorithm in which the desired image's subbands are modeled as stationary generalized Gaussian processes. The intuition underlying our modeling is that images typically have spatial structure consisting of smooth areas dispersed with occasional edges. When the image is subjected to a subband decomposition, band pass and high pass pixels of the image are seen to have significantly non-Gaussian behavior with a distribution function sharply peaked about zero (contributions coming from smooth areas) and heavy tails which are mainly due to the edges [5]. Gaussian modeling puts too much penalty on the sharp transitions (edges) in the image therefore the restorations based on Gaussian modeling look smooth. Several authors used generalized Gaussian distributions to model subband statistics in the context of image coding and denoising. However, to our knowledge such models have not been used in connection with image restoration yet. In this paper, we develop new multiresolution edge preserving image models for image restoration applications and solve the nonlinear optimization and parameter selection issues associated with the new modeling.

2 Wavelet Transform of Images

The wavelet transform is a time-frequency representation which has good localization in both domains. We restrict our attention to orthonormal separable wavelet transforms and assume that the reader is familiar with the wavelet theory [1]. The following discussion is intended to establish the notation that will be used in the rest of the paper.

Following the wavelet literature, elements of the image $x(m, n)$, $1 \leq m, n \leq 2^J$, are called the finest scale scaling coefficients. From these finest scale scaling coefficients we obtain the scaling coefficients at a lower resolution level which represent a coarser scale version of x and a set of wavelet coefficients which represents the horizontal, vertical and diagonal details lost in moving from a finer scale to a coarser scale. By repeating this procedure on the coarse scale coefficients $k-1$ times, we can obtain a k -level wavelet decomposition. We denote the scaling coefficients at scale j_0 and position (m, n) by $x_{j_0,0}(m, n)$, $1 \leq m, n \leq 2^{j_0}$ and the wavelet coefficients at scale j , $j_0 \leq j \leq J-1$, orientation i , $1 \leq i \leq 3$ and position (m, n) by $x_{j,i}(m, n)$.

The wavelet transform can be represented as an orthonormal matrix \mathcal{W} , which transforms an image \mathbf{x} into its wavelet domain representation through matrix multiplication. That is, $\hat{\mathbf{x}} = \mathcal{W}\mathbf{x}$. By using this, (1) can be transformed into wavelet domain as

$$\begin{aligned}\mathcal{W}\mathbf{y} &= (\mathcal{W}\mathcal{H}\mathcal{W}^T)\mathcal{W}\mathbf{x} + \mathcal{W}\mathbf{n} \\ \hat{\mathbf{y}} &= \hat{\mathbf{H}}\hat{\mathbf{x}} + \hat{\mathbf{n}},\end{aligned}\quad (4)$$

where $\mathcal{W}^T\mathcal{W} = \mathbf{I}$ follows from the orthonormality of the wavelet transformation.

3 A Multiscale Image Model for Use in Image Restoration

Following the work in [5] on image coding and denoising, we model the distribution of wavelet coefficients of images by a Generalized-Gaussian (GG) density [9]. The following distribution function serves as the prior model

$$P(x_{j,i}(m, n)|p, \theta_{j,i}(m, n)) \propto \exp \left\{ -\frac{1}{p} \left| \frac{x_{j,i}(m, n)}{\theta_{j,i}(m, n)} \right|^p \right\}, \quad (5)$$

where $1 \leq p \leq 2$ is a parameter which determines the tail behavior of the density function and $\theta_{j,i}(m, n)$ is the *scale parameter* similar to the variance of a Gaussian density. For simplicity, we will refer to the density in (5) as $GG(0, \theta_{j,i}(m, n), p)$. For $p = 1$ we have the Laplacian or double exponential density and for $p = 2$ we have the familiar Gaussian density. Intermediate values of p correspond to increasingly heavier tails. We assume that the scaling coefficients $x_{j_0,0}$, $1 \leq m, n \leq 2^{J-L}$, are i.i.d. with $GG(0, \theta_{j_0,0}(m, n), p)$ density.

The specification of one θ parameter for every wavelet coefficient results in an image model far too complex to be of use in a restoration procedure. Generally, there is neither sufficient information *a priori* nor in the data to reliably specify or estimate a

model with as many degrees of freedom as unknowns in the original image. Nonetheless, the structure of the model in (5) coupled with the specification of the problem in the wavelet domain does suggest a variety of simplifications which are of use for the restoration problem. In this work, we consider the following three:

1. **Model 1:** The variance of the wavelet coefficients decreases exponentially with the scale, i.e., $x_{j,i}(m, n) = GG(0, \theta 2^{-\alpha(j-j_0)}, p)$, $i = 1, 2, 3$, $1 \leq m, n \leq 2^j$ with j_0 the coarsest scale, θ the scale parameter corresponding to j_0 and $\alpha \geq 0$. The rationale behind this model is that it is equivalent to a deterministic modeling of the image as a member of Besov spaces [1].
2. **Model 2:** Scaling coefficients are i.i.d. with $GG(0, \theta_{j_0,0}, p)$ distribution and wavelet coefficients at a particular scale are i.i.d. with $GG(0, \theta_j, p)$, $j = J-L, \dots, J-1$. This model is useful in cases where the variance of the wavelet coefficients at different scales can not be well-approximated by a simple exponential law.

For the remainder of this paper, we assume that the two parameters α and p in the above models are known and fixed *a priori*. Generally, the performance of the regularizer is impacted to a greater extent by the on-line identification of the θ parameters and it is here where we choose to concentrate our efforts.

4 A Multiscale Bayesian Image Restoration Algorithm

Motivated by the above discussion, we consider a restoration scheme for obtaining an estimate of the image's wavelet coefficients $\hat{\mathbf{x}}^*$ which is defined in terms of the following optimization problem:

$$\begin{aligned}\hat{\mathbf{x}}^* &= \arg \min_{\hat{\mathbf{x}}} \{ \|\hat{\mathbf{y}} - \hat{\mathbf{H}}\hat{\mathbf{x}}\|_2^2 + \lambda_{j_0,0} \|\hat{\mathbf{x}}_{j_0,0}\|_p^p \\ &\quad + \sum_j \sum_i \lambda_{j,i} \|\hat{\mathbf{x}}_{j,i}\|_p^p \},\end{aligned}\quad (6)$$

where the notation $\hat{\mathbf{x}}_{j,i}$ denotes the vector of wavelet (scaling) coefficients at scale j and orientation i and $\lambda_{j,i} = \frac{2\sigma^2}{p\theta_{j,i}}$. The cost function in (6) has a unique minimum for a given $\boldsymbol{\lambda}$. This follows from the fact that it is just the summation of an L^2 and an L^1 norm which are both convex in $\hat{\mathbf{x}}$. The first order conditions that must be satisfied by $\hat{\mathbf{x}}^*$ is found by differentiating (6) with respect to $\hat{\mathbf{x}}$. However, the cost function in (6) is not differentiable in its current form due to the fact that the l_p norm term present in the cost function is not differentiable at zero. To alleviate this problem, we slightly perturb the l_p norm term to obtain a differentiable approximation: $\|\mathbf{x}\|_p^p \approx \sum_i (x_i^2 + \beta)^{p/2}$, where x_i is the i th element of the vector \mathbf{x} and $\beta > 0$ is a small constant. Substituting this approximation into (6) and differentiating with respect to $\hat{\mathbf{x}}$ we obtain

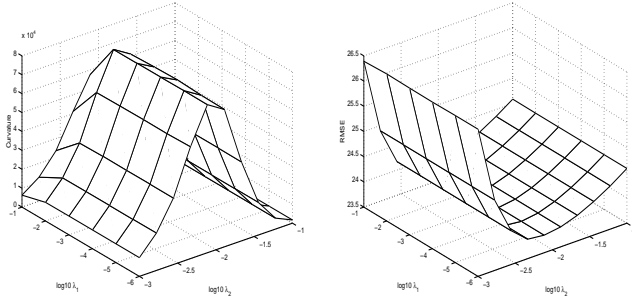


Figure 1: Curvature of the L-hypersurface and the corresponding RMSE plot.

the following equation

$$\mathbf{D} = \text{diag} \left[\frac{\lambda_{j,i}}{(|\hat{x}_{j,i}^*(m,n)|^2 + \beta)^{1-p/2}} \right] \quad (7)$$

$$\left(\hat{\mathbf{H}}^T \hat{\mathbf{H}} + \frac{p}{2} \mathbf{D} \right) \hat{\mathbf{x}}^* = \hat{\mathbf{H}}^T \hat{\mathbf{y}} \quad (8)$$

An iterative algorithm based on (8) can be developed to approximate the solution $\hat{\mathbf{x}}^*$. Starting from an initial estimate $\hat{\mathbf{x}}^{(0)}$, we solve the following equation for $\hat{\mathbf{x}}^{(k+1)}$ until convergence is achieved

$$\left(\hat{\mathbf{H}}^T \hat{\mathbf{H}} + \frac{p}{2} \mathbf{D}^{(k)} \right) \hat{\mathbf{x}}^{(k+1)} = \hat{\mathbf{H}}^T \hat{\mathbf{y}}, \quad (9)$$

where $\mathbf{D}^{(k)}$ denotes the diagonal matrix obtained by replacing \hat{x}^* in (7) by $\hat{x}^{(k)}$. It can be shown that the iterative algorithm in (9) converges to the unique minimum of the perturbed cost function in (6). The stabilization constant β affects the convergence rate. The bigger the β , the faster the convergence. However, using a relatively large β has the side effect of smoothing out the edges in the image. Therefore, β should be set to balance the compromise between the convergence rate of the algorithm and sharpness of the resulting solution. Based on our experience, we recommend $\beta \approx 1$ for problems involving real life images.

4.1 Determination of Regularization Parameters

A noteworthy feature of the multiscale image restoration algorithm introduced in the previous section is the need for multiple regularization parameters. In this paper, we utilize a multidimensional extension of the popular L-curve method [7] called the ‘‘L-hypersurface’’ to determine multiple regularization parameters appearing in (6).

The L-hypersurface is a plot of the log of residual norm, $\|\mathbf{y} - \mathbf{H}\mathbf{x}^*(\boldsymbol{\lambda})\|_2^2$, against the log of side constraint norms $\|\hat{\mathbf{x}}_{j,i}^*(\boldsymbol{\lambda})\|_p^p$, $j_0 \leq j \leq J-1$, for a range of regularization parameters. It has been argued and shown through numerical experiments that the points on the L-hypersurface where the Gaussian curvature reaches

a local maxima are closely tied to the points on the restoration error, $\|\mathbf{x} - \mathbf{x}^*(\boldsymbol{\lambda})\|_2^2$, hypersurface where the restoration error reaches a local minima. The L-hypersurface selection for $\boldsymbol{\lambda}$ corresponds to that $\boldsymbol{\lambda}$ for which the Gaussian curvature of the L-hypersurface is a maximum. Monte Carlo simulations indicates that [8] the L-hypersurface method produces regularization parameters which are almost as good as the optimal (i.e. minimizing the restoration error). The necessary formulas for the computation of Gaussian curvature of the L-hypersurface for the problem in (6) are provided in [8].

5 Simulation Study

In this section, we illustrate the performance of our proposed multiscale image restoration algorithm for both real and synthetic images. In the first example, we used a 2-D Gaussian convolutional kernel, with $\sigma_x = \sigma_y = 2.0$ to blur 256×256 Mandrill image. Zero mean white Gaussian noise was added to set the SNR to 30dB. The top two plots in Figure 2 display the original and the blurred, noisy images.

We restored the degraded Mandrill image using three regularization techniques: our proposed multiscale regularization scheme, the Constrained Least Squares (CLS) algorithm with a 2-D Laplacian regularizer [3], and the TV algorithm [6]. The relevant regularization parameters were determined using either the L-curve or the L-hypersurface method. For the TV algorithm and our algorithm we used $\beta = 1.0$ as the stabilization constant. For our multiscale image restoration algorithm, we used the Daubechies eight tap most symmetrical wavelets [2].

In Figure 2 we display the restored Mandrill images. For our multiscale image restoration method we computed two restorations according to the regularization schemes Model 1 and Model 2 described in section 3 with $p = 1.0$. For Model 1, we used a 5-level wavelet decomposition and set the exponential parameter to $\alpha = 1.2$. For Model 2, we used a 3-level wavelet decomposition and set the regularization for the scaling coefficients to 10^{-5} . For the Model 1 case the L-hypersurface was used to determine two parameters corresponding to the scaling coefficients and the coarsest scale wavelet coefficients. In this case, L-hypersurface is a 2-D function of regularization parameters as seen in Figure 1. Also shown in Fig. 1 is a plot of the root mean square error (RMSE), $\sqrt{\frac{1}{N^2} \|\mathbf{x} - \hat{\mathbf{x}}^*\|_2^2}$, as a function of these regularization parameters. Examining these images shows that the L-surface has a distinct extended maxima along which the RMSE is very close to being a minimum. Thus, we see that the restoration algorithm is not overly sensitive to the scaling coefficient regularization parameter and locating the correct regularization parameter for the wavelet coefficients is more important.

In Model 2 restoration, we employed a 3-level wavelet decomposition and assumed that each scale gets a different regularization parameter. Based on the insensitivity of the restoration to the scaling coefficient regularization parameter we set this value to 10^{-5} .

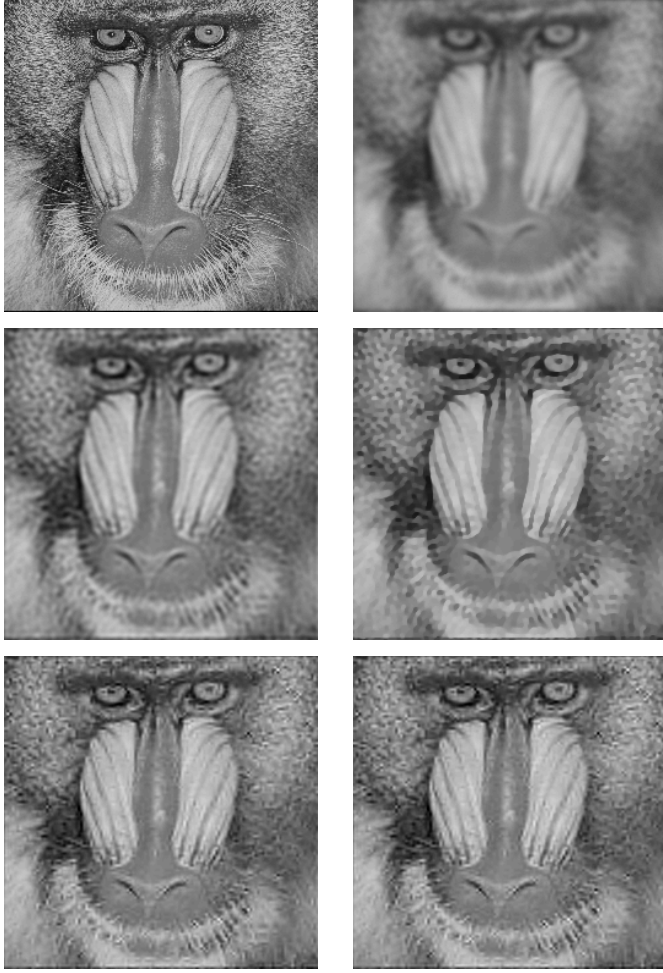


Figure 2: Original, blurred, restored by CLS, restored by TV and restored by Model 1 and Model 2 techniques.

Fig. 2 shows that both the TV algorithm and our algorithm produce restored images visually superior to the CLS algorithm. We also observe that the images restored by our algorithm are a little sharper than that of the TV and the texture-like regions abundant in the Mandrill image (eg. the hairs around the mouth of the Mandrill) are better preserved by our algorithm. The RMSE values are 23.73 for the TV algorithm, 24.11 for the CLS algorithm and 23.75 and 23.84 for the Model 1 and Model 2 restorations respectively. Finally, for this image we see little difference either in terms of the error norm or in terms of visual quality between the the restorations using our method. The primary advantage of the Model 2 approach is that we no longer need to independently set the α parameter.

In our final example, we first blurred the original Bridge image with a 9×9 uniform motion blur and added white Gaussian noise to the degraded image



Figure 3: Original, blurred, restored by the TV, restored by Model 1 scheme.

to set the SNR at 40dB. Figure 4 shows the original and the degraded images. We display the restorations obtained by the TV and the proposed algorithm in Figure 4. For our restoration, we applied the Model 1 regularization scheme with $\alpha = 1.2$, $p = 1.2$ and a 5-level wavelet decomposition. We used the appropriated 2-D L-hypersurface to determine the relevant regularization parameters. Although, in this case the RMSE values were similar (19.43 for TV and 19.72 for the multiscale algorithm), the two restored images in Figure 4 exhibit vastly different visual characteristics. The TV algorithm completely wipes out the small features in the image and produces an overly homogenized restoration resembling an “oil painting” of the original scene. On the other hand, the wavelet-based algorithm is able to reproduce finer detail thereby yielding a more visually appealing restoration.

6 Summary and Conclusions

In this paper, we introduced a wavelet domain multiscale image restoration algorithm for use in linear



Figure 4: Original, blurred, restored by the TV, restored by Model 1 scheme.

image restoration problems. Following the recent results in the area of image denoising and coding, we developed a statistical prior model for the wavelet coefficients of images. Our priors are able to capture spatial, scale space characteristics of images accurately. We developed an efficient iterative optimization algorithm to solve the nonlinear optimization problem resulting from using such priors and introduced the L-hypersurface method to choose the multiple hyperparameters governing the structure of our priors. Comparison of our multiscale image restoration algorithm with either the traditional image restoration algorithms or the more recent edge-preserving image restoration algorithms shows that our algorithm can produce restorations which are visually significantly better than that of the traditional techniques and at least comparable, if not better, than that of the edge-preserving algorithms.

References

[1] Y. Meyer, *Wavelets and Operators*, Cambridge

Univ. Press, 1992.

- [2] I. Daubechies, *Ten Lectures on Wavelets*. NewYork: SIAM Press, 1992.
- [3] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*, Englewood Cliffs, NJ:Prentice-Hall, 1977.
- [4] P. Charbonnier, L. Blanc-Feraud, G. Aubert and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging", *IEEE Trans. Image Processing*, vol. 6, no. 2, pp. 298-311, February 1997.
- [5] E. P. Simoncelli, E. Adelson, "Noise removal via Bayesian wavelet coring," in *Proceedings 3rd International Conference on Image Processing*, Lausanne, Switzerland, September 1996.
- [6] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," unpublished.
- [7] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Review*, vol. 34, pp. 561-580, 1992.
- [8] M. Belge, M. E. Kilmer and E. L. Miller, "Simultaneous multiple regularization parameter selection by means of the L-hypersurface with applications to linear inverse problems posed in the wavelet domain," to be published in *Proceedings of SPIE*, San Diego, U.S.A, July 1998.
- [9] M. Belge, M. E. Kilmer and E. L. Miller, "Wavelet Domain Image Restoration With Adaptive Edge-Preserving Regularization," submitted to *IEEE Trans. Image Processing*.

Statistical Methods for Shape-Based Inverse Scattering

Eric L. Miller¹

235 Forsyth Building

Dept. of Electrical and Computer Engineering

Northeastern University

Boston, MA. 02115

Email:elmiller@cdsp.neu.edu

Misha Kilmer

235 Forsyth Building

Dept. of Electrical and Computer Engineering

Northeastern University

Boston, MA. 02115

Email:mkilmer@ece.neu.edu

Abstract — Estimation and decision-theoretic methods are presented for the localization and characterization of an object positioned in the interior of a region given sparse, noisy observations of scattered radiation obtained at the surface. Using a B-spline model for the perimeter of the object, a sequential generalized likelihood ratio approach is used to deform the contour, add and remove knots, and determine the internal structure of the object. Prior models on the structure of the object are introduced to regularize the overall problem. Examples are provided for a linearized inverse scattering problem.

I. OVERVIEW

The problems of image formation and target identification from limited observations of scattered wavefields arises in areas such as non-destructive testing, medical imaging, and land-mine remediation. Often the physical quantity of interest (electrical conductivity, acoustic velocity etc.) is pixelated and a gradient decent approach based on the scattering physics is used to find an image of the parameter [1]. The resulting image is then post-processed to extract the desired targets. To bypass the computational burden and ill-posedness of this imaging approach, there has also been much work on directly estimating target shape and location information [2]. The underlying assumption is that the target represents a homogeneous perturbation on a homogeneous background.

Here, we discuss an alternate method designed to solve simultaneously the image formation and object detection problems. The unknown is modeled as a superposition of a slowly varying, compactly supported perturbation on a background of partially known structure. Mathematically, we have

$$g(\mathbf{r}) = S(\mathbf{r})\mathbf{B}_1(\mathbf{r})\mathbf{a}_1 + (1 - S(\mathbf{r}))\mathbf{B}_2(\mathbf{r})\mathbf{a}_2 \quad (1)$$

where g is the object of interest, $S(\mathbf{r})$ is 1 over the (unknown) support of the object and 0 elsewhere, \mathbf{B}_1 (\mathbf{B}_2) are row vectors holding expansion functions describing the texture of the object (background), and \mathbf{a}_1 (\mathbf{a}_2) are the associated expansion coefficients. Assuming the \mathbf{B}_i are

known, the objective of the problem is to determine the structure of S and the \mathbf{a}_i given a data vector y related to g via $y = Ag + n$ where A is the forward scattering operator (here taken to be linear) and n is a vector of additive white Gaussian noise.

We use an iterative algorithm to solve the target id and imaging problems. First, given an estimated perimeter curve at step k (denoted \hat{S}^k), we employ estimation theoretic methods to determine a low-order, optimal image of the perturbation and background. That is, we take as estimates of the \mathbf{a}_i at step k ,

$$(\hat{\mathbf{a}}_1^k, \hat{\mathbf{a}}_2^k) = \arg \min_{\mathbf{a}_1, \mathbf{a}_2} \|y - Ag(\mathbf{a}_1, \mathbf{a}_2, \hat{S}^k)\|_2^2 + \Omega_a(\mathbf{a}_1, \mathbf{a}_2)$$

where $\Omega_a(\mathbf{a}_1, \mathbf{a}_2)$ is derived from a prior statistical model on the behavior of the perturbation and the background. Second, given an estimated image, we use an M-Ary hypothesis testing approach to determine how best to modify the structure of the perimeter curve. The m th hypothesis, H_m^k , is of the form $S_m^k = \hat{S}^k + \delta S_m$ where δS_m is generated by perturbing the current estimate of the object support either by moving a control point in the representation of the boundary or by adding a new knot to this representation. The new curve then is defined by the following

$$\hat{S}^{k+1} = \arg \max_m \left[\log \text{Prob}(y | H_m^k, \hat{\mathbf{a}}_1^k, \hat{\mathbf{a}}_2^k) + \Omega_S(S_m^k) \right]$$

where Ω_S is obtained from a prior model on the structure of S which is constructed to reflect our expectations concerning the geometric properties of the “blobs” likely to be found in the application of interest.

In the talk accompanying this abstract we describe in greater detail the prior models associated with Ω_a and Ω_S , the methods for determining $\hat{\mathbf{a}}_i$, and the manner in which the H_m^k are generated. Finally, we present examples of this algorithm applied to a medical imaging problem in which diffused photon wavefields are used to probe the sub-dermal structure of the human body.

REFERENCES

- [1] R. E. Kleinman and P. M. van den Berg, “An extended range modified gradient technique for profile inversion,” *Radiation Science*, vol. 29, 1994, pp. 877-884.
- [2] A. Litman, D. Lesselier, and F. Santosa, “Reconstruction of a two-dimensional binary obstacle by controlled evolution of a level-set,” *Inverse Problems*, vol. 14, pp. 685-706.

¹This work was supported by an ODDR&E MURI under Air Force Office of Scientific Research contract F49620-96-1-0028, a CAREER Award from the National Science Foundation MIP-9623721, and the Army Research Office Demining MURI under Grant DAAG55-97-1-0013.

A STATISTICAL APPROACH TO MULTICHANNEL BLIND SIGNAL DETECTION FOR GROUND PENETRATING RADAR ARRAYS

Xiaoyin Xu, Eric L. Miller, and Gary Sower

235 Forsyth Building
Northeastern University
Boston, MA, 02115
elmiller@ece.neu.edu

ABSTRACT

We address the problem of detecting buried mines or objects from multichannel sequentially collected ground penetrating radar (GPR) data for applications such as landmine and unexploded ordnance remediation, utility line mapping, and archaeology. Usually, the exact form of target signal is not known. We develop and analyze a low-complexity, decision-theoretic, sequential detection strategy to solve this problem. At each stop of the array, a windowed ANOVA test is designed to look for significant difference among the multichannel signals and suppress accumulated noise. A sequential estimator-detector then processes the ANOVA results to detect mines as array proceeds down track. Its performance in terms of mean time between false alarms and probability of detection is analyzed. Results from applying the method on real-world data are presented.

1. INTRODUCTION

The use of ground penetrating radar (GPR) arrays for detecting buried objects has received considerable attention in recent years in areas such as landmine and unexploded ordnance remediation, utility line mapping, and archaeology [9]. A typical GPR configuration for such applications is shown in Fig. 1. Here one array of transmitters and a second array of receiving elements are simultaneously moved down a linear track. At every stop of the system, each transmitter emits a short pulse of electromagnetic energy which interacts with the surrounding medium. Based on observations of scattered fields collected by the receivers the objective of the problem is to determine if an object is present in the field of view of the array.

Current processing methods for this problem fall into one of three categories. First, pattern matching methods [5]

employ techniques such as fuzzy set theory and neural networks. Such methods can be fast but also require extensive training to function well. Moreover, performance analysis is limited to Monte-Carlo simulations. Second, image-then-detect techniques [4] employ a beamforming or backpropagation approach to build an image of the subsurface which is then post-processed to detect objects. Such methods generally require the data from the full GPR scan to form an image and are thus not well suited to on-line computations in which information is processed sequentially as the array proceeds down track. More subtly, the attenuation associated with the propagation of the GPR signal typically results in useful signal only over receivers located closest to the firing transmitter. In many functioning systems in fact only a single receiver is employed per transmitter. Thus, methods based on beamforming which require array-based observations are not really appropriate for this problem. Finally, there has been some very interesting work done in the area of statistical processing methods [3] where one can examine quantities such as detection rates, false alarm probabilities, etc.; however the techniques in [3] for instance are based on highly complex electromagnetic models for the GPR sensor and are thus computationally intensive.

With this as background, here we view the GPR detection problem in a multichannel (one channel per transmit/receiver pair) blind signal detection framework and employ statistical methods to process the GPR returns. This approach allows us to exploit two generic properties of the signal transmission process. First, for any given stop of the array, the presence of the mine close to a transmitter-receiver (T/R) pair results in a jump in the mean value of the observed signal relative to that seen in other pairs. Thus, we develop a windowed ANOVA test [7] to detect this change. Second, physical principles dictate that as the array moves from one stop to the next, this jump will be detectable only for those positions of the array close to the object; that is, the jump will be transient [9]. This behavior is exploited in the synthesis of a sequential detector to process the output of the ANOVA test as the GPR system moves down the track.

This work was supported by an OSD MURI on Demining under Grant DAAG55-97-1-0013

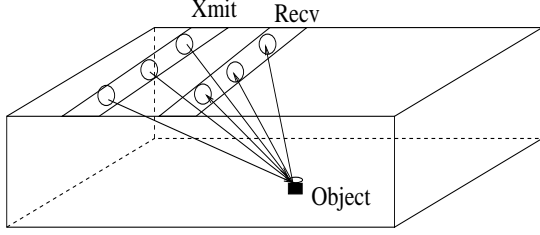


Figure 1: Setup of GPR array.

The paper is organized as follows. Section 2 discusses physical model of a GPR system and corresponding mathematic model. In Section 3, we present a statistical algorithm to detect mines and analyze its performance in mean time between false alarms between false-alarms and probability of detection. Conclusion and direction of future work is given in Section 4.

2. MODEL

2.1. GPR Model

To start, we consider a simple model of the GPR returns [2]. In addition to sensor noise, there are two primary components in the GPR signal. One always sees a signal due to the specular reflection of the transmitted signal from the ground. While there are many ways of modeling this ground bounce signal [3], here we assume that it varies slowly as the array moves down the track so that it can be estimated and subtracted from the data.

If a subsurface scatterer is present, the receiver records signal from that source too. In Fig. 2 we plot the ground-bounce corrected data for single T/R pair obtained from field data as a function of down-track position of the sensor. That is, each column of this image is composed of the samples in the received waveform for a given stop of the array. As can be seen from this figure the target signal is transient in two ways. First, it appears only when the sensor system is in the vicinity (≈ 0.5 meter) of the mine. For most systems this implies that there are on the order of tens of stops of the array when the mine can be detected out of hundreds to thousands of array positions in a typical GPR survey. Second, when the system is positioned close to the object, the effects are only seen in a small number of samples of the received waveform.

Before proceeding, we make two observations regarding the signals in Fig. 2. First in general the precise form of the signal of interest is typically unknown due to variations caused by fluctuations in the electrical properties of the soil, unmodeled physical effects such as surface roughness, and variability in the signature caused by the unknown orientation of the target relative to the sensor. Indeed, Fig. 2 is typical of the variability seen in practice. Second for clar-

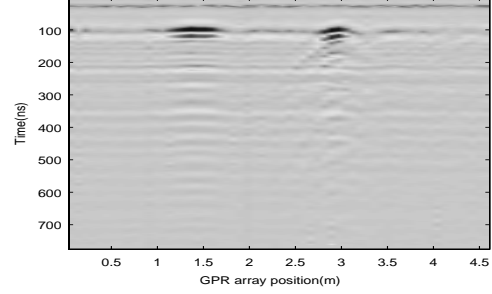


Figure 2: Signal generated from two objects buried at position 1.5m and 3.0m.

ity, we have displayed signals from two targets which are relatively easy to detect. Generally, even after removing the ground bounce other sources of noise and clutter can significantly reduce the detectability of the desired signal and increase number of false-alarms.

2.2. Mathematic Model

Based on this discussion, at down track position n of the array, we have the classic binary hypothesis model for the GPR signal returns

$$\begin{aligned} H_0 : \quad & \mathbf{x}_i(n) = \mathbf{w}_i(n) \\ H_1 : \quad & \mathbf{x}_i(n) = \mathbf{s}_i(n) + \mathbf{w}_i(n) \end{aligned} \quad (1)$$

for $i = 1, \dots, M$ where $\mathbf{x}_i(n)$ is the length K vector of time samples for the received signal from the i th T/R pair, $\mathbf{s}_i(n)$ represents the signal due to the target, and $\mathbf{w}_i(n)$ is measurement noise, assumed to be white and Gaussian with zero-mean and variance σ_i^2 . While the exact structure of \mathbf{s}_i is unknown, we see that the effects of this signal are to cause a jump in the mean of the \mathbf{x}_i for those n where the target is in the field of view of the GPR. Thus the statistical problem of interest in this work is to detect this transient signal based on data vectors from M sensors sequentially obtained as the system moves down track.

As detailed in the remainder of this paper, we approach this problem in two steps. First, we develop a method for detecting the presence of \mathbf{s}_i for a fixed location of the GPR. Second, we use this test in a sequential manner to process the returns as the system acquires new data.

At a given position of the array since we do not know of \mathbf{s}_i the statistical problem we pose is

$$\begin{aligned} H_0 : \quad & \boldsymbol{\mu}_i = \mathbf{0} \\ H_1 : \quad & \boldsymbol{\mu}_i \neq \mathbf{0}, \quad i = 1, \dots, M \end{aligned}$$

with $\boldsymbol{\mu}_i = E[\mathbf{x}_i]$. In many cases, analysis of variance (ANOVA) is used to solve this problem. Recently however, Fan [6], Fan and Lin [7], has noted that the performance of ANOVA suffers for problems when the signal of interest is

limited to a small number of samples in the overall observation vector as is the case for the GPR problem of interest here. Fan's original work was limited to problems in which the first m samples were used, with m found from the data. Here we consider a generalization of Fan's work to take into account the fact that for the GPR problem the transient object is significant over a window not generally starting with the first sample. Following the windowed ANOVA, we turn to a sequential probability ratio test (SPRT) to make on-line detection as new data is acquired.

3. ALGORITHM

3.1. Estimation and detection

Before describing windowed ANOVA, it is helpful to see why in some cases ANOVA loses its discrimination power. For example, assume we have one observation vector of size $K \times 1$, $\mathbf{x}_1 \sim N(\boldsymbol{\mu}, \sigma_1^2 \mathbf{I})$ and we wish to test $H_0 : \boldsymbol{\mu} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\mu} = \boldsymbol{\mu}_1$. Standard ANOVA is to estimate $\boldsymbol{\mu}$ by \mathbf{x}_1 and use the testing procedure $X_1^2 = \|\mathbf{x}_1\|^2$. The approximate power of the standard ANOVA estimator-correlator test is

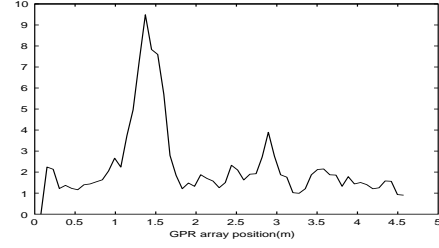
$$P_d(H_1|H_1) = Q\left(\gamma - \frac{\sum_{k=1}^K \mu_1^2(k)}{\sigma_1^2 \sqrt{2K}}\right) \quad (2)$$

where γ the test threshold and Q the complementary cumulative distribution function and strictly decreasing. If $\boldsymbol{\mu}_1$ is different from $\mathbf{0}$ for only a small number of k then as K goes large, $\sum_{k=1}^K \mu_1^2(k) \ll \sqrt{K}$. From (2), we then conclude that testing all dimensions of the data actually causes the test to lose power due to the accumulation of stochastic noise. This deterioration is reflected in the factor $1/\sqrt{2K}$.

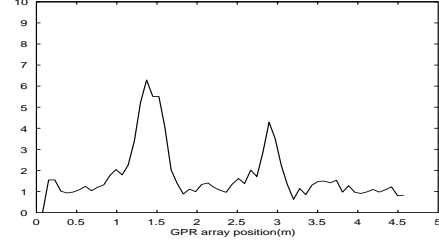
Based on this observation, Fan *et al.* developed an adaptive windowed ANOVA test of the form described previously. Fig. 3 shows results from an ANOVA and a windowed ANOVA test for the signals in Fig. 2. The windowed ANOVA truncates observation vectors of length K to length L by two steps 1) discarding the first m_1 components which are consisted of ground-bounce only (the target reflected signal always comes after the ground-bounce) and 2) discarding the last m_2 elements which represent signal that has been attenuated too much to be meaningful for signal processing. Both the ANOVA and windowed ANOVA tests easily detect the stronger signal, at position 1.5m. For the weak signal at about position 3.0m, windowed ANOVA shows improved performance. Specifically, the peak in the windowed case is more clearly visible above the nominal "noise floor."

Generalizing windowed ANOVA to multiple observations, we build the test statistic as

$$X^2 = \sum_{i=1}^M \sigma_i^{-2} \|\mathbf{x}_i - \bar{\mathbf{x}}\|^2 \quad (3)$$



(a)



(b)

Figure 3: Test statistic, a) result of an ANOVA, using all dimensions of observations, b) result of a windowed ANOVA, using truncated observations.

with $\bar{\mathbf{x}} = \sum_{i=1}^M \sigma_i^{-2} \mathbf{x}_i / \sum_{i=1}^M \sigma_i^{-2}$. It is easily shown that

$$X^2 \sim \chi_{ML}^2(\delta^2) \quad (4)$$

where

$$\delta^2 = \sum_{i=1}^M \sigma_i^{-2} \|\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}\|^2 \quad (5)$$

with $\bar{\boldsymbol{\mu}} = \sum_{i=1}^M \sigma_i^{-2} \boldsymbol{\mu}_i / \sum_{i=1}^M \sigma_i^{-2}$ and $\chi_{ML}^2(\delta^2)$ is the χ^2 distribution with ML degrees of freedom and non-centrality parameter δ^2 .

While windowed ANOVA detects difference among observations at one stop of the array, it does not capture the structure seen as the array moves down track. To improve detection performance, we employ a sequential detection scheme based on the processing of the windowed ANOVA statistic to look for the transient signal from one stop of the GPR array to the next. Hence, the two hypotheses are

$$\begin{aligned} H_0 : & \quad X^2(n) \sim \chi_{ML}^2(0) \\ H_1 : & \quad X^2(n) \sim \chi_{ML}^2(\delta^2(n)) \end{aligned} \quad (6)$$

for $n = 1, \dots, N$ where $\delta^2(n)$ is defined in (5). At stop n , the log likelihood ratio for this problem is

$$u(n) = \ln \frac{p_n(X^2(n))}{p_0(X^2(n))} \quad (7)$$

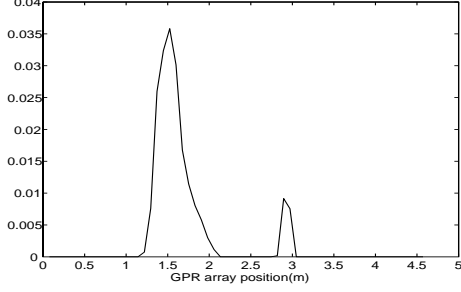


Figure 4: Sequential detection test statistic $U(n)$.

with $p_n(X^2(n))$ the $\chi_{ML}^2(\delta^2(n))$ PDF evaluated at $X^2(n)$ and $p_0(X^2(n))$ the $\chi_{ML}^2(0)$ PDF evaluated at $X^2(n)$. One difficulty with generating $u(n)$ is that $\delta^2(n)$ is typically not known *a priori* since the underlying μ_i are not assumed known. Here we construct a generalized log likelihood statistic. Typically, this is done by replacing $\delta^2(n)$ by its maximum likelihood estimate which for this problem is $X^2(n)$. After some experimentation in real data and computer simulation, we have found that performance can be improved by incorporating a one-step delay into the processing. Specifically, we estimate $\delta^2(n)$ as $X^2(n-1)$.

The full algorithm then is essentially a repeated sequential probability ratio test (SPRT) [1] and is summarized by the following steps,

- $U_1 = 0$.
- FOR $n = 2, \dots, N$
 - $\hat{\delta}^2(n) = X^2(n-1)$
 - Form $u(n)$ according to (7)
 - $U(n) = \max(0, U(n-1) + u(n))$.
 - IF $U(n) > \alpha$, declare object.
- ENDFOR

where α is a preset threshold. Fig. 4 shows the result of the algorithm applied on windowed ANOVA output of Fig. 3. Clearly, setting α below about 0.01 will allow both objects to be detected.

3.2. Performance

Performance of this algorithm is studied in terms of mean time between false alarms and probability of detection. Under H_0 , the test statistic changes as a Markov chain, with update $u(n)$ at each step. Under H_1 , the update $u(n)$ is changing, and therefore, the test statistics $U(n)$ can be described as an inhomogeneous Markov chain. One method of analyzing Markov chain is a matrix approach, which we will take in this paper. First, since ML is usually very large,

χ^2 distribution can be approximated by a Gaussian distribution. We then have

$$\begin{aligned} H_0 : X^2(n) &\sim N(ML, 2ML) \\ H_1 : X^2(n) &\sim N(ML + \delta^2(n), 2ML + 4\delta^2(n)). \end{aligned}$$

Second, suppose the update $u(n)$ and the interval $[-\alpha, \alpha]$ is uniformly quantized to $2k+1$ levels such that $v_0 = 0$, and $v_k = \alpha$. Let $\pi_{nl} = \Pr\{U_n = v_l | N > n\}$ for integer $l \in [0, k-1]$, and $\pi_n = [\pi_{n0}, \pi_{n1}, \dots, \pi_{n(k-1)}]^T$ denote the probabilities of $U_n = v_k$ on the condition that there is no alarm until after sample time n . We can write the transition matrix $C^{(n)}$ at step n as [8]

$$C^{(n)} = \begin{bmatrix} p_n^0 & p_n^{-1} & \dots & p_n^{-\gamma+1} \\ p_n(1) & p_n(0) & \dots & p_n(-\gamma+2) \\ p_n(2) & p_n(1) & \dots & p_n(-\gamma+3) \\ \vdots & \vdots & \ddots & \vdots \\ p_n(\gamma-1) & p_n(\gamma-2) & \dots & p_n(0) \end{bmatrix} \quad (8)$$

where

$$p_n(l) = \Pr(\log \frac{f_1^{(n-1)}(X^2(n))}{f_0(X^2(n))} = v_l) \quad (9)$$

and elements in first row of Eq. 8 are $p_n^i = \sum_{l=-\infty}^i p_n(l)$. Here $f_1^{(n-1)}(X^2(n))$ denotes the PDF of $X^2(n)$ under the alternative hypothesis H_1 and using $X^2(n-1)$ to estimate $\delta^2(n)$. Under H_1 , the transition matrix $C^{(n)}$ can be built using Eq. 8 for different n , since signal is time varying. The probability of detecting a signal of length n_d is then found from

$$P_d(n_d) = 1 - \mathbf{1}^T \prod_{n=1}^{n_d} C^{(n)} \pi_0 \quad (10)$$

where $\mathbf{1}$ is a column vector of all ones. And the mean time between false alarms \bar{T} is found to be [8]

$$\bar{T} = \mathbf{1}^T (\mathbf{I} - C^{(0)})^{-1} \mathbf{e}_1 \quad (11)$$

in which $C^{(0)}$ is the transition matrix under the signal-absent statistics and \mathbf{e}_1 is a column vector containing a one in position 1 and zeros elsewhere. With more levels of quantization, we obtain more accurate estimation of P_d and \bar{T} . Fig. 5 shows simulated results of the mean time between false alarms between false alarms under the null hypothesis H_0 . With a changing measurement noise level, probability of detection changes, as shown in Fig. 6.

Fig. 7 shows the probability of detection when the length of signal is changing. As expected, a signal of large transient length can be detected at a higher probability.

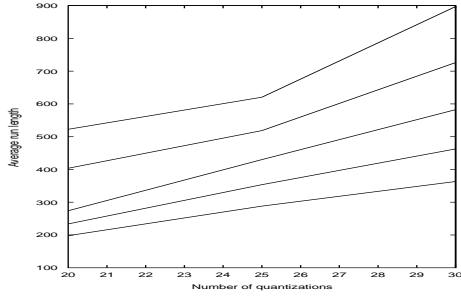


Figure 5: Mean time between false alarms with different threshold and number of quantization levels.

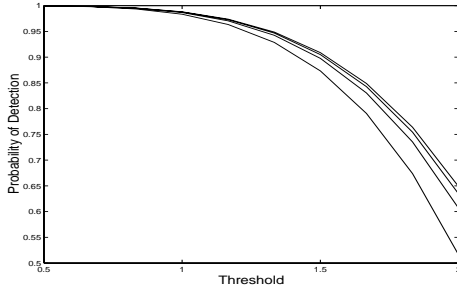


Figure 6: Probability of detection at different signal to noise ratio.

4. CONCLUDING REMARKS

In this paper, we have proposed a sequential, windowed ANOVA to process multichannel GPR returns. The method has a relatively low computational complexity and can be implemented in real-time. We have demonstrated the performance of this technique on a sample of field data. Future research will focus on adaptive sampling of received signal to take into consideration of roughness of ground-air interface; on-line localization will be integrated to allow detect-localize-detect and thus provides higher probability of detection.

5. REFERENCES

- [1] M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, New York, 1993.
- [2] J. M. Bourgeois and G. S. Smith. A fully three-dimensional simulation of a ground-penetrating radar: FDTD theory compared with experiment. *IEEE Trans. Geosci. Remote Sensing*, 34(1):36–44, January 1996.
- [3] T. Dogaru and L. Carin. Time-domain sensing of targets buried under a rough air-ground interface. *IEEE Trans. Antennas and Propagation*, 46(3):360–372, March 1998.

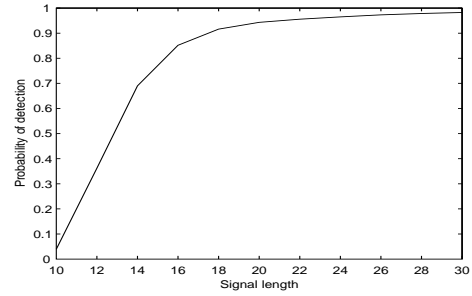


Figure 7: Probability of detection versus transient length of the signal.

- [4] Marshall Bradley et al. Mine detection with a multi-channel stepped-frequency ground-penetrating radar. In *Detection and Remediation of Mines and Minelike Targets IV*, volume SPIE 3710, pages 953–960, 1999.
- [5] Paul Gader et al. New results in fuzzy set based detection of landmines with gpr. In *Detection and Remediation of Mines and Minelike Targets IV*, volume SPIE 3710, pages 1075–1084, 1999.
- [6] J. Fan. Test of significance based on wavelet thresholding and Neyman’s truncation. *Journal of American Statistical Association*, 91:674–688, 1996.
- [7] J. Fan and S. Lin. Test of significance when data are curves. *Journal of American Statistical Association*, 93:1007–1021, 1998.
- [8] C. Han, P. K. Willett, and D. A. Abraham. Some methods to evaluate the performance of Page’s test as used to detect transient signals. *IEEE Trans. Signal Processing*, 47(8):2112–2127, August 1999.
- [9] L. Peters Jr., J. J. Daniels, and J. D. Young. Ground penetrating radar as a subsurface environmental sensing tool. *Proceedings of the IEEE*, 82(12):1802–1822, December 1994.

SPECIFYING PML CONDUCTIVITIES BY CONSIDERING NUMERICAL REFLECTION DEPENDENCIES

Scott C. Winton

The Mitre Corporation

202 Burlington Road

Bedford, MA 01730

Carey M. Rappaport

Center for Electromagnetics Research

Northeastern University

Boston, MA 02115

Abstract Berenger's Perfectly Matched Layer (PML) Absorbing Boundary Condition (ABC) has greatly enhanced FDTD scattering analysis. In a discretized domain, however, performance is signal-dependent and large angle performance is poor due to a rapid reduction in layer decay rate. Increasing the conductivity to offset this reduction increases the discretization errors, especially at near-normal incidence angles. However, by carefully specifying the conductivity in each of the PML sub-layers, it is possible to balance the small and large angle performance. The signal-dependence of reflections may be described in terms of the number of spatial points per wavelength. This lends itself to an overall strategy for which to search for PML profiles that provide superior performance for waves incident on a PML at angles between 0° and 75° and signals that have at least 15 spatial points per wavelength sampling. A 1-dimensional projection method may be employed to allow an exhaustive search to become a viable alternative to optimization. Such a search provides profile pa-

rameters, that while not necessarily 'optimal', give excellent wide angle, wide band reflection performance.

I. INTRODUCTION

The Finite Difference Time Domain (FDTD) method for electromagnetic scattering problems is particularly useful because it is wideband, easily implemented and adaptable to a wide variety of problems. To keep FDTD simulations from becoming computationally expensive, edges of the computational domain must be kept as close to the scatterer as possible. To avoid reflections from the edges of the domain from scattered fields propagating in all directions, an Absorbing Boundary Condition (ABC) must absorb waves incident from all angles. Historically, ABCs have had only limited success in absorbing waves incident at all angles[2-5], but this recently changed with the introduction of the Berenger Perfectly Matched Layer (PML)[6].

The PML works perfectly in continuous analysis, absorbing waves incident at arbitrary angles with arbitrary attenuation. Unfortunately, in order to be used in computer simulations, the PML must be discretized. This discretization limits the conductivity increase for offsetting the reduction in decay rate as the angle of incidence increases and thus limits the overall performance of the PML. Although many efforts improve the PML[7-15], few of these deal with performance as a function of angle, choosing instead to concentrate on improving performance at normal incidence. Although large angle effects may be minor if waves incident on a PML at grazing angles reflect nearly normally incident on an adjacent PML, it has been shown [7] that there are several applications where this is not the case and that the performance at large incidence angles is important to the overall performance of the ABC. Consequently, the improvement of large angle performance of the PML is an important problem.

Discretization also prevents the PML from being signal-independent, as it is in the continuous case. Most investigators choose to describe this signal-dependence as a function

of frequency. Because FDTD involves spatial and temporal sampling, we suggest that it may be simpler to describe the signal-dependence in terms of spatial points per wavelength.

There has also been much effort to analyze numerical reflections in the PML in order to optimize[16-20] wave absorption. Obtaining a closed form expression for the reflection from a PML as a function of the PML parameters, however, is a difficult problem. This difficulty lies in the fact that numerical reflections are created each time a wave passes from one layer to another. These reflections, in turn, create more reflections as they encounter different layers. Keeping track of all reflections becomes increasingly difficult as the number of layers is increased. Furthermore, this type of analysis can only be performed for one frequency at a time, complicating any effort for optimization.

The approach being used for this current work is largely heuristic. We wish to understand the mechanisms that affect reflections from a PML in order to develop a strategy that will allow us to identify PML parameters that will provide enhanced performance over the widest range of angles and signals. With the aid of computational analysis tools, an automated search has been developed for PML parameters . Although the 'optimal' profile may depend on the specific application, the parameters determined in this analysis provide the best wideband, wide-angle performance reported in the standard literature.

II. REFLECTION DEPENDENCIES

The decay rate of the PML in continuous time and space is given as $\alpha = \sigma \eta_0 \cos \theta$ [9], where θ is the incidence angle, σ is the conductivity of the PML and η_0 is the impedance of free space. Note that the decay rate is proportional to $\cos \theta$. Since there is no reflection at the PML interface, σ may be chosen arbitrarily large in order to offset the loss of decay rate due to the $\cos \theta$ factor.

The geometry for the discrete PML is shown in Figure 1. The decay rate in the discretized PML also decreases with increasing angle. Unlike with the continuous case, the discretized σ_i in the i^{th} PML sub-layer may not be chosen arbitrarily large. In general,

since FDTD simulations calculate electric and magnetic fields on complementary overlapping lattices, they are sensitive to changes in the parameters of adjacent layers. Stated another way, at any air-PML or PML-PML interface, the discrete equations for the fields at the interface are dependent on unmatched conductivities.

As an example, assume that σ_i are constant over a PML half space. An incident wave encounters only one interface, that between free space and the PML. The difference equation for the field at the interface will be dependent on the conductivity of the free space ($\sigma=0.0$) and the conductivity of the PML ($\sigma=\text{constant}$). If σ is very small, the resulting reflection will also be very small. If σ is increased, the decay rate increases, but there will also be a larger reflection from the interface. If a gaussian pulse, $\text{Exp}[-c^2(t - t_0)^2/W^2]$, where t_0 is the time at which the pulse is at its maximum, excites the first row of an FDTD grid, Figure 2 shows the total reflection from a constant conductivity PML half space versus angle for values of σ equal 0.15, 0.3, and 0.45 S/m. For these simulations, the temporal increment $\Delta t = 10\text{pS}$, spatial increment $\Delta = 0.012\text{m}$ and $W = \frac{50}{3}\Delta = 0.2\text{m}$. It is apparent that the reflection increases with an increase in σ , and that as θ increases the discrete change in conductivity at the interface is moderated and the reflection decreases.

It should be noted that the time and space steps need not have physical units. Instead, the entire FDTD simulation including the PML may be entirely specified by the Courant number $R = c\Delta t/\Delta$ and the i^{th} layer decay rate $S_i = \sigma_i\Delta\eta_0$ and the performance of the PML may be parameterized in terms of two unitless parameters: $S = (S_1, S_2, \dots, S - N)$, λ/Δ and the angle of incidence, θ . Given these, Δ and the wave phase velocity all other physical quantities can be determined.

One way reflection from the PML is reduced is to vary the sub-layer conductivities from small values near the free space interface to larger values toward the final sub-layer. This profiling of PML conductivity improves absorption, providing an alternative to increasing the number of sub-layers, which is computationally expensive. Since incoming waves do not encounter larger values of σ until well within the layer, the reflections caused by these

larger jumps in conductivity are attenuated by the initial lossy layers both for forward and backward propagation. Yet even with the parabolic conductivity profile $\sigma_i = \sigma_f (i/N)^2$ [6], the total reflection from an 8-PML at 75° incidence is shown to be about -33dB , compared to the -100dB achieved at normal incidence. It is noted that as the angle of incidence is increased, the numerical experimental results more closely match the theoretical loss, which for the case of [6] at 75° , was -31dB . This can be explained by the fact that as the angle of incidence is increased, the projected spatial increment $\Delta \cos \theta$ decreases, which has the effect of decreasing the discretization error, and the jumps in conductivity variations at every layer interface are smaller, giving lower sub-layer reflections.

Since $\sigma^*/\mu_0 = \sigma/\epsilon$, where σ^* is the magnetic conductivity, σ/ϵ may be used in both Ampere's and Faraday's laws [13]. Discrete conductivity values σ_n can be assigned to every half sub-layer. Using the half-layer profile specification and assuming no discretization error, the two-way loss equation is:

$$L = e^{-2 \sum_{n=1}^{2N} \sigma_n \eta_0 \cos \theta \frac{\Delta}{2}} \quad (1)$$

with conductivity profile taking the form $\sigma_n = \sigma_f (n/2N)^p$, $n = 1, 2, \dots, 2N$. For this profile form, the decay rate S may be described by $S_f = \sigma_f \Delta$ and p . A conductivity profile with a power dependence is simple to implement and does not change as rapidly as a profile with an exponential dependence, which was found to not perform as well.

The three different time-domain signals shown in Figure 3 have different temporal increments (Sig.1 $\Delta t = 2ps$, Sig.2 $\Delta t = 1.8ps$, Sig.3 $\Delta t = 1.7ps$), as well as the different waveforms which account for the different frequency contents. These signals have been used as the excitation for a 1-dimensional FDTD simulation to test various PMLs with perfectly conducting terminations. In each case, the Fourier transform of the reflected field divided by the Fourier transform of the incident field has been plotted versus the inverse of the discrete frequency, $2\pi/\omega\Delta t$, times the Courant number, R , which is also the number of spatial points per wavelength, λ/Δ . The use of non-physical units to describe

PML reflection has also been used in [22]. The incident waveform was obtained by using a much larger FDTD simulation where the wave passes over the receiver location just outside the PML and the simulation is stopped before reflection can occur. Results for the 1-dimensional FDTD simulations at normal incidence in Figure 4 show all three signals practically overlapping for all three PML layer configurations of $N = 8, 10, 12$. The layer size and normalized conductivity profile is indicated for each set of graphs. Note that since the curves coincide for each PML configuration, PML reflections are independent of R and of the waveform, but dependent primarily on the relationship between the two (i.e λ/Δ).

To further investigate the PML reflection dependencies, we consider the same three signals, but in this case maintain the temporal increment and conductivity profile, but use three different spatial increments and hence different Courant numbers and different values of the decay rate S . Here we test the 8-layer PML only, the results are in Figure 5. Note that the magnitudes of the reflection coefficients for the different signals are very different.

These figures clearly indicate that it is not the physical quantities ω and σ that determine the reflection from the PML, but rather the non-physical quantities S and λ/Δ . This intuitive result follows since the PML works 'perfectly' in the continuous domain specified by physical parameters, while reflections that arise from the discretized PML depends on discretization parameters.

III. EXHAUSTIVE SEARCH VS. OPTIMIZATION

One obvious method for improving PML performance is to optimize the PML parameters S_f and p over all angles and signals. To do this would require an analytic expression for reflections from the PML as a function of the size of the PML, incident angle, incident signal, S_f and p . Such analytic expressions which have been numerically verified for small N PMLs appear in the literature [18,22]. For an 8-layer PML, analytic expressions in the literature have not been found to accurately predict numerical reflections for all angles and

incident waveforms.

Even assuming confidence with an analytic expression of the reflection coefficient, optimization of the PML over all profiles, signals and angles is a non-trivial problem. By relaxing the 'optimality' requirement, we may utilize an exhaustive search to identify profiles that will provide superior performance over a wide range of signals and angles.

IV. EXHAUSTIVE SEARCH

The basic idea behind the exhaustive search is simple: FDTD simulations are run for each PML profile under consideration. The search continues until suitable profile parameters have been identified. Because we must account for both the angle-dependence and signal-dependence, however, steps must be taken to make the search viable and manageable.

Having described the PML signal-dependence, we now consider angle-dependence. Similar to an optimization problem, we must create a criterion on which to judge the results. Furthermore, recall that the decay rate, α in the PML is given as $\alpha = \sigma \eta_0 \cos \theta$. Therefore, regardless of the profile, we can expect significant reflection from a PML with a perfectly conducting termination at very large angles. It is therefore appropriate to introduce a weighting function. To improve large angle performance without sacrificing near-normal performance, we have chosen the simple 'minimax' criteria with a weighting function:

$$C(S_f, p, \lambda/\Delta) = \max_{\theta} \{ |\Gamma(\theta, \sigma_f, p, \lambda/\Delta) W(\theta)| \} \quad (2)$$

where $W(\theta)$ is a weighting function and $\Gamma(\theta, \sigma_f, p, \lambda/\Delta)$ is the reflection from the PML. Note that C is not a function of θ . We seek the values of S_f and p that minimize C . This function is well-suited to the wide-angle minimum reflection because it does not penalize for a realization of $\Gamma(\theta)$ that has a large variation. The weighting function, $W(\theta)$ is equal to unity (0dB) up to and including 60° , at which point it drops off at $-12dB/5^\circ$ to a

minimum of -36dB at 75° . This drop off from 60° to 75° accounts for the reduction in performance predicted by Eq. (1). Incident angles larger than 75° are not considered.

To deal the the signal-dependence, we describe reflection as a function of spatial points per wavelength and break incident signals into three groups. These groups have, respectively, a minimum of 15, 20 and 30 spatial point per wavelength sampling for frequencies that have power densities within -6dB of the maximum power density. The results reported in this work are therefore worst case reflections for signals adhering to these groups.

V. ONE DIMENSIONAL PROJECTION METHOD

Running a 2-dimensional FDTD simulation for every S_f and p is time prohibitive. Furthermore, depending on the type of excitation used, analysis of the reflection at a single angle is quite complicated. Fortunately, it has been shown 2-dimensional uniform plane waves may be represented using 1-dimensional FDTD simulations[21]. Reflections from a PML may also be analyzed using 1-dimensional FDTD simulations. Basically a 1-dimensional “slice” of a 2-dimensional wave is taken in the direction normal to the PML under test. This 1-D wave moves with a velocity $c/\cos\theta$, where c is the velocity of the 2-D wave in the propagation direction perpendicular to the planar wavefront. The PML only attenuates in the direction normal to its interface, so there are no transverse variations in the PML, but since the normal variations are specified by the same FDTD formalism as the 2-D, the 1-D FDTD simulation efficiently demonstrates the complete reflection and transmission characteristics of the PML.

This can be shown analytically. It has been shown [13] that the time harmonic curl equations for TM waves inside a continuous PML may be written as:

$$-\frac{\partial E_z}{\partial y} = j\omega\mu_0 H_x \quad (3a)$$

$$\frac{\partial E_z}{\partial x} \left(\frac{1}{1 - j\sigma/\omega\epsilon_0} \right) = j\omega\mu_0 H_y \quad (3b)$$

$$\frac{\partial H_y}{\partial x} \left(\frac{1}{1 - j\sigma/\omega\epsilon_0} \right) - \frac{\partial H_x}{\partial y} = j\omega\epsilon_0 E_z \quad (3c)$$

The time-harmonic plane wave solution for Eq. (3) is of the form:

$$E_z = E_0 e^{-jk_x^{pml}x - jk_y y + j\omega t} \quad (4a)$$

where the time dependence $e^{j\omega t}$ will be suppressed in the following:

$$\mathbf{H} = \frac{E_z}{\eta_0} (\hat{x} \sin \theta - \hat{y} \cos \theta) \quad (4b)$$

with

$$\begin{aligned} k_x^{pml} &= k_x \left(1 - j \frac{\sigma}{\omega\epsilon_0} \right) \\ &= \left(\frac{\omega}{c} - j\sigma\eta_0 \right) \cos \theta \end{aligned} \quad (5)$$

Note that the velocity of the 1-D slice inside the PML is $c/\cos \theta$. So the 1-D solution should take the form:

$$E_{1Dpml} = \hat{z} E_z \quad (6a)$$

$$H_{1Dpml} = \hat{y} H_y \quad (6b)$$

at $y = 0$.

Because the solution in Eq. (6) does not satisfy Maxwell's curl equations, Ampere's Law must be modified. Taking the partial derivative of H_x and H_y from Eq. (4b) with respect to y and x respectively at $y = 0$, with $k_y = (\omega/c) \sin \theta$ gives:

$$\frac{\partial H_x}{\partial y} = -jk_0(\sin^2 \theta) \frac{E_o}{\eta_0} e^{-jk_x^{pml}x} \quad (7a)$$

$$\frac{\partial H_y}{\partial x} = jk_0(\cos^2 \theta)(1 - j\sigma/\omega\epsilon_0) \frac{E_o}{\eta_0} e^{-jk_x^{pml}x} \quad (7b)$$

These are equated as:

$$\frac{\partial H_x}{\partial y} = \frac{-(1 - \cos^2 \theta)}{\cos^2 \theta(1 - j\sigma/\omega\epsilon_0)} \frac{\partial H_y}{\partial x}$$

Thus, the left hand side of Eq. (3c) becomes:

$$\frac{1}{\cos^2 \theta (1 - j\sigma/\omega\epsilon_0)} \frac{\partial H_y}{\partial x}$$

Now Eqs. (3) for the 1-D curl equations in the time domain becomes:

$$\frac{\partial E_{z1Dpml}}{\partial x} = \mu_0 \frac{\partial H_{y1Dpml}}{\partial t} + \frac{\sigma\mu_0}{\epsilon_0} H_{y1Dpml} \quad (8a)$$

$$\frac{1}{\cos^2 \theta} \frac{\partial H_{y1Dpml}}{\partial x} = \epsilon_0 \frac{\partial E_{z1Dpml}}{\partial t} + \sigma E_{z1Dpml} \quad (8b)$$

Eq. (8) should be recognized as the lossy Faraday and Ampere Laws with an impedance η^{pml} matched to that of free space: $\sqrt{\mu_0/\epsilon_0} = E_z/H_y$.

It can be shown that there are 1-dimensional equivalents for TE waves inside the PML as well. The derivation proceeds along similar lines.

To verify the proceeding derivation, we have run 2-dimensional FDTD simulations and compared them with the 1-dimensional equivalents. The results are found in Figure 6. In Figure 6, a TM plane wave incident at 30° is partially absorbed by an 8-layer PML at the back wall. The incident wave is propagating toward the left, rear corner. Figure 6b shows the reflected field. Note that the incident and reflected fields obeys Snells's law and that the reflected field is several orders of magnitude smaller than the incident field. Figure 6c shows a cross-section of the reflected field from the 2-dimensional grid taken at 50Δ from the right hand side as well as the reflected field from 1-dimensional simulation. The time stamp of the plots of Figure 6c are 150 timesteps from those of 6b. Note the similarities in the pulse shape and amplitude. The 2-dimensional plane wave was created by using 1-dimensional FDTD simulations on the left and right hand sides of the grid as described in [21]. The absence of additional reflection artifacts along the left and right sides indicate that the existing reflected field is generated from the PML only.

The 1-dimensional method comprises a simple and efficient means to test PML reflections. The attenuation is in one direction, the direction normal to the boundary. A careful

dispersion analysis shows that the error from representing a 2-dimensional wave with a 1-dimensional slice is comparable to the error involved in numerical dispersion which is inherent to 2-dimensional FDTD simulations. In two dimensions, Maxwell's curl equations for the PML require either a supplemental equation, as used in [13], or split-field equations, as used by [6].

Armed with this powerful tool, the search for S_f and p is greatly simplified and hence can be automated. A search program may iterate through different values of θ , S_f and p . Furthermore, since reflections can be described in terms of spatial points per wavelength, we may use a single wide-band incident wave as a test signal and be able to account for the signal-dependence of the PML as described in the previous sections.

VI. RESULTS

Using the techniques described in the previous sections, several automated searches were performed to identify the PML profile parameters S_f and p that will yield superior performance.

For each profile tested, a 1-dimensional FDTD simulation was run for every angle between and including 0° and 75° at 5° increments. In each case a wide-band test signal was the incident wave. The reflection coefficient is calculated by:

$$|\Gamma(\lambda/\Delta)| = 20 \log \max \left\{ \left| \frac{fft(E^s)}{fft(E^i)} \right| W(\theta) \right\}$$

where $\Gamma(\lambda/\Delta)$ is the reflection, E^s is the scattered field and E^i is the incident field.

Our experiments indicate that $\Gamma(\lambda/\Delta)$ becomes non-increasing after a certain value of points per wavelength, usual between 100 and 150. Therefore, once $\Gamma(\lambda/\Delta)$ has been calculated for a given profile and angle, the maximum between 15, 20, 30 and 200 points per wavelength respectively are recorded. Then these maxima are multiplied by the weighting function as necessary to yield the overall maximum for a given profile. The parameters that yield the smallest overall maximums are found in Table 1 through Table 4.

As the results reported in the tables are 'worst case' for the discrete frequency with the greatest reflection, depending on the incident time domain signal, results may be significantly better than those reported in the tables. For example, a gaussian pulse with a $10\Delta t$ time constant has been used as the incident signal to test plane waves incident on the PML at various angles for some of the profiles found in Tables 1 and 2. This is the same test signal used in [6]. The reflections coefficient as calculated by

$$|\Gamma| = 20 \log \max_{\theta} \left| \frac{E^s}{\max_t E^i} \right|$$

have been plotted versus angle in Figure 7 for the 8-layer PML with conductivity profile parameters $S_f = 0.016$ and $p = 3.74$. The results reported by Berenger [6] have also been plotted. Note that there is more than an order of magnitude improvement at both 45° and 75° . Using the same test signal, the reflection coefficient for a 10-layer PML with conductivity profile parameters $S_f = 0.025$ and $p = 3.91$ has also been plotted on Figure 7. Using an additional two layers, we may further decrease $\Gamma(\theta)$ by 2 orders of magnitude at 75° .

The improved performance shown by the 10-layer PML at large angles suggests that the weighting function may not be needed for larger PMLs. To explore this idea, several additional searches were performed without including the weighting function, the results for these searches are found in Tables 5 through 7. Although the results for searches that do not employ the weighting function are not quite as good overall, these profiles provide excellent performance across the entire range of 0° to 75° .

VII. CONCLUSIONS

We have explored the mechanisms that govern reflection from PML absorbing boundaries. In so doing, we have presented a simple means of accounting for the signal-dependence of PML reflections. Furthermore, we have shown that it is the 'discrete'

parameters points per wavelength λ/Δ and loss per PML sublayer S which govern this signal-dependence and not the 'physical' parameters ω and σ .

We have presented a fast and simple method to determine the reflection from a PML due to uniform plane wave at arbitrary angle. This method does not require the use of an anisotropic media or a split-field formulation. Plane wave decomposition may be used to gain insight into the overall reflection from arbitrary waves.

Employing these methods, we have conducted searches for conductivity profiles that provide improved performance. Assuming adequate sampling, profile parameters have been provided which will insure exceptional performance over a wide variety of scattering problems. This approach removes the burden of seeking adequate ABC absorption from the FDTD user.

Despite the obvious utility of the method, caution must be exercise to ensure that extreme angles ($> 75^\circ$) are minimized. The method developed here also does not account for evanescent waves.

ACKNOWLEDGMENT

The authors are grateful to Ann W. Morgenthaler and Eric L. Miller for many helpful discussions.

REFERENCES

- [1] Yee, K.S., "Numerical Solution of Initial Boundary-Value Problems Involving Maxwell's Equations in Isotropic Media," *IEEE Trans. Ant. Prop.*, Vol. AP-14, No. 1, pp. 302-307, May 1966.
- [2] Engquist, B., and Majda, A., "Absorbing Boundary Conditions for the Numerical Simulation of Waves," *Mathematical Computation*, Vol. 31, pp. 629-651, 1977.
- [3] Mur, G., "Absorbing Boundary Conditions for the Finite-Difference Approximation of the Time-Domain Electromagnetic-Field Equations," *IEEE Transactions on Electromagnetic Compatibility*, Vol. EMC-23, No. 4, pp. 377-382, November 1981.

- [4] Mei, K., and Fang, J., “Superabsorption—A Method to Improve Absorbing Boundary Conditions,” *IEEE Transactions on Antennas and Propagation*, Vol. 40, No. 9, pp. 1001–1010, September 1992.
- [5] Rappaport, C., and Bahrmassel, L., “An Absorbing Boundary Condition Based on Anechoic Absorber for EM Scattering Computation,” *Journal of Electromagnetic Waves and Applications*, Vol. 6, No. 12, pp. 1621–1634, December 1992.
- [6] Berenger, J., “A Perfectly Matched Layer for the Absorption of Electromagnetic Waves,” *Journal of Computational Physics*, Vol. 114, No. 1, pp. 185–200, October 1994.
- [7] Gedney, S., “An Anisotropic Perfectly Matched Layer-Absorbing Medium for the Truncation of FDTD Lattices,” *IEEE Transactions on Antennas and Propagation*, Vol. 44, No. 12, pp. 1630–1639, December 1996.
- [8] Chew, W., and Weedon, W., “A 3-D Perfectly Matched Medium of Modified Maxwell’s Equations with Stretched Coordinates,” *Microwave and Optical Technology Letters*, Vol. 7, No. 13, pp. 559–604, September 1994.
- [9] Rappaport, C., “Perfectly Matched Absorbing Boundary Conditions Based on Anisotropic Lossy Mapping of Space,” *IEEE Microwave and Guided Wave Letters*, Vol. 5, No. 3, pp. 90–92, March 1995.
- [10] Katz, D., Thiele, E. and Taflove, A., “Validation and Extension to Three Dimensions of the Berenger PML Absorbing Boundary,” *IEEE Guided Wave Letters*, Vol. 4, No. 8, pp. 268–270, August 1994.
- [11] Gribbon, M., Lee, S., and Cangellaris, A., “Modification of Berenger’s Perfectly Matched Layer for the Absorption of Electromagnetic Waves in Layered Media,” *11th Annual Review of Progress in Applied Computational Electromagnetics Symposium Digest*, pp. 498–503, March 1995.
- [12] Gedney, S., and Roden, A., “The Uniaxial Matched Layer (UPML) Truncations of FDTD Lattices for Generalized Media,” *URSI Symposium Digest*, pp. 366, 1996.
- [13] Rappaport, C., “Interpreting and Improving the PML Absorbing Boundary Condition Using Anisotropic Lossy Mapping of Space,” *IEEE Trans. on Magnetics*, pp. 968–974, May 1996.
- [14] Wu, Z., and Fang, J., “High-Performance PML Algorithms,” *IEEE Microwave and Guided Wave Letters*, Vol. 6, No. 9, pp. 335–337, September 1996.
- [15] Gianluca, L., and Gandhi, O., “On the Optimal Design of the PML Absorbing Boundary Condition

- for the FDTD Code,” *IEEE Transactions on Antennas and Propagation*, Vol. 45, No. 5, pp. 914-916, May 1997.
- [16] Wu, Z., and Fang, J., “Numerical Implementation and Performance of Perfectly Matched Layer Boundary Condition for Waveguide Structures,” *IEEE Transactions on Microwave Theory and Techniques*, Vol. 43, No. 12, pp. 2676–2683, December 1995.
 - [17] Wu, J., Lee, R., and Lee, J., “The Use of Higher Order Edge-Based Finite Elements to Improve the Accuracy of the Anisotropic Perfectly Matched Layer,” *URSI Symposium Digest*, pp. 361, 1996.
 - [18] Fang, J., and Wu, Z., “Closed Form Expression of Numerical Reflection Coefficient of Perfectly Matched Layers,” *URSI Symposium Digest*, pp. 364, 1996.
 - [19] Chew, W., and Jin, J., “Perfectly Matched Layers in the Discretized Space: An Analysis and Optimization,” *personal communication*, 1996.
 - [20] Wu, Z., and Fang, J., “Closed-Form Expression of Numerical Reflection Coefficient at PML Interfaces and Optimization of PML Performance,” *IEEE Microwave and Guided Wave Letters*, Vol. 6, No. 9, pp. 332-334, September 1996.
 - [21] Winton, S., and Rappaport, C., “Implementation of a Two Dimensional Plane Wave FDTD Using One Dimensional FDTD on the Lattice Edges,” *97 ACES Symposium Digest*, March 1997.
 - [22] De Moerloose, J., and Stuchly, M., “An Efficient Way to Compare ABCs,” *Antennas and Propagation Magazine*, Vol. 38, No. 1, Feb 1996.

λ/Δ	S_f	p	Γ (dB)
15	0.0152	3.77	-90.27
20	0.0160	3.74	-93.71
30	0.0177	3.78	-98.83

Table 1 8-layer PML

λ/Δ	S_f	p	Γ (dB)
15	0.0215	3.93	-106.17
20	0.0179	3.99	-108.95
30	0.0193	6.98	-114.17

Table 2 10-layer PML

λ/Δ	S_f	p	Γ (dB)
15	0.026	4.20	-119.69
20	0.020	4.50	-122.37
30	0.021	4.40	-127.29

Table 3 12-layer PML

λ/Δ	S_f	p	Γ (dB)
15	0.020	4.70	-130.73
20	0.023	5.00	-138.12
30	0.023	5.00	-138.12

Table 4 14-layer PML

λ/Δ	S_f	p	Γ (dB)
15	0.0260	3.90	-102.30
20	0.0295	3.90	-106.42
30	0.0345	3.90	-107.43

10-layer PML No Weighting

Table 5

λ/Δ	S_f	p	Γ (dB)
15	0.0270	4.2	-116.64
20	0.0305	4.1	-120.83
30	0.0365	4.1	-122.49

12-layer PML No Weighting

Table 6

λ/Δ	S_f	p	Γ (dB)
15	0.028	4.5	-129.13
20	0.028	4.5	-129.13
30	0.031	4.9	-132.66

14-layer PML No Weighting

Table 7

FIGURE CAPTIONS

Figure 1 Lattice geometry and typical plane wave incident on N-layer conductivity profiled PML ABC.

Figure 2 Reflection as a function of angle for PML half spaces with constant conductivity profile, $\sigma_c = 0.45(*)$, $0.30(o)$, and $0.15(+)$.

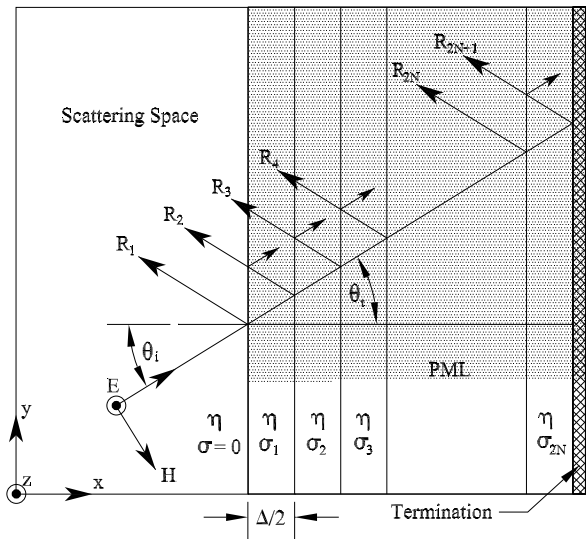
Figure 3 Three time domain signals and their associated power spectral densities. For Signal 1, $\Delta t = 2ps$, Signal 2, $\Delta t = 1.8ps$, Signal 3, $\Delta t = 1.7ps$.

Figure 4 Reflection coefficient as a function of λ/Δ for various PMLs as determined using three different signals. $\Delta = .6mm$. Signal 1(-)($\Delta t = 2ps$), Signal 2 (..)($\Delta t = 1.8ps$), Signal 3 (-)($\Delta t = 1.7ps$).

Figure 5 Reflection coefficient as a function of λ/Δ for various PMLs as determined using three different signals. Signal 1(-)($\Delta t = 2ps$, $\Delta = .6mm$, $S_f\Delta = 0.018$, $p=3.675$), Signal 2(..) ($\Delta t = 1.8ps$, $\Delta = .54mm$, $S_f\Delta = 0.0162$, $p=3.675$), Signal 3(-)($\Delta t = 1.7ps$, $\Delta = .51mm$, $\sigma_f\Delta = 0.0153$, $p=3.675$).

Figure 6 Gaussian plane wave incident on an 8-layer PML located at $x = 392\Delta$ at 30° (from lower right to left). These are views of a portion of a 400×100 grid. a) Incident field, b) scattered field, showing specular reflection from PML five orders of magnitude lower than the incident field and c) cross-section of reflected field from 2-D simulation and reflected field from 1-D simulation. Time slice of graphs in (c) +150 time steps of those in (a) and (b).

Figure 7 Reflection coefficients of N-layer PMLs as a function of θ . $(N, p, \sigma_f\Delta) = (8, 3.74, 0.016)(-), (10, 3.91, 0.025)(--)$ and the results reported in [6]($*$). The values of W and Δ used in the simulations were chosen to match those used in [6].



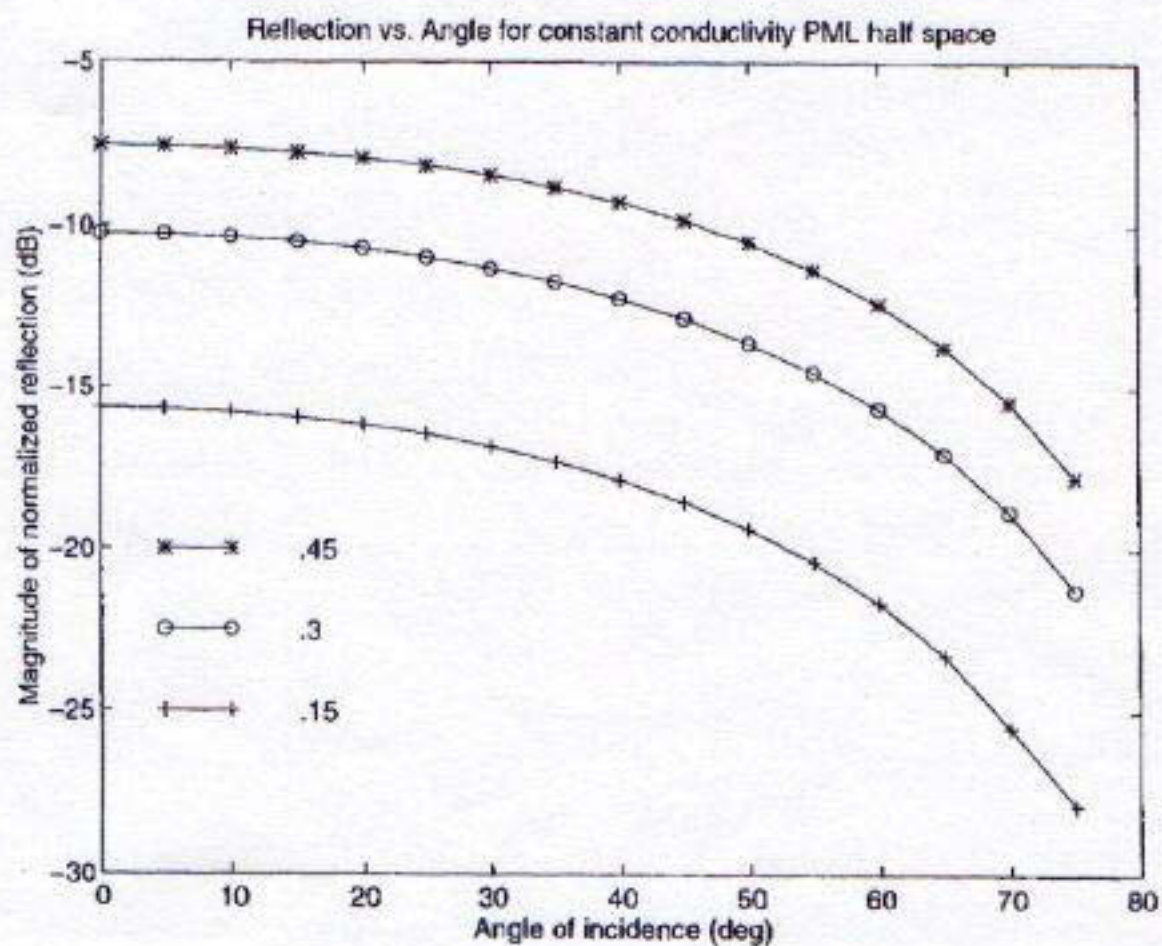
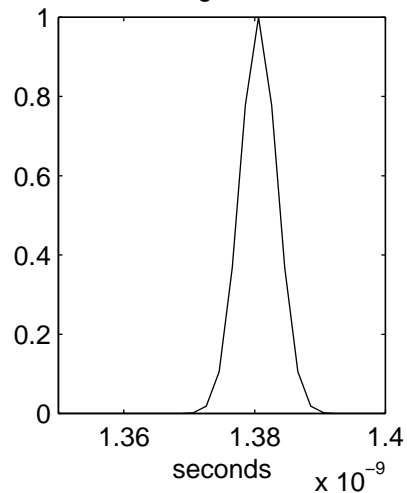
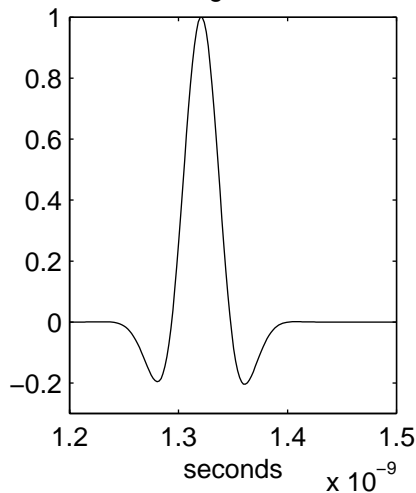


Figure 2

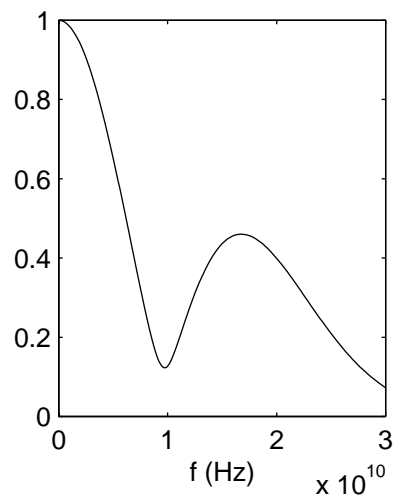
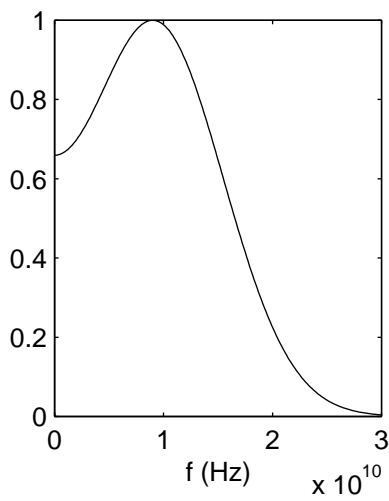
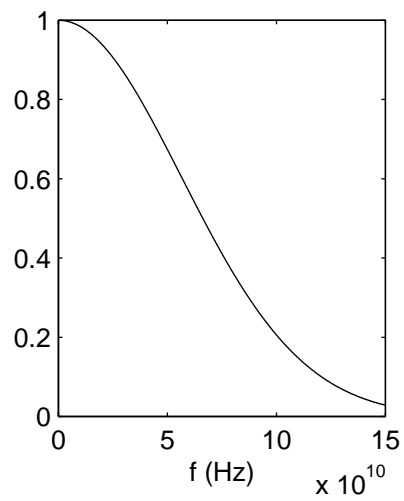
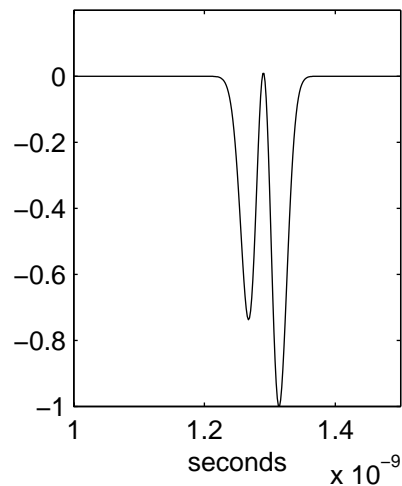
Signal 1

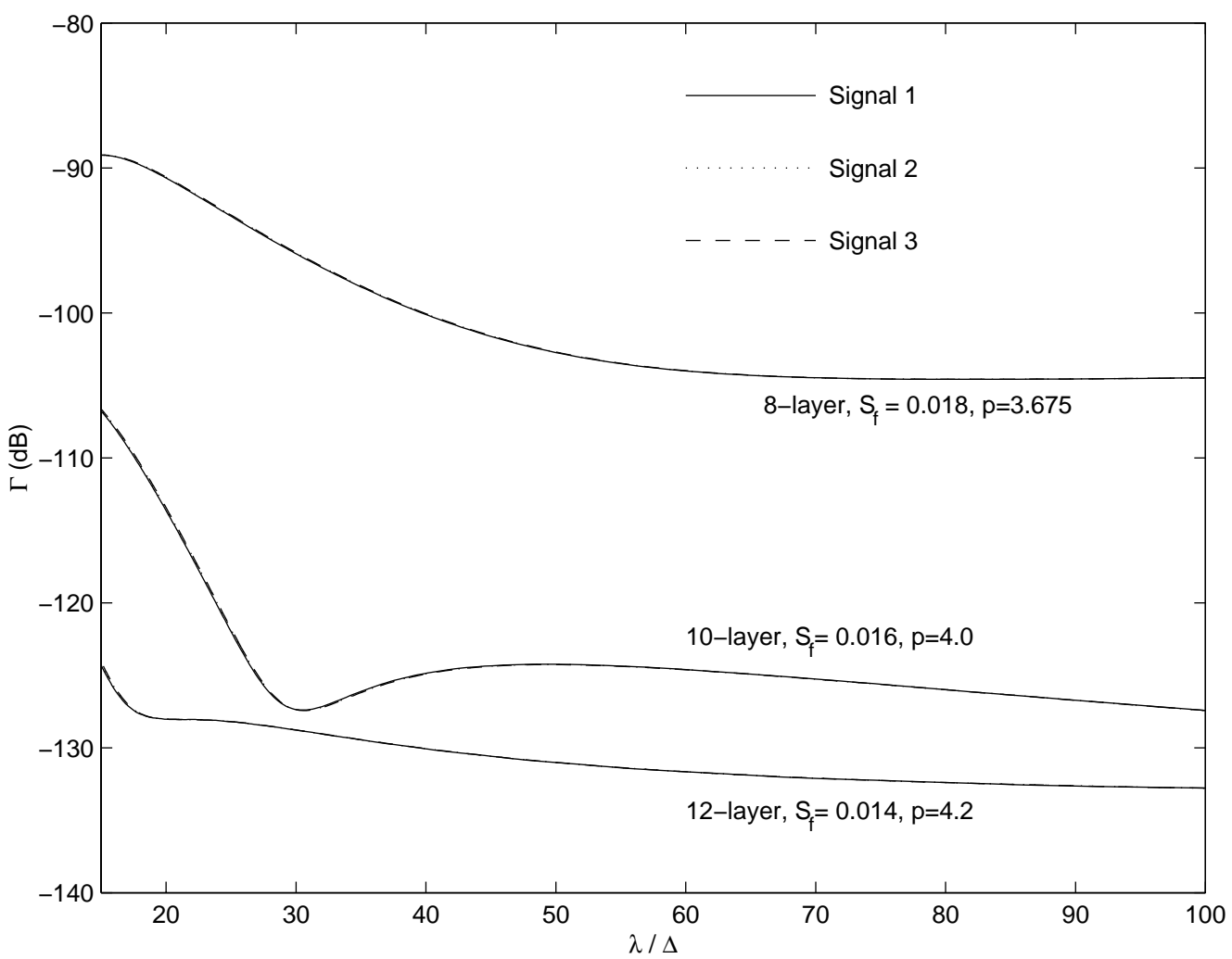


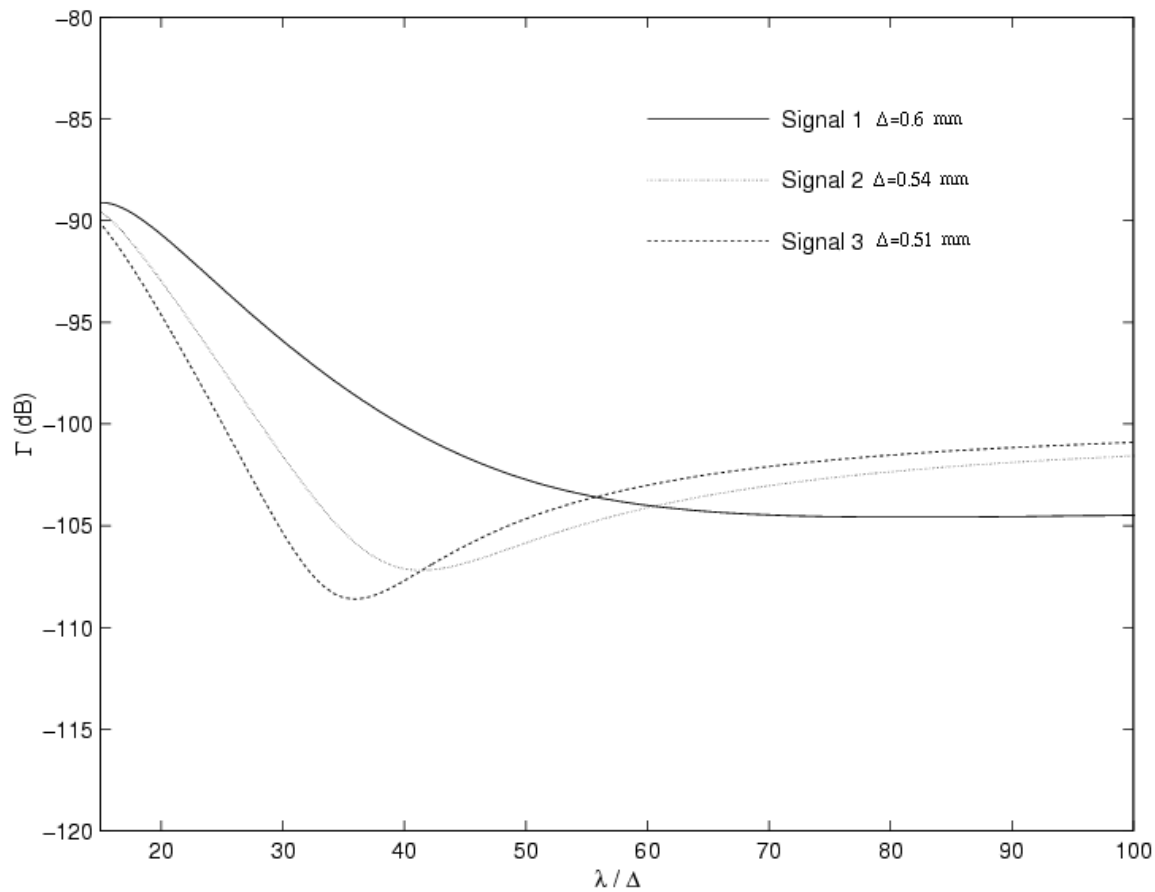
Signal 2



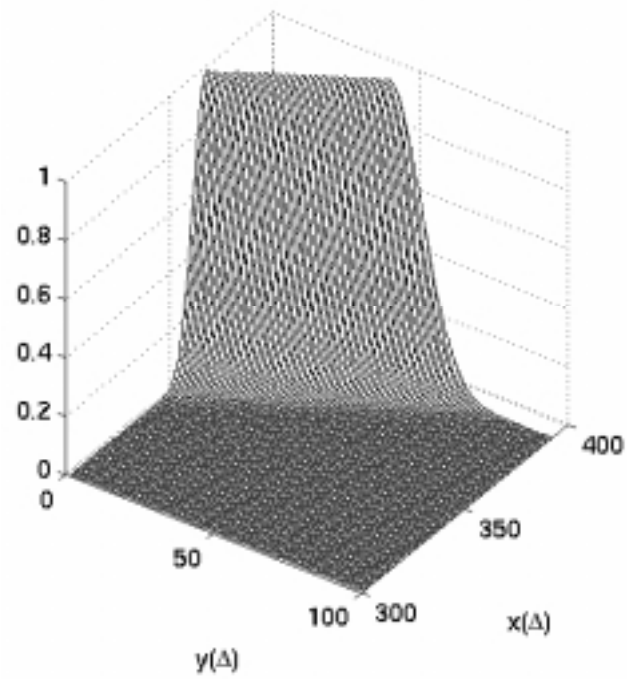
Signal 3



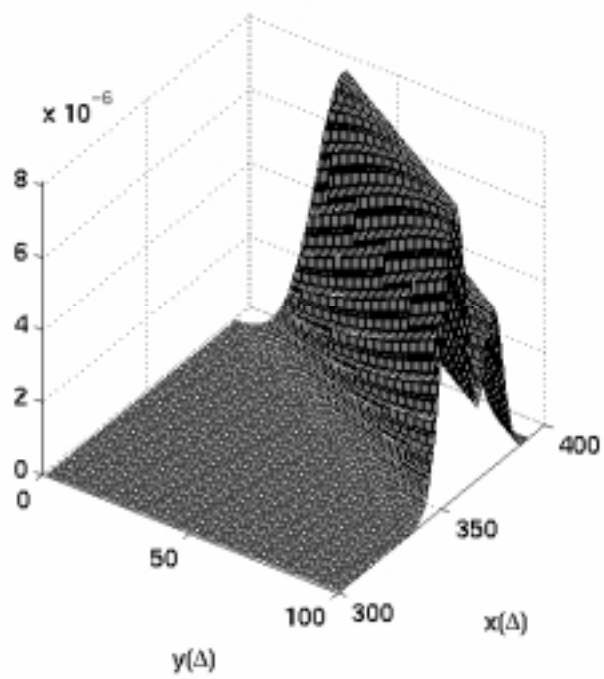


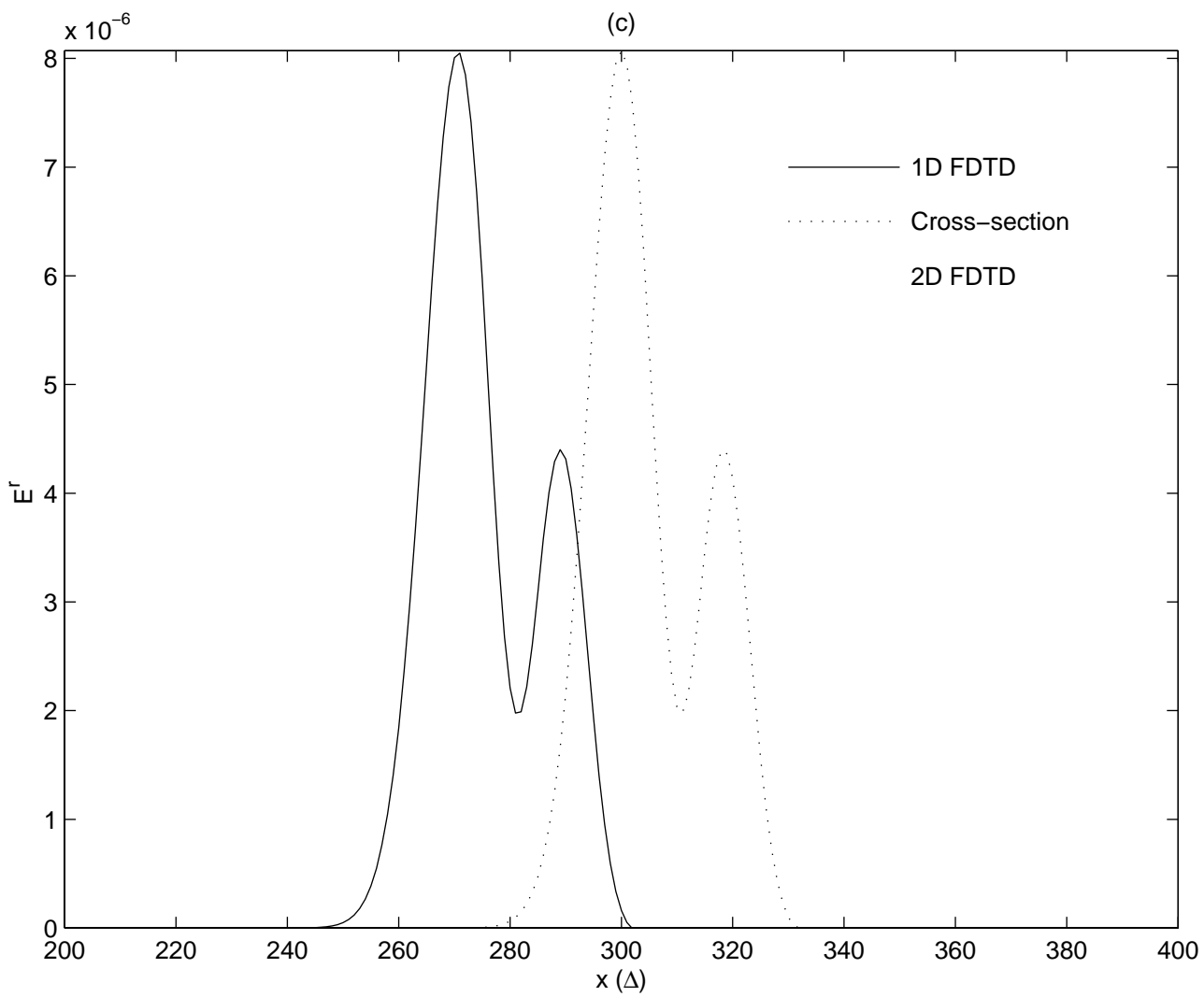


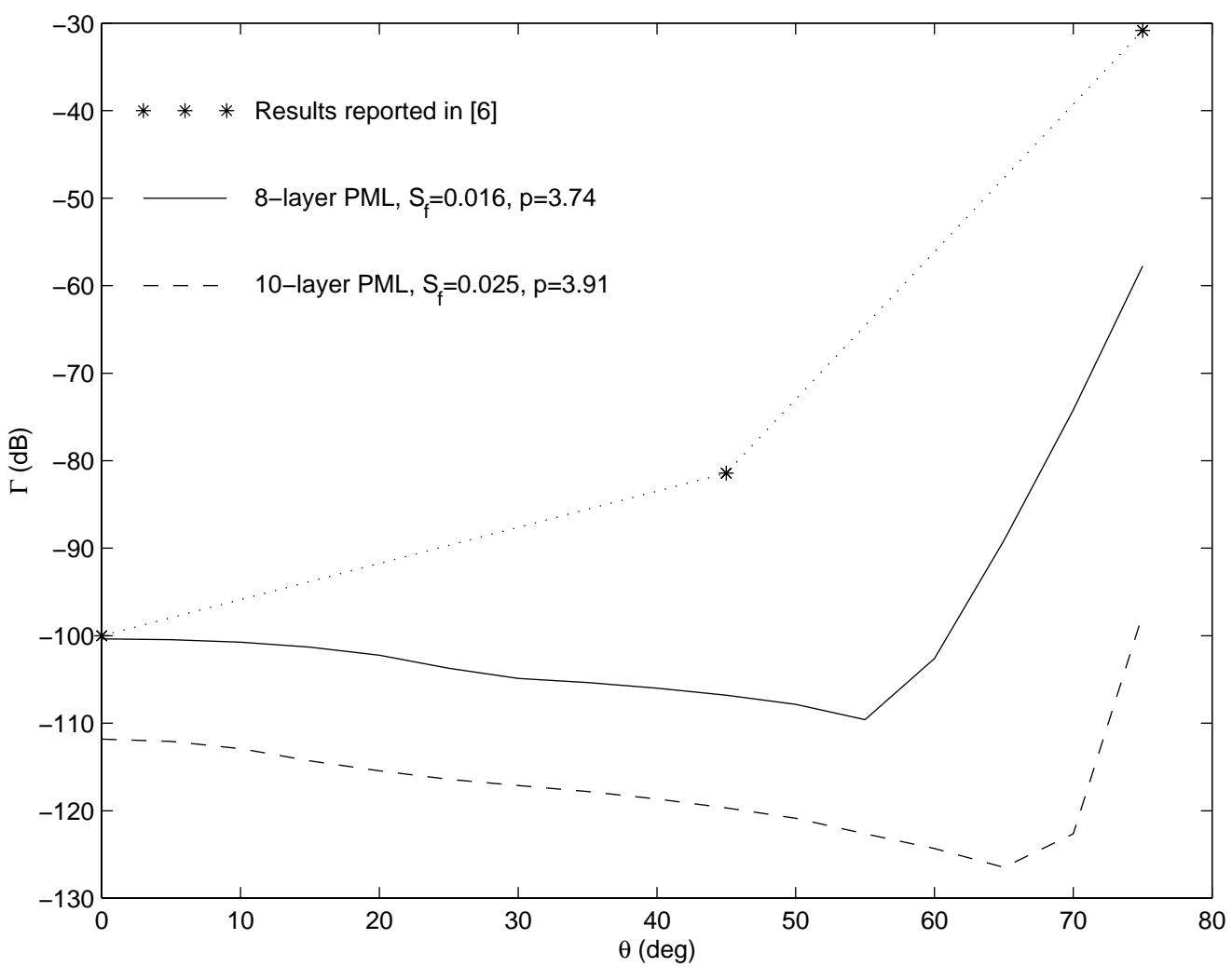
(a)



(b)







Handheld forward-looking focused array mine detection with plane wave excitation

Carey Rappaport^{*a}, Stephen Azevedo^b, Tom Rosenbury^b, Jamie Gough^b, and Dongping Jin^a

^aDept. Elect. and Comp. Eng., Northeastern University, Boston, MA 02115

^bLawrence Livermore National Laboratory, Livermore, CA

ABSTRACT

A novel handheld time-domain array GPR antipersonnel mine detection system prototype has been developed. Using an offset paraboloidal reflector antenna to collimate rays from an ultra-wideband feed, the transmitted microwave impulse is concentrated forward, in front of the antenna structure. The resulting wave is a non-uniform plane wave over the portion of ground being investigated, and is incident at 45 deg. to normal. As such, much of the ground reflected wave is directed further forward, away from the operator, the reflector, and the receiving antennas, thereby reducing clutter. However, the wave transmitted into the ground, which interacts with the target, tends to have significant backscatter returning toward the receiving antennas. These receiving antennas are configured in a 2 by 2 array to provide spatial focusing in both along and cross-track directions. This is accomplished by measuring and comparing the backscattered signals at each receiver in the narrow time window between the times when the ground reflected wave passes the receiver and before this wave re-reflects from the reflector components.

Two-dimensional FDTD simulation of this parabolic reflector transmitter indicates that it generates a beam with a non-uniform planar wavefront, which scatters from rough ground primarily in the forward direction. The wave transmitted into the ground is also planar, propagating at the angle of refraction, and scattering fairly isotropically from a small penetrable target.

This system has been built and tested at Lawrence Livermore National Laboratory, using a very narrow pulse shape. LLNL's Micro-Impulse Radar (MIR) and custom-built wideband antenna elements operate in the 1.5 to 5 GHz range. One particular advantage of using the MIR module is its low cost: an important feature for mine detectors used in developing countries. Preliminary measured data indicates that the surface clutter is indeed reduced relative to the target signal, and that small non-metallic anti-personnel mines can be reliably detected at burial depths as shallow as 1 inch in both dry.

Keywords: GPR, mine detection, focused array, reflector antenna.

1. INTRODUCTION

There is considerable interest in the detection of shallow buried objects, such as mines and buried pipes, many of which are non-metallic. The detection of non-metallic objects, such as anti-personnel mines is challenging, since the differences between the dielectric constant of the target and the surrounding soil is relatively low. Since this contrast is low, the target scattering is small, in particular compared to the scattering by the ground surface.

To address the detection of plastic antipersonnel mines, which is also challenging for conventional metal detector, a novel handheld time-domain array GPR system prototype has been developed. Designed at Northeastern University, this detector makes use of an offset paraboloidal reflector antenna and an array of wideband receiving elements to provide moderate forward-looking capability with reduced surface clutter.

This new GPR system reduces this ground clutter by illuminating the sample ground surface with a forward propagating, quasi-planar wave, and receiving the scattered signals with a two-dimensional multistatic array. Since the scattering by a small target is relatively isotropic, while scattering by the ground is primarily specular, a planar transmitted signal is well suited for shallow GPR detection. Plane wave illumination has another advantage beside clutter reduction compared to point source excitations: for a given target burial depth, the wave incident on a target from a plane wave source will always scatter the same way. For a point source, the incident wave on a given target will be illuminated from the side for one transmitter position and directly above for another. The constant exposure angle for a planar wave excitation makes processing the returned signals more straightforward.

We have used the finite-difference time-domain (FDTD)¹ method to electromagnetically model our novel GPR configuration. We implemented the 2-D FDTD code to simulate the generation of the non-uniform plane wave, the scattering by the modeled dispersive soil ground surface, the scattering by the target, and the retransmission back into the air, confirming the clutter minimizing characteristics of this mine detector. Computer simulation guides optimum size and offset section selection of the reflector, the choice of height of reflector off the ground, and the best positions of receiving elements.

A prototype system has been fabricated and tested at Lawrence Livermore National Labs (LLNL) using an excitation signal that is sufficiently short in time duration to resolve small targets and discriminate the ground surface from a shallow buried target. The multistatic array concept provides for additional clutter rejection and time-domain focusing². This focusing is accomplished by measuring, comparing, and summing the backscattered signals at each receiver in the narrow time window between the times when the residual ground reflected wave passes the receiver and before this wave re-reflects from the reflector components.

2. GROUND PENETRATING RADAR GEOMETRY

The radar system makes use of an inclined transmitted plane wave impulse excitation in conjunction with a multistatic array of receivers. It is an essential aspect of the radar design to send out a very short pulse, and accurately measure the returned signal in very small time increments. The radar system would not function with stepped frequency excitation.

2.1 Parabolic Reflector Transmitter

In order to generate a plane wave near the ground surface, the transmitting antenna must collimate outgoing rays in a beam that illuminates the sampled region of ground. Not only must this antenna generate a planar wavefront, but also its aperture must extend a distance comparable to the distance to the ground. In other words, the ground must be in the near field of the antenna.

A reflector whose surface is a section of a paraboloid of revolution collimates rays emanating from a feed at its focal point. As long as the section of the paraboloid is as deep and wide as its height above ground, the wave it generates will remain planar over the illuminated region of ground. Outside the illuminated spot on the ground, the transmitted wave is not planar, but since most of the wave power is concentrated within the beam, this diverging field is relatively unimportant.

The orientation of this quasi-planar transmitted wave must be carefully selected to balance the clutter reduction and safety advantages of forward-look with packaging disadvantages. For greater wave incidence angles, the forward edge of the reflector must either be higher or farther forward. And since greater height requires greater aperture size, both options demand increased reflector size. As a compromise, the incidence angle was chosen to be at 45 deg. to normal. To maintain close proximity of the reflector to the ground surface, an offset section of the parent paraboloid is chosen for the reflector antenna.

Figure 1 shows the geometry of the parabolic reflector and the way it directs rays from the transmitting feed to the ground. Diverging rays leaving the transmitter reflect from the paraboloidal surface, emerge as parallel rays, in such manner as to keep the path length from the feed to an inclined wavefront constant. The inclined wavefront is perpendicular to -- and propagates along -- the axis of revolution of the parent paraboloid, which includes the parabola focus and vertex. Also, the reflector produces a beam of microwave energy with an abrupt drop in power outside the ray tube bounded by the perimeter of the reflector (indicated with rays in Figure 1). The distance along the rays from reflector to ground S is equal to the projected reflector diameter D (distance between the furthest rays), so the rays representing the transmitted wave will be parallel, and the wave will be planar. The governing equation for the nearfield of the reflector is: $S \ll 2 D^2 / \lambda$, so when the reflector is positioned close to the ground, the radiated wave is in the nearfield, and concepts of antenna gain and radiation pattern are irrelevant.

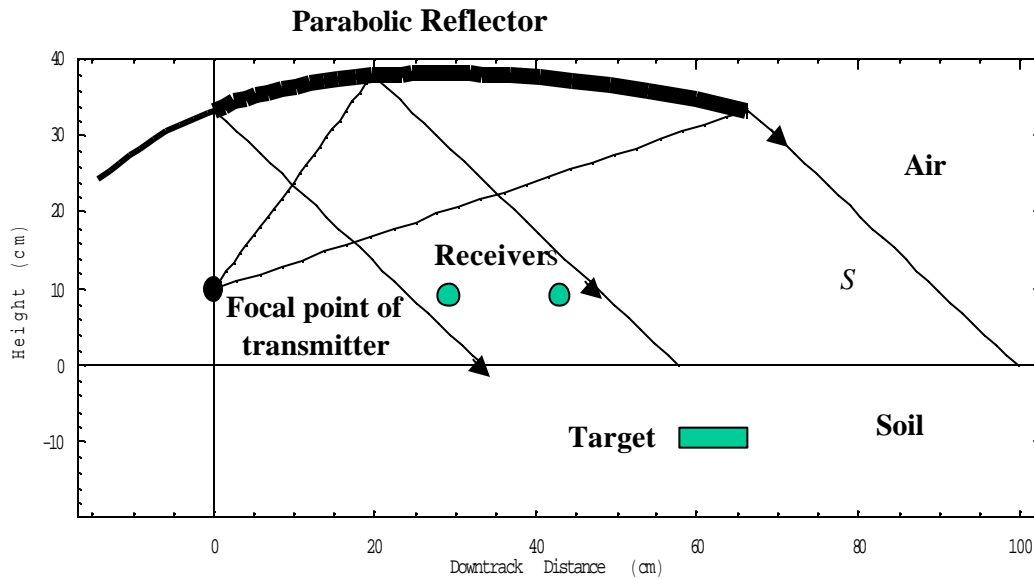


Figure 1. Geometry of the offset parabolic reflector

Because the transmitted wave diverges very little from reflector to the ground, most of the power incident from the illuminating feed is transferred to the ground. Much of the ground reflected wave is directed further forward, away from the operator, the reflector, and the receiving antennas. In addition, the wave transmitted into the ground is incident on the target in the same manner for any antenna position: always as a plane wave with constant soil path length and incident angle $\theta_i = \sin^{-1}(1/\sqrt{2\epsilon_r})$. Since the scattering from an electrically small buried target is primarily isotropic, there will be a significant backscattered signal, propagating oppositely to the surface clutter signal, returning toward the receiving antennas.

An offset section of the paraboloid is selected to avoid blockage of rays by the feed structure. In contrast to offset reflectors used in communications applications, this offset section is particularly deep, extending from the vertex past the focal point by twice the focal length, giving an F/D ratio of the parent paraboloid of about 0.15. The best offset section extends from 45 deg. to about 115 deg. from the symmetry axis, which ensures that the front and rear edges are at the same height above ground. For a parabolic focal length of 20 cm, the projected aperture diameter of the reflector is about 47 cm, which nominally illuminates an elliptical spot of ground with axes 47 and 67 cm. For a reflector positioned 33.5 cm above the ground, the center of this elliptical region is immediately below the front edge of the reflector, and all of the collimated rays from the reflector would reflect from a flat ground just missing the front of the reflector.

2.2. Receiver Array

The receivers are positioned under the reflector, but behind the point on the ground at the center of the illuminated spot. The 4 receivers are arranged in a rectangular 2 by 2 array, with the forward pair separated by about 40 cm, and the backward pair by the same distance, 20 cm behind the forward pair. Since the receivers are displaced from the centerline, they do not appreciably block any of the wave from the reflector to the ground. Since the receivers are behind the point on the ground where the central ray strikes the ground, they will completely miss the strongest ground reflection signal. The entire structure is less than 70 cm deep and stands within 40 cm of the ground, clearly small enough to be operated as a handheld sensor.

2.3. Excitation Pulse

The LLNL Micro-Power Impulse Radar (MIR) was used as the transmitter source for the fabricated mine detection prototype system². This radar source generates an impulse with pulse width of about 300 ps and frequency range from about 1.5 to 5 GHz (see Figure 2). The duration of the main pulse peak is short enough to resolve about 1cm in dry sand and 5 mm in moist loam. However, this short pulse also enhances the electrical size of the roughness of uneven ground surface. Without the radar's clutter

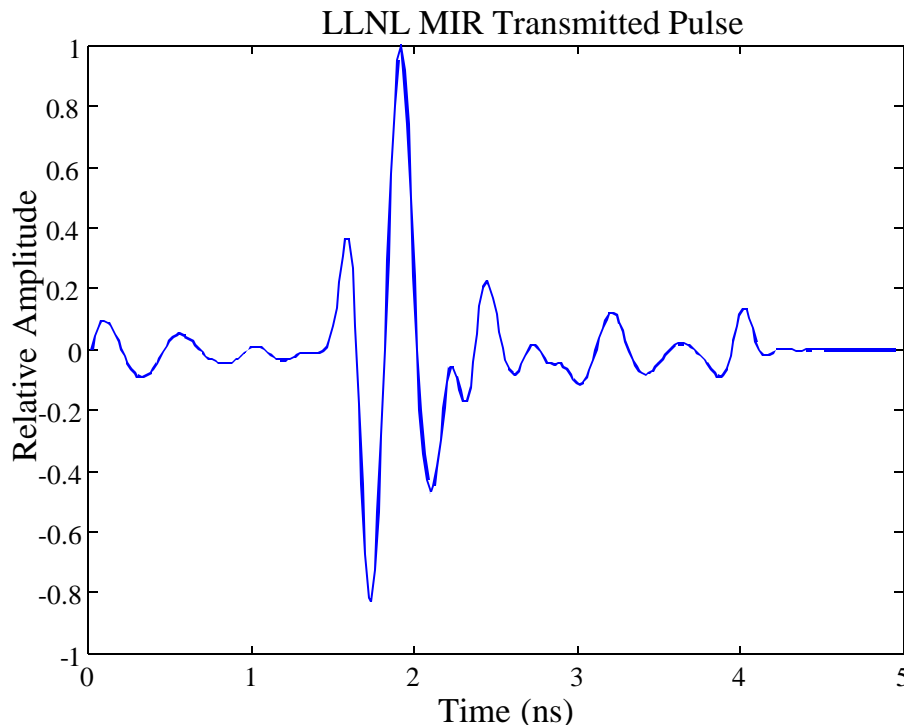


Figure 2. Measured excitation pulse of MIR radiated by the reflector feed.

suppression advantage of forward specular reflection of the ground reflected wave, it would be very difficult to distinguish the target signal. Although it would be desirable to minimize the continued “ringing” of the pulse, the large relative amplitude of the main peak and the characteristic shape of the adjacent peaks makes the processing of returned signals to identify target scattering relatively straightforward.

3. FDTD MODELING

In order to confirm the concept that a plane wave impulse will scatter from rough ground primarily in the forward direction, while scattering from a small buried target is relatively isotropic, we computationally model the problem using the Finite Difference Time Domain method (FDTD). Real soils have dielectric characteristics that are frequency dependent, and as such require convolution of the permittivity and conductivity with electric field for modeling propagation and scattering in the time domain. Since convolution over all time is computationally prohibitive, other modeling methods must be used³⁻⁵. We avoid time domain convolution by modeling the frequency dependent conductivity as a rational function of the Z-transform variable⁶ $Z = \exp(j\omega \mathbf{D} t)$:

$$\mathbf{s} (Z) = \frac{b_0 + b_1 \cdot Z^{-1} + b_2 \cdot Z^{-2}}{1 + a_1 \cdot Z^{-1}} \quad (1)$$

and assume a constant average dielectric constant. The parameters in (1) are chosen to match the experimental data, such as Puerto Rican clay loam (PRCL)⁷ with density 1.4 g/cc and moisture 10%, with a time step $\mathbf{D} t = 20\text{ps}$ ⁸: $a_1 = -0.88$, $b_0 = 0.9162$, $b_1 = -1.6766$, $b_2 = 0.7611$, and $\mathbf{e}_{\text{Avg}} = 4.2$. The expression for conductivity in Equation (1) is equivalent to the ratio $J(Z)/E(Z)$. By multiplying through by the denominators, and using the fact that Z^{-1} corresponds to a unit time delay, a finite difference equation relating current to E-field in the time domain is obtained. This relation, along with $D(t) = \mathbf{e}_{\text{Avg}} E(t)$ models the dispersive nature of the soil in a manner that is easily implemented in FDTD. The perfectly matched layer (PML) absorbing boundary condition is used in the FDTD model to terminate the computational grid^{9,10}.

The transmitting antenna is modeled in two dimensions with a point source feed backed with absorber, and a staircase approximated perfectly conducting parabolic arc reflector. The absorber backing prevents any direct radiation toward the ground, which would scatter, then reflect from the parabola and contribute to the overall clutter. Both TE and TM polarizations were modeled, with special care taken for the TE computations to avoid electric field singularities at the ends of the reflector. To test the detectability of buried objects, two cases were run for each polarization: one with a target present and one without. We examined rectangular targets with dimension 5cm by 10cm buried 5cm below the nominal ground surface. Both metallic and non-metallic targets under rough ground with mean height variation of 3cm were considered.

Figure 3 presents three images for the TM simulation for the metallic mine target. By convention, for two-dimensional numerical calculations, the transverse plane is the plane of the calculation. For this TM calculation, the electric field is perpendicular to the plane of the paper. The first image shows the total scattered field, with reflection and diffraction from the reflector, scattering from the ground, and the scattering of the mine. The second shows the field distribution for the same ground, but without the buried mine. The planar wave generated by the parabola clearly scatters from the illuminated region of rough ground in primarily the specular direction, remaining roughly planar. The third image shows the difference between the first two, representing the field scattered just by the mine, and then rescattered from the air/soil interface. This mine-scattered signal appears to be circular, showing that the electrically small target scatters almost isotropically, with very little specular reflection from the top of the flat metallic target.

It is clear which features due solely to the mine are present in the first image. Those features due to scattering from the ground, shown in the second image, are also easily distinguished in the first image. Since these two sets of features are clearly separable in space for this particular time sample, it is demonstrated that the ground surface clutter is suppressed by the geometry of the radar system. Positioning the receivers behind the point where the central ray of the reflector intersects the ground ensures that the signals received will be primarily from the mine target and not from the ground surface.

Figure 4 is similar to Figure 3, with a non-metallic mine target. In this more challenging detection case, the dielectric contrast is quite low, so the target scattering is much smaller. In fact, the contrast between the TNT target ($\epsilon_{\text{Avg}}=2.9$, $\sigma = 0.0005$) and the soil is less than between the soil and the air. It is thus much more important with the non-metallic target that the ground surface clutter be reduced with good sensor design. Figure 4 shows that again the mine target signal is spatially separated and easily distinguishable from the clutter signal.

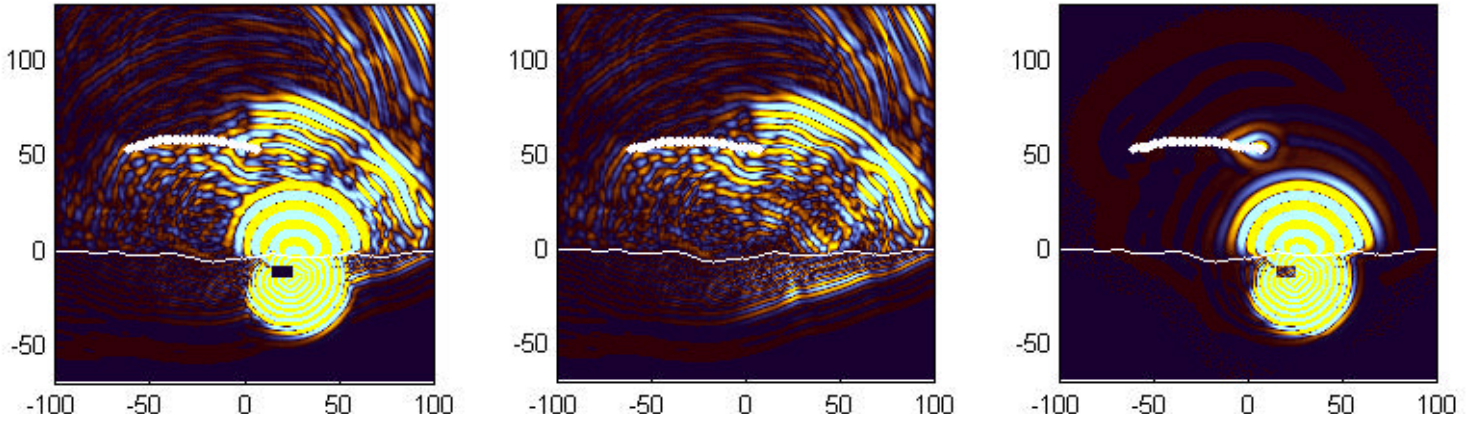


Figure 3. TM 2-D FDTD simulation of the LLNL MIR pulse reflected from the indicated 45 deg. inclined offset parabolic reflector and interacting with rough dispersive PRCL: a) Total field with rectangular metal mine target buried 5 cm, b) Ground scattered field with no target, c) Mine scattered field resulting from the difference between a) and b).

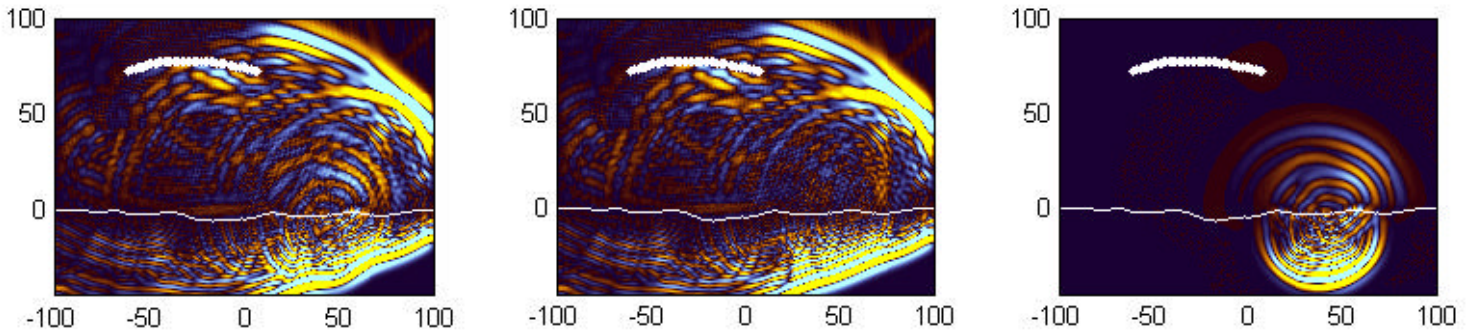


Figure 4. Same TM FDTD simulation as Figure 4, for non-metallic AP mine. Reflector height above ground is increased 25 cm to show backscattering details.

4. PROTOTYPE SYSTEM AND RESULTS

The Lawrence Livermore National Laboratory MIR source is well-suited for the mine detection problem, having the particular advantages of being small and extremely low cost, which are important features for mine detectors used in developing countries. This source was assembled with the custom-built metallic offset paraboloidal reflector. Figure 5 shows the full mine detector prototype; Figure 6 shows the device performing measurements at the test site at LLNL. In this test, a non-metallic antipersonnel mine simulant was buried in dry sand 1 in. below a very rough surface. This is a particularly challenging detection problem, because the dielectric constants of the plastic body TNT filled mine and the surrounding soil are very close. In addition, the random rough surface height variation is of the order of the height of the mine, and its burial depth. Thus, the anomaly detection is frustrated by low signal to clutter both in terms of size and contrast.

The result of processing the measured signals is shown in Figure 8, with bright areas signifying anomalies. To process the receiver signals, the direct signal from the transmitter to each receiver is removed, and the resulting signals are aligned in time to focus on suspected targets. Although there is still appreciable clutter from the rough ground, the target is still visible in the center of the image, two-thirds of the way up. The extent of the rough ground variation precludes clutter suppression using purely signal processing means. However, by ensuring that the ground-scattered signal specularly reflects away from the receivers, the target signal can be discriminated from the clutter.

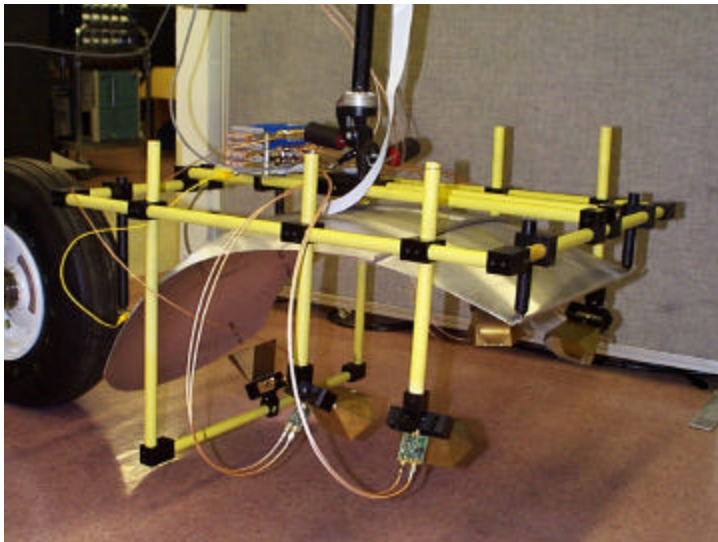


Figure 5 Offset parabolic mine detector with LLNL MIR sources and antenna elements

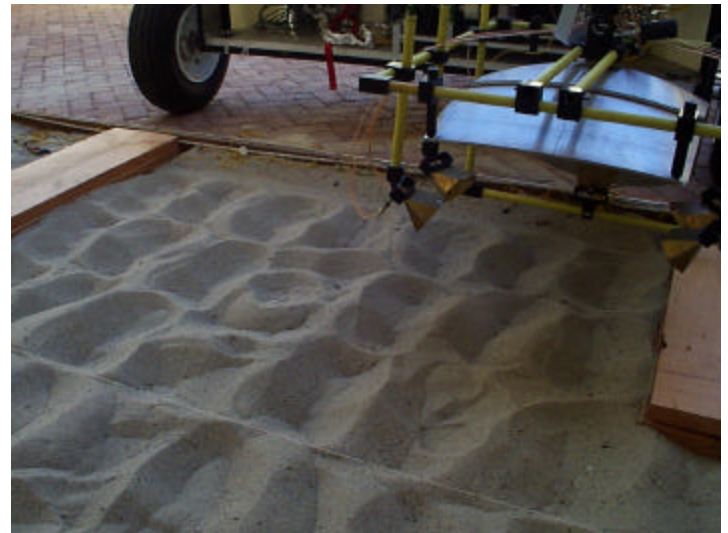


Figure 6 Parabolic mine detector under test at LLNL test site

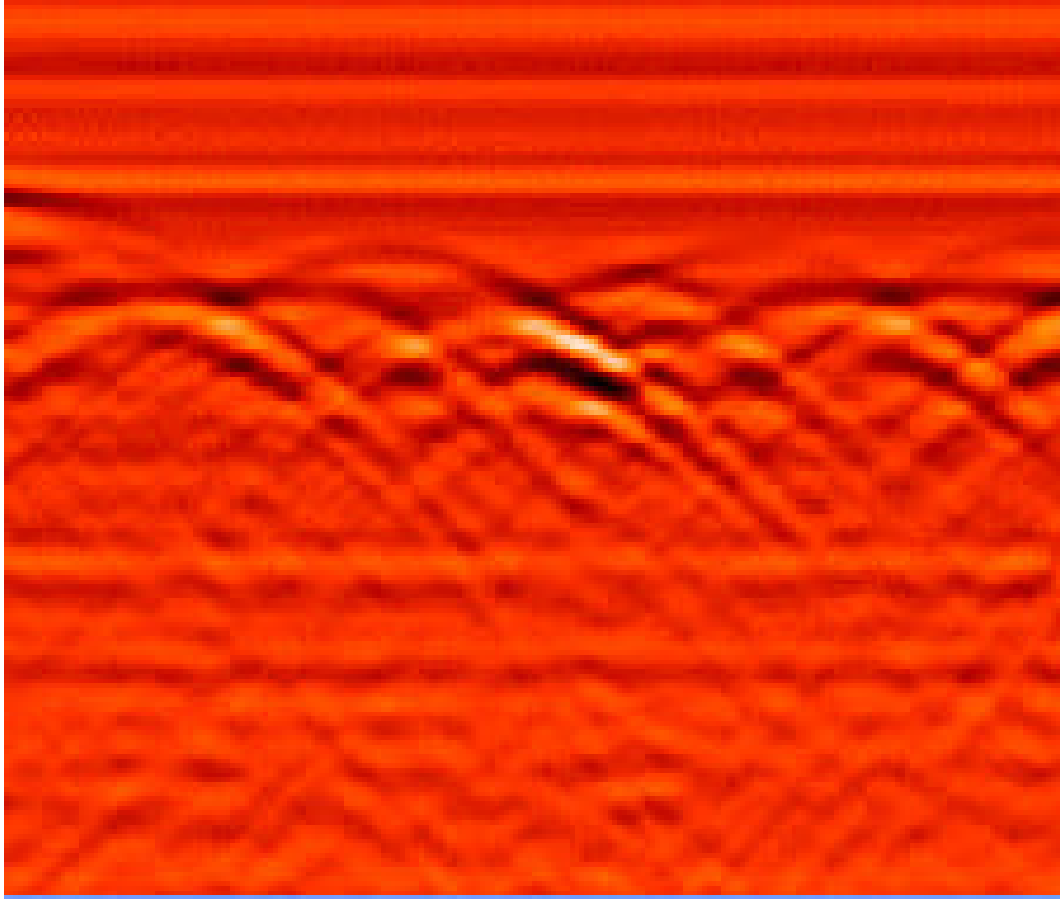


Figure 7 Detection results for the parabolic transmitter with LLNL MIR radar elements for rough dry sand with non-metallic AP mine buried 1 in.

5. CONCLUSIONS

A novel GPR mine detection system that reduces ground surface clutter has been developed and tested. The detector uses an offset parabolic reflector to generate a forward propagating plane wave to illuminate the ground and a 2-dimensional multistatic array to focus and enhance the received backscattered signal. Numerical simulations support several design assumptions:

- 1) The measured excitation gives good resolution and sufficient penetration in realistically modeled soil.
- 2) The offset parabolic reflector converts circular waves emanated by a point source feed into a non-uniform plane wave.
- 3) Choosing an offset parabolic section avoids blockage in generating a forward propagating plane wave.
- 4) A forward-looking planar transmitted wave takes advantage of specular ground reflection, spatially separating the target signal from the clutter signal.

- 5) Positioning receivers behind the central transmitted ray allows the monitoring of the backscattered mine signal while avoiding the ground clutter signal.
- 6) Even for a rough surface with a buried non-metallic mine, FDTD predicts distinct target and surface scattered signals.

Measured signals using the LLNL MIR radar for rough dry sand indicate that small non-metallic mine targets can be detected and discriminated from the cluttered background, even when shallow buried to a depth of 1 in. below a rough ground surface.

ACKNOWLEDGEMENT

This work was supported by The Army Research Office Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013.

REFERENCES

1. Yee, K., "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media," *IEEE Trans. Antennas and Propagation*, vol. 14, 1966, pp. 302-307.
2. Warrick, A., Azevedo, S., and Mast, J., "Prediction of Buried Mine-like Target Radar Signatures using Wideband Electromagnetic Modeling," *Proc. SPIE Aerosense*, vol. 3392, 1998, pp. 776-783.
3. Gandhi, O., "A Frequency-Dependent Finite Difference Time Domain Formulation for General Dispersive Media," *IEEE Transactions on Microwave Theory and Techniques*, vol. 41, April 1993, pp. 658-665.
4. Luebbers, R., Hunsberger, F., Kunz, K., Standler, R., Schneider, M., "A Frequency-Dependent Finite Difference Time Domain formulation for Dispersive Media," *IEEE Transactions on Electromagnetic Compatibility*, vol. 32, August 1990, pp. 222-227.
5. D. M. Sullivan, "Frequency-dependent FDTD methods using Z transforms," *IEEE Trans. Antennas and Propagation*, vol. 40, no. 10, pp. 1223--1230, 1992.
6. Rappaport, C., and Weedon, W., "A general method for FDTD modeling of wave propagation in arbitrary frequency dispersive media," *IEEE Trans. Antennas & Propagation*, March, 1997, pp. 401-410.
7. Hipp, J., "Soil Electromagnetic Parameters as Functions of Frequency, Soil Density, and Soil Moisture," *Proc. of the IEEE*, vol. 62, no. 1, January 1974, pp. 98-103.
8. Rappaport, C., Wu, S., and Winton, S., "FDTD Wave Propagation Modeling in Dispersive Soil Using a Single Pole Conductivity Model," *IEEE Trans. Magnetics*, vol. 35, May 1999, pp. 1542--1545
9. Berenger, J., "A Perfectly Matched Layer for the Absorption of Electromagnetic Waves," *J. of Computational Physics*, vol. 114, no 1, October 1994, pp. 185-200.
10. Rappaport, C., and Winton, S., "Using the PML ABC for Air/Soil Wave Interaction Modeling in the Time and Frequency Domains," accepted for publication in *International Journal of Subsurface Sensors and Applications*, April 2000.

Accuracy Considerations in Using the PML ABC with FDFD Helmholtz Equation Computation

Carey M. Rappaport, Misha Kilmer and Eric Miller

Center for Electromagnetics Research, Northeastern University, Boston, MA

Abstract—A careful analysis of discretization error in various implementations of Berenger’s PML absorbing boundary condition for the FDFD method applied to the Helmholtz equation is presented. Analytic and numerical formulations indicate the clear superiority of specifying the material characteristics in terms of the coupled Ampere’s and Faraday’s laws first, and then combining the first order differential equations into a discrete Helmholtz-like difference equation.

Index terms—FDFD, PML, Absorbing Boundary Conditions

I. INTRODUCTION

The introduction of Berenger’s perfectly matched layer (PML) absorbing boundary condition (ABC) [1] has significantly increased the performance and popularity of the finite difference methods of electromagnetic field computation. The time domain version (FDTD) [2-4] has seen explosive acceptance of the PML lattice termination in the past three years, while its use in the frequency domain (FDFD) [5,6] as well as in finite element methods [7-9] are considerably less popular.

One reason why FDFD is avoided is the perception that it is computationally expensive, requiring solutions of thousands or millions of simultaneous equations. Another flaw with this method is its reliance on highly non-reflective ABCs in preventing reflection from any point on the grid perimeter. In the frequency domain, small imperfections in the ABC can greatly degrade the true field solution in the propagation domain. The first concern is overcome by combining the complementary sampled Faraday’s and Ampere’s Laws into the Helmholtz Equation (thus reducing the number of unknowns by 67% in two dimensional problems), and by using preconditioned iterative methods [10]. The latter issue is successfully addressed with the PML ABC, but only with careful considerations. In particular, when using the Helmholtz equation, it is no longer possible to assume that the PML layer conductivity profile is piecewise constant. It is the intention of this study to demonstrate the subtleties required in applying the FDFD PML procedure to the Helmholtz Equation while preventing significant numerical errors.

II. FDFD PML FORMULATION

The essential aspect of the PML is the additional loss factor $S(n)$ —which depends on the direction n normal to a boundary. This unitless quantity alters the wave number in the PML according to the relation:

$$\begin{aligned} k_{\text{PML}n} &= k^{\text{rel}}(n)k_{0n} \\ k^{\text{rel}}(n) &= 1 - jS(n) \end{aligned} \tag{1}$$

This work has been supported by The Army Research Office, Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013

(where n is the distance from to the layer boundary, $k_{\text{PML}n}$ and k_{0n} are the normal components of the PML and propagation space wave numbers) which multiplies the differential element of the direction normal to the layer boundary [6]. This PML factor can be considered a complex anisotropic mapping of the normal coordinate, or a stretching of the coordinate system in the normal direction [11,12], and follows from Maxwell's equations with the assignment $S(n) = \sigma(n)/\omega\epsilon_0$, where $\sigma(n)$ is the spatially dependent PML conductivity. With this choice, using the inverse dependence on frequency ω , the time domain transforms have frequency independent exponential decay with distance into the PML. However, in most lossy layer ABCs, the conductivity $\sigma(n)$ is profiled from small values closest to the propagation space interface to a maximum value at the PML termination. Both power and geometric series profiling have been studied for optimal absorption of incident signal [13-15].

In the frequency domain, the equations for a two-dimensional transverse electric (relative to the longitudinal direction z into the paper) PML region terminating a lattice at $x = x_{max}$ (assume, without loss of generality that $x_{max} = 0$) have the form:

$$\frac{\partial H_z}{\partial y} = j\omega\epsilon_0 E_x \quad (2a)$$

$$-\frac{\partial H_z}{\partial x} \left(\frac{1}{1 - j\sigma_e(x)/\omega\epsilon_0} \right) = j\omega\epsilon_0 E_y \quad (2b)$$

$$\frac{\partial E_y}{\partial x} \left(\frac{1}{1 - j\sigma_m(x)/\omega\epsilon_0} \right) - \frac{\partial E_x}{\partial y} = -j\omega\mu_0 H_z \quad (2c)$$

For the continuous case, $\sigma_e(x) = \sigma_m(x)$. Since the PML conductivity changes with distance, so does the loss factor $S(x)$ and the normal wave number $k_{\text{PML}x}$, and therefore the propagation behavior of the wave in the PML will not have a simple linear phase dependence. For continuously varying conductivity, the normal wave number for the boundary $x = x_{max}$ can be replaced by:

$$k_{\text{PML}x} \rightarrow \int_0^x k_0 k^{\text{rel}}(x) dx \quad (3)$$

where the conductivity $\sigma(x)$ is labeled differently for Ampere's and Faraday's laws. For the both the coupled first-order FDFD and FDTD formulations, in which the electric and magnetic fields are sampled on interlocking grid points [16], the values of PML conductivity can be thought of as piecewise constant, changing every one-half grid cell: $\sigma_m(x) = \sigma_i$, for $|x - i\Delta| < \Delta/4$ and $\sigma_e(x) = \sigma_{i+\frac{1}{2}}$, for $|x - (i + \frac{1}{2})\Delta| < \Delta/4$. Although a parabolic variation of conductivity with distance $\sigma_i = \sigma_f(i/N)^2$ for an N sublayer PML gives acceptable absorption [1], it has been determined empirically that an excellent choice of profile for wide-incidence angle matching to free space using an eight grid cell PML layer in both the time and frequency domains is [5,13]:

$$\sigma_i = \left(\frac{2i-1}{16} \right)^{3.7} \frac{0.02}{\Delta}, \quad i = 1, \dots, 8 \quad (4a)$$

$$\sigma_{i+\frac{1}{2}} = \left(\frac{2i}{16} \right)^{3.7} \frac{0.02}{\Delta}, \quad i = 0, \dots, 8 \quad (4b)$$

Using the Helmholtz equation with dependent variable H_z , instead of separate Ampere's and Faraday's laws with additional dependent variables E_x , E_y , leads to a three-fold reduction in the number of unknown discretized field values. Even using fast solvers for the sparse FDFD matrix equations, this reduction leads to great improvements in computation time. It is important to maintain the distance dependencies when combining the three first order TE equations (1) into the Helmholtz equation:

$$\frac{1}{k^{\text{rel}}(x)} \frac{\partial}{\partial x} \frac{1}{k^{\text{rel}}(x)} \frac{\partial H_z}{\partial x} + \frac{\partial^2 H_z}{\partial y^2} + k_0^2 H_z = 0 \quad (5)$$

for the continuous case. Note that if $\sigma(x)$ (and hence $k^{\text{rel}}(x)$) is treated as piecewise constant, an important first derivative term would be absent.

Discretizing Equation (5) for FDFD computation can be done in three ways:

(1.) Analytically expand the second partial derivative term to give:

$$-\left(\frac{1}{k_i^{\text{rel}}}\right)^3 \frac{\partial k^{\text{rel}}}{\partial x}|_{x=i\Delta} \left(\frac{H_{i+1,j} - H_{i-1,j}}{2\Delta}\right) + \left(\frac{1}{k_i^{\text{rel}}}\right)^2 \left(\frac{H_{i+1,j} - 2H_{i,j} + H_{i-1,j}}{\Delta^2}\right) + \left(\frac{H_{i,j+1} - 2H_{i,j} + H_{i,j-1}}{\Delta^2}\right) + k_0^2 H_{i,j} \approx 0 \quad (6)$$

where $H_{i,j} = H_z(i\Delta, j\Delta)$ and $k_i^{\text{rel}} = 1 - jS(i\Delta)$

(2.) Expand as in (1.), but use the discretized values of conductivity. The first term becomes:

$$-\frac{k_{i+1}^{\text{rel}} - k_{i-1}^{\text{rel}}}{2(k_i^{\text{rel}})^3 \Delta} \left(\frac{H_{i+1,j} - H_{i-1,j}}{2\Delta}\right) \quad (7)$$

(3.) Discretize the first-order equations (2) on one-half grid cell increments first, then combine them into the discrete Helmholtz equation:

$$\frac{1}{k_i^{\text{rel}} \Delta} \left(\frac{H_{i+1,j} - H_{i,j}}{k_{i+\frac{1}{2}}^{\text{rel}} \Delta} - \frac{H_{i,j} - H_{i-1,j}}{k_{i-\frac{1}{2}}^{\text{rel}} \Delta}\right) + \left(\frac{H_{i,j+1} - 2H_{i,j} + H_{i,j-1}}{\Delta^2}\right) + k_0^2 H_{i,j} \approx 0$$

which combines to give:

$$\left(\frac{H_{i+1,j} k_{i-\frac{1}{2}}^{\text{rel}} - H_{i,j} (k_{i+\frac{1}{2}}^{\text{rel}} + k_{i-\frac{1}{2}}^{\text{rel}}) + H_{i-1,j} k_{i+\frac{1}{2}}^{\text{rel}}}{k_i^{\text{rel}} k_{i+\frac{1}{2}}^{\text{rel}} k_{i-\frac{1}{2}}^{\text{rel}} \Delta^2}\right) + \left(\frac{H_{i,j+1} - 2H_{i,j} + H_{i,j-1}}{\Delta^2}\right) + k_0^2 H_{i,j} \approx 0 \quad (8)$$

Although each method is accurate to order $(k_0 \Delta)^2$ for functions which have Taylor series representations, the first method would appear to be the most exact, since it delays the discretization step to last, requiring the material parameters only at the central point of the three-point stencil. However, the third method incorporates three material parameters instead of just two (as with the other two methods), and will be shown to have the greatest accuracy.

III. ERROR ANALYSIS FOR FDFD DIFFERENCE EQUATIONS

To quantify the accuracy of the three discretization methods above, assume the continuous version of the 3.7 power conductivity profile given in (4) $S(x) = S_0 x^p$, where S_0 is constant with units m^{-p} . With this loss factor form, the corresponding exact plane wave solution to Equation (5) is readily computed using Equations (1) and (3):

$$H_z(x, y) = H_0 e^{-jk_{0x} x - jk_{0y} y - \frac{k_{0x} S_0}{1+p} x^{1+p}} \quad (9)$$

where the incident wave propagation vector is $\mathbf{k} = k_{0x} \hat{x} + k_{0y} \hat{y}$.

Applying this exact solution to each of the discretized Helmholtz equations (6-8) gives a measure of the discretization error. Care must be used in deriving the expansion of the exponential power functions for non-integer values of p . First, it must be understood that the small argument $\Delta \ll 1/k_0$ expansion relative to the central position (say x_i) in the stencil must take into account that the central position is itself an integer multiple of that small argument: $x_i = i\Delta$. Next, it is important to expand in terms which include both Δ and Δ^p . For *constant* S_0 , and non-integer p , the resulting lowest order error in approximating the differential Helmholtz equation (5) with the particular finite difference equation (6-8) (each divided by k_0^2 to normalize to distance) are respectively:

$$\frac{(k_{0x}^4 + k_{0y}^4)\Delta^2}{12k_0^2} + \frac{k_{0x}S_0\Delta^{p-1}}{k_0^2} \left(i^{p-1}p - \frac{(i-1)^{p+1} - 2i^{p+1} + (i+1)^{p+1}}{p+1} \right) \quad (10)$$

$$\frac{(k_{0x}^4 + k_{0y}^4)\Delta^2}{12k_0^2} + \frac{k_{0x}S_0\Delta^{p-1}}{k_0^2} \left(\frac{(i+1)^p - (i-1)^p}{2} - \frac{(i-1)^{p+1} - 2i^{p+1} + (i+1)^{p+1}}{p+1} \right) \quad (11)$$

$$\frac{(k_{0x}^4 + k_{0y}^4)\Delta^2}{12k_0^2} + \frac{k_{0x}S_0\Delta^{p-1}}{k_0^2} \left(\left(i + \frac{1}{2}\right)^p - \left(i - \frac{1}{2}\right)^p - \frac{(i-1)^{p+1} - 2i^{p+1} + (i+1)^{p+1}}{p+1} \right) \quad (12)$$

Note that in each of these expressions the leading term constitutes the standard FDFD error:

$$\frac{k_0^2\Delta^2(\sin^4\theta + \cos^4\theta)}{12} \quad (13)$$

for incidence angle θ and that the second terms are zero for $p = 1$ or 2 . For subsequent terms, the error is of higher order for $p > 3$. For non-integer p , no Taylor series exists for $H_z(x, y)$ as $x \rightarrow 0$, so the dependence on Δ can have order lower than $(k_0\Delta)^2$.

However, for predefined, frequency-independent loss of the form of Equation (4), $S_0 = \sigma_f/(N\Delta)^p\omega\epsilon_0 = \sigma_f\eta_0/k_0(N\Delta)^p$, a different expansion is needed. In this case, the dependence on Δ must be evaluated before the exponential function of Equation (9) is approximated. The first term of each error expression of Equations (10-12) remain the same, the lowest order dependence on Δ at the first layer ($n = 1$), for three cases now becomes:

$$\frac{U(\lambda/\Delta)}{2\pi} (f_1(p) + V f_2(p)) \quad (14)$$

$$f_1(p) = \begin{cases} p + \frac{2(1-2^p)}{1+p} & \text{method 1} \\ 2^{p-1} + \frac{2(1-2^p)}{1+p} & \text{method 2} \\ 2^{-p}(3^p - 1) + \frac{2(1-2^p)}{1+p} & \text{method 3} \end{cases}$$

$$f_2(p) = \begin{cases} p \left(1 - \frac{2^p}{(1+p)}\right) & \text{method 1} \\ 2^{p-1} \left(1 - \frac{2^p}{(1+p)}\right) & \text{method 2} \\ 2^{-p} \left(\frac{p(3^p-1)-2(1-2^p)^2}{1+p}\right) & \text{method 3} \end{cases}$$

$$U = \frac{\sigma_f\Delta\eta_0(\lambda/\Delta)}{2\pi N^p}$$

$$V = \frac{jU}{\cos\theta - jU}$$

with dependence only on the conductivity-grid product $\sigma_f\Delta$, the number of grid points per wavelength (λ/Δ) , the number of PML sublayers N , the and incidence angle θ . Note that the numerical error is not proportional to $\cos\theta$, indicating that although the numerical error is greatest for normal incidence, predicting the wide-angle absorption of a PML with given parameters is non-trivial.

Figure 1 shows the magnitude of the numerical error at normal incidence, for each of the three Equations (6-8), as a function of depth into the PML region, for the power parameter given in Equation (4), $p = 3.7$. The value of S_0 is chosen using $\sigma_f = .02/\Delta$, as in Equation (4) to provide an idealized two-way absorption of 223 dB, corresponding to an eight-layer loss of 2.67×10^{-6} . From Equation (9), $S_0 = -\ln(2.67 \times 10^{-6})(1+p)/(k_0(8\Delta)^{1+p})$. In this figure, $\Delta = \lambda/40$, implying $S_0 = 18511$ when $\lambda = 1$ m. The solid curve represents the first discretization method, Equation (6), the dotted curve represents Equation (7), and the dashed line corresponds to Equation (8). Clearly, the third method is superior. All curves indicate considerable error, over 100% at sublayer 5. These errors levels are due to difficulty in approximating the rapidly changing values of field within the PML layer. It is noted that if the conductivity variation is not included in the Helmholtz equation, the numerical errors are more than an order of magnitude greater than these presented in Figure 1.

For integer power $p \geq 2$ and constant S_0 , the numerical error for all three methods is $O(\Delta^2)$, but depends on the particular value of p . For the parabolic variation $p = 2$, the numerical error for each of the first two methods is given by:

$$\frac{(k_{0x}^4 + k_{0y}^4)\Delta^2}{12k_0^2} + j \frac{8k_{0x}^2 S_0 \Delta^2}{k_0^2} + \frac{2k_{0x} \Delta^3 S_0 i (k_{0x}^2 - j S_0)}{3k_0^2} \quad (15)$$

while for the third method, it is:

$$\frac{(k_{0x}^4 + k_{0y}^4)\Delta^2}{12k_0^2} + j \frac{5k_{0x}^2 S_0 \Delta^2}{k_0^2} + \frac{k_{0x} \Delta^3 S_0 i (4k_{0x}^2 - j S_0)}{6k_0^2} \quad (16)$$

Using the loss parameter for 120 dB two-way loss, PML(p, 8, 0.0001) as in [1], with the loss dependence specified in Equation (9): $S_0 = -\ln(10^{-3}) 3/(k_0(8\Delta)^3)$. When $\Delta = \lambda/40$ and $\lambda = 1$ m, $S_0 = 412$. In terms of conductivity,

$$\sigma_i = \sigma_f \left(\frac{2i-1}{16} \right)^2, \quad i = 1, \dots, 8 \quad (17a)$$

$$\sigma_{i+\frac{1}{2}} = \sigma_f \left(\frac{2i}{16} \right)^2, \quad i = 0, \dots, 8 \quad (17b)$$

with $\sigma_f = -3 \ln 10^{-6}/8\eta_0\Delta$

Figure 2 shows this numerical error dependence with sublayer number. The error for the third method is lowest for both the $p = 2$ and $p = 3.7$ cases.

IV. TWO DIMENSIONAL SIMULATION RESULTS FOR PML ABCS

The results of the previous section have been tested with a specific two dimensional FDFD scattering problem. The solution to the Helmholtz equation with a two dimensional point source:

$$(\nabla^2 + k_0^2)H_z = \delta(x - x_0, y - y_0) \quad (18)$$

(where δ is the 2D Dirac delta function) is given by the Hankel function $(j/4)H_0^{(2)}(k_0\sqrt{(x-x_0)^2 + (y-y_0)^2})$. To approximate the solution to Equation (18) a 3λ by 2λ space is discretized with an $121\Delta \times 81\Delta$ grid. The grid is surrounded by a PML layer of width 8Δ (that is, $N = 8$). The origin of the point source is $x_0 = -14\Delta, y_0 = -6\Delta$, where the origin (0,0) is assumed to be at row 41 and column 61 of the grid. The PML is terminated with a perfect magnetic conductor.

Discretization of Equation (18) together with either Equations (6) or (8) and lexicographical ordering of the unknowns leads to a system of linear equations $Ax = b$. The resulting matrix is sparse, complex, non-symmetric and non-hermitian. For this particular problem, A has dimension 13,192 but only 65,494 non-zero elements and is reasonably well conditioned; thus it was feasible to solve the system naively by Gaussian elimination. We note that a stable and fast approach effective at solving large systems of this type is given in [10].

Two separate experiments were performed, one in which we set the conductivity profile as in Equation (4) and then calculated solutions using Equations (6) and (8), respectively, and one in which we set the parabolic conductivity profiles as in Equation (17). All experiments were performed in Matlab in double precision floating point arithmetic. We used the Matlab command `besselh` to compute the analytic solution at our grid-points. For each of the two experiments, we compare our computed solutions with the analytic in terms of relative numerical error at each grid-point. That is, Figures 3 and 4 show values of $|H_{i,j} - H_0^{(2)}(i,j)|/|H_0^{(2)}(i,j)|$ corresponding to the two discretizations of the PML given in (6) and (8), respectively, where the profile is as given in Equation (4). Here, $H_0^{(2)}(i,j)$ is the value of the analytical solution at (x_i, y_j) and $H_{i,j}$ is our computed estimate of H_z at the same point. Figures 5 and 6 show the relative numerical errors when the profile is given by $p = 2$ and equations (6) and (8) are used for the PML, respectively.

Note that in both Figures 4 and 6 the symmetry of the numerical error patterns indicates almost no appreciable reflections from any of the grid edges. In addition, the pattern of numerical error is consistent with the standard angle dependent free-space FDFD sampling error, given by Equation (13); in other words, the PML is absorbing the wave as it should. However, the numerical error in Figures 3 and 5 is noticeably larger and the patterns seem to indicate reflection from the four side boundaries. These results are consistent with the asymptotic PML error results of the previous section.

V. CONCLUSIONS

The Perfectly Matched Layer can be efficiently implemented as the absorbing boundary condition for the Helmholtz equation FDFD method as long as the profiling of PML conductivities is incorporated into the computational stencil. A comparison of three methods of discretizing the Helmholtz equation with non-linearly increasing loss has shown that the numerical errors are smallest when the loss values are discretized first, at the first-order Maxwell's equations level, before combination into the Helmholtz equation. The error analysis is more difficult when the conductivity profile dependence has non-integer powers p of distance into the PML.

ACKNOWLEDGEMENTS

The authors are grateful to Ann Morgenthaler for many helpful discussions.

REFERENCES

- [1] J. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves", *J. Math. Phys.*, **114**, 185-200, (1994).
- [2] L. Zhao and A. Cangellaris, "GT-PML: Generalized Theory of Perfectly Matched Layers and Its Application to the Reflectiveness Truncation of Finite Difference Time-Domain Grids", *IEEE Transactions on Microwave Theory and Techniques*, **44**, 2555-2563, (1996).
- [3] S. Gedney, "An Anisotropic Perfectly Matched Layer-Absorbing Medium for the Truncation of FDTD Lattices", *IEEE Transactions on Antennas and Propagation*, **44**, 1630-1639, (1996).
- [4] Z. Sacks, D. Kingsland, R. Lee, and J. Lee, "A Perfectly Matched Anisotropic Absorber for Use as an Absorbing Boundary Condition", *IEEE Transactions on Antenna and Propagation*, **43**, 1460-1463, (1995).
- [5] E. Marengo and C. Rappaport and E. Miller, "Optimum PML ABC Conductivity Profile in FDFD", in review *IEEE Transactions on Magnetics*, **35**, 1506-1509, (1999).
- [6] C. Rappaport, "Interpreting and improving the PML absorbing boundary condition using anisotropic lossy mapping of space", *IEEE Transactions on Magnetics*, **32**, 968-974, (1996).
- [7] M. Kuzuoglu, R. Mittra, "Investigation of Nonplanar Perfectly Matched Absorbers for Finite-Element Mesh Truncation", *IEEE Transactions on Antennas and Propagation*, **45**, 474-486, (1997).
- [8] J. Tang, K. Paulsen, S. Haider, "Perfectly Matched Layer Mesh Terminations for Nodal-Based Finite-Element Methods in Electromagnetic Scattering", *IEEE Transactions on Antennas and Propagation*, **46**, 507-516, (1998).
- [9] J. Wu, R. Lee, and J. Lee, "The Use of Higher Order Edge-Based Finite Elements to Improve the Accuracy of the Anisotropic Perfectly Matched Layer", *URSI Symposium Digest*, 364, (1996).
- [10] M. Kilmer, E. Miller, and C. Rappaport, "Preconditioners for structured matrices arising in subsurface object detection", *SIAM Journal on Scientific Computation*, in review
- [11] C. Rappaport, "Perfectly Matched Absorbing Boundary Conditions Based on Anisotropic Lossy mapping of Space", *IEEE Microwave and Guided Wave Letters*, **5**, 90-92, (1995).
- [12] W. Chew and W. Weedon, "A 3-D Perfectly Matched Medium of Modified Maxwell's Equations with Stretched Coordinates", *Microwave and Optical Technology Letters*, **7**, 559-604, (1994).
- [13] S. Winton and C. Rappaport, "Profiling the Perfectly Matched Layer to Improve Large Angle Performance", *IEEE Transactions on Antennas and Propagation*, accepted for publication (September 2000).
- [14] J-P. Berenger, "Improved PML for the FDTD Solution of Wave Structure Interaction Problems", *IEEE Transactions on Antenna and Propagation*, **45**, 466-473, (1997).
- [15] Z. Wu and J. Fang, "High-Performance PML Algorithms", *IEEE Transactions on Microwave Theory and Techniques*, **6**, 2676-2683, (1995).
- [16] K. Yee, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media", *IEEE Transactions on Antennas and Propagation*, **AP-14**, 302-307, (1966).

FIGURE CAPTIONS

Figure 1. Numerical error in the discretized Helmholtz equation (residual value when exact H-field is used in difference equation, normalized to distance and initial value at space/PML boundary) for eight sublayer PML with loss factor $S_0 x^{3.7}$, $S_0 = 18511$ ($\sigma_f \Delta = 0.02$), for continuously varying conductivity (solid curve), conductivity discretized at the second-order equation level (dotted curve), and conductivity discretized at the first order equation level (dashed curve). $\Delta = \lambda/40$.

Figure 2. Numerical error in the discretized Helmholtz equation for eight sublayer PML with loss factor $S_0 x^2$, $S_0 = 412$ ($\sigma_f \Delta = 0.00687$), for continuously varying conductivity (solid curve), conductivity discretized at the second-order equation level (dotted curve coincident with solid curve), and conductivity discretized at the first order equation level (dashed curve). $\Delta = \lambda/40$.

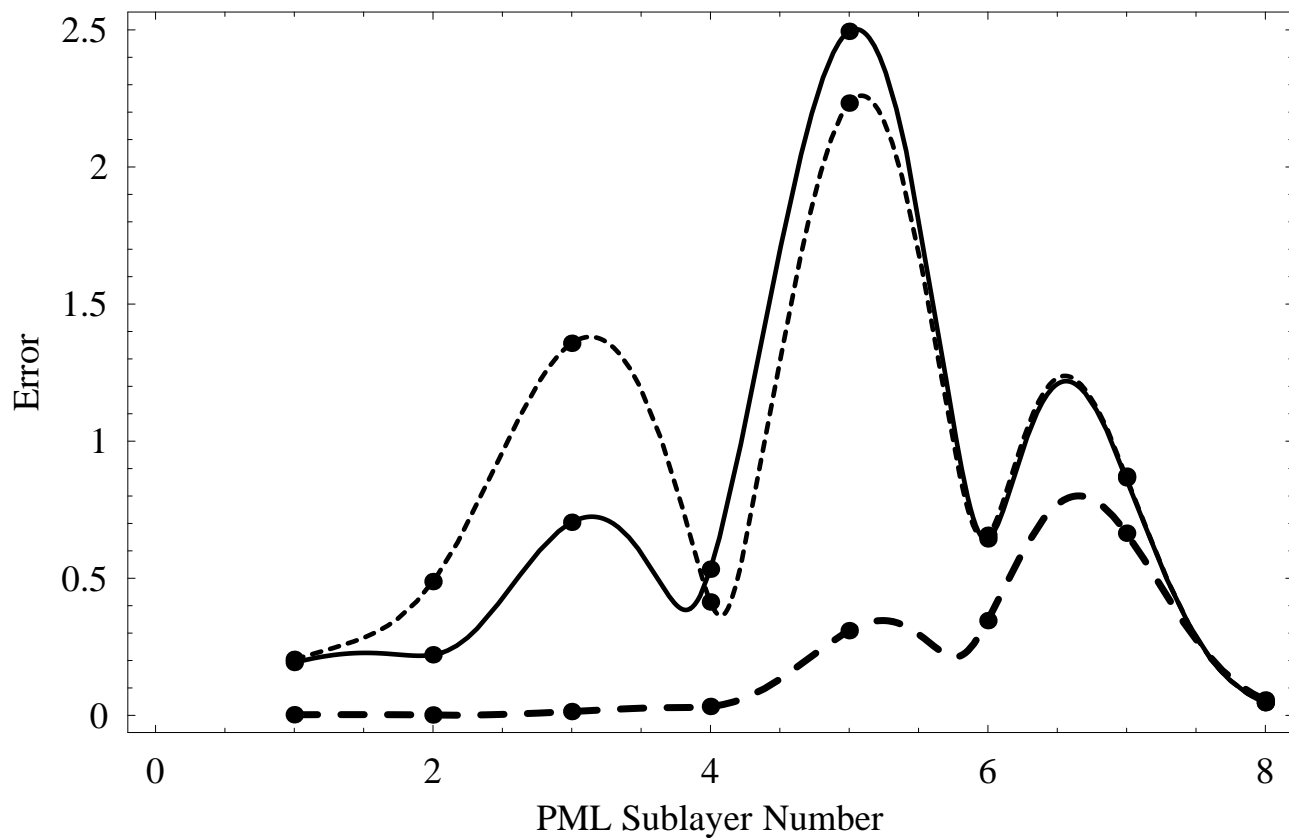
Figure 3: Relative error when FDFD computed solution has PML discretized as in Equation (6), wide-angle conductivity profile as in Equation (4).

Figure 4: Relative error when FDFD computed solution has PML discretized as in Equation (8), wide-angle conductivity profile as in Equation (4).

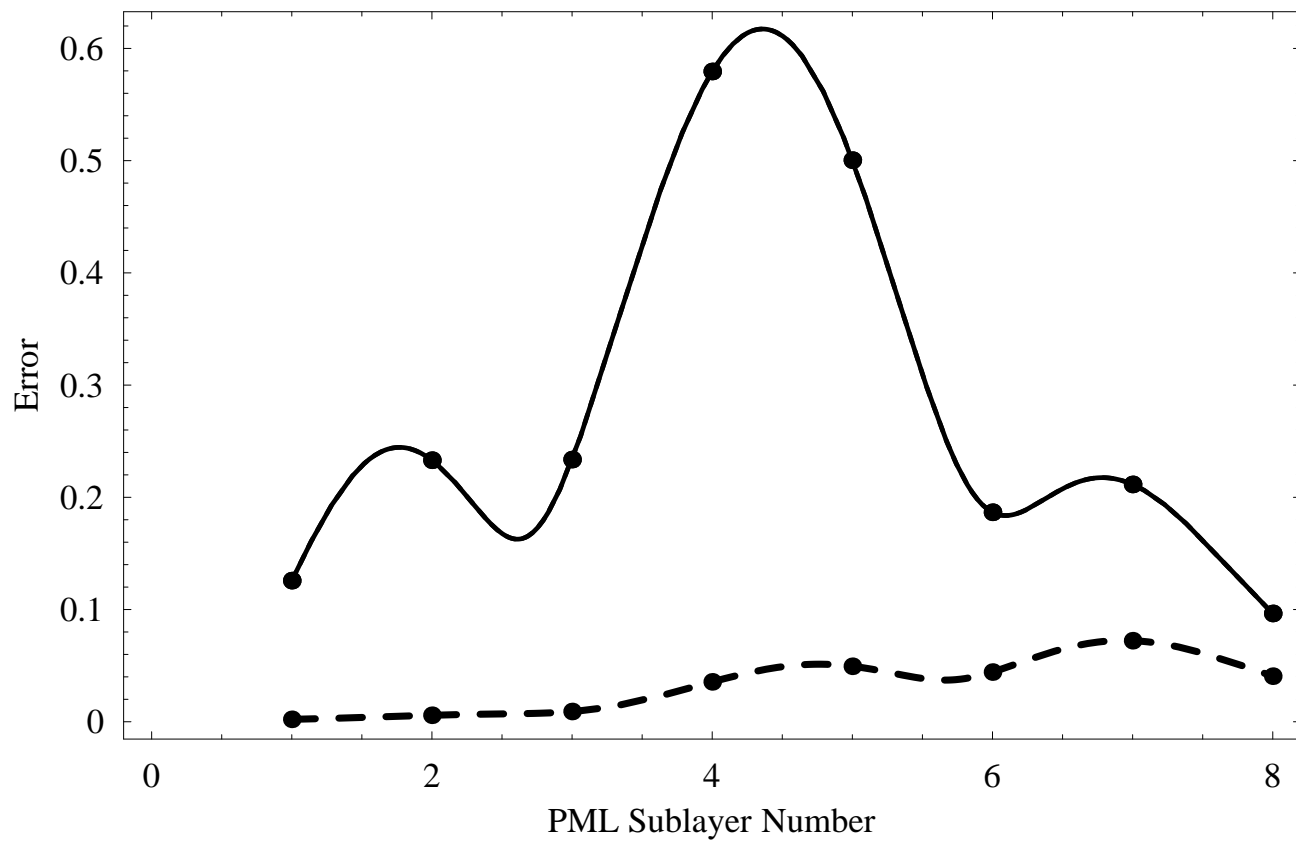
Figure 5: Relative error when FDFD computed solution has PML discretized as in Equation (6) with Berenger type parabolic conductivity profile in Equation (17).

Figure 6: Relative error when FDFD computed solution has PML discretized as in Equation (8) with Berenger type parabolic conductivity profile in Equation (17).

$p=3.7$, $D = .0251$, $\text{Loss} = 1.7 \times 10^{-12}$



p=2, D = .0251, Loss = 10^{-6}



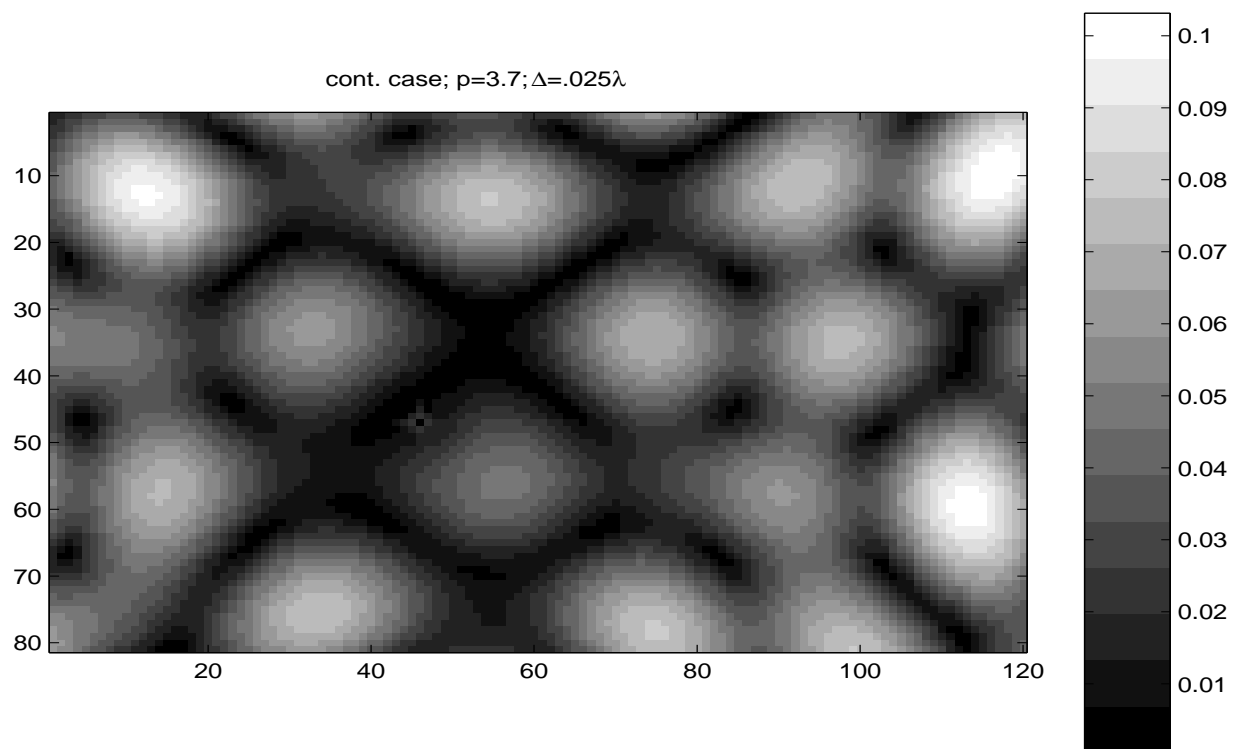


Figure 3:

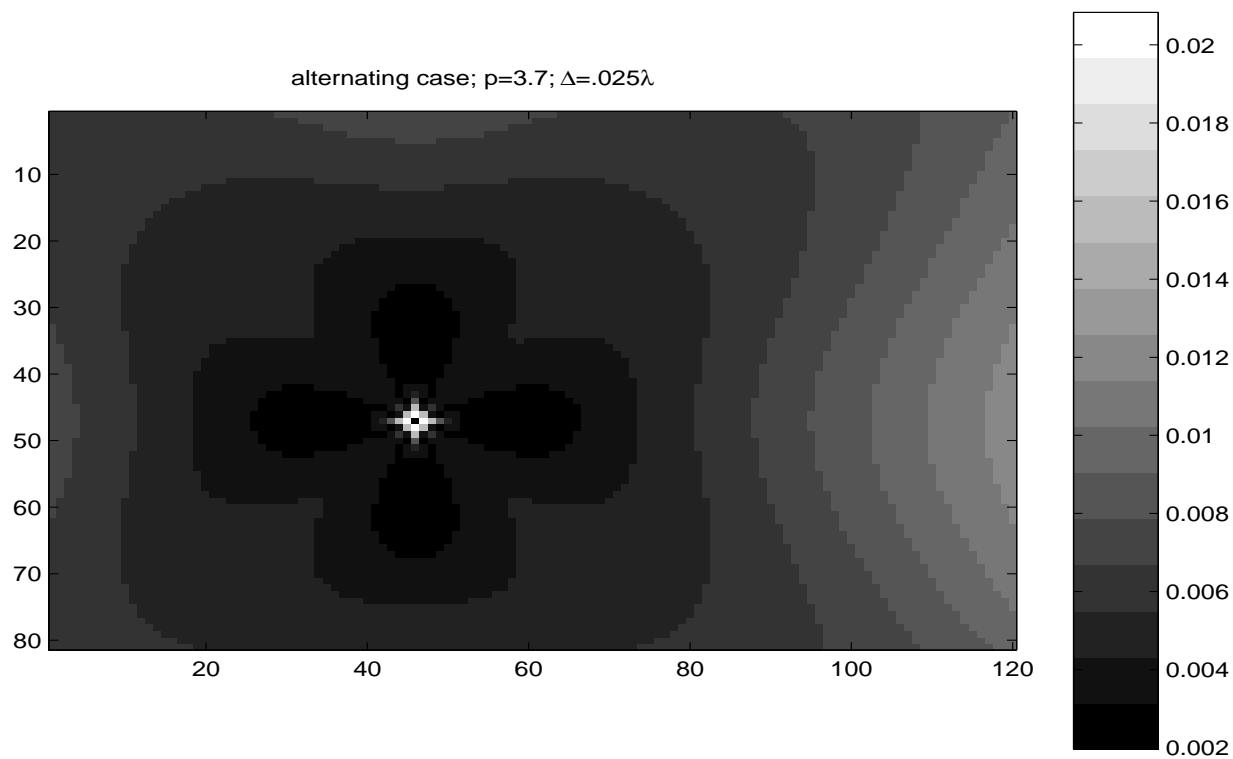


Figure 4:

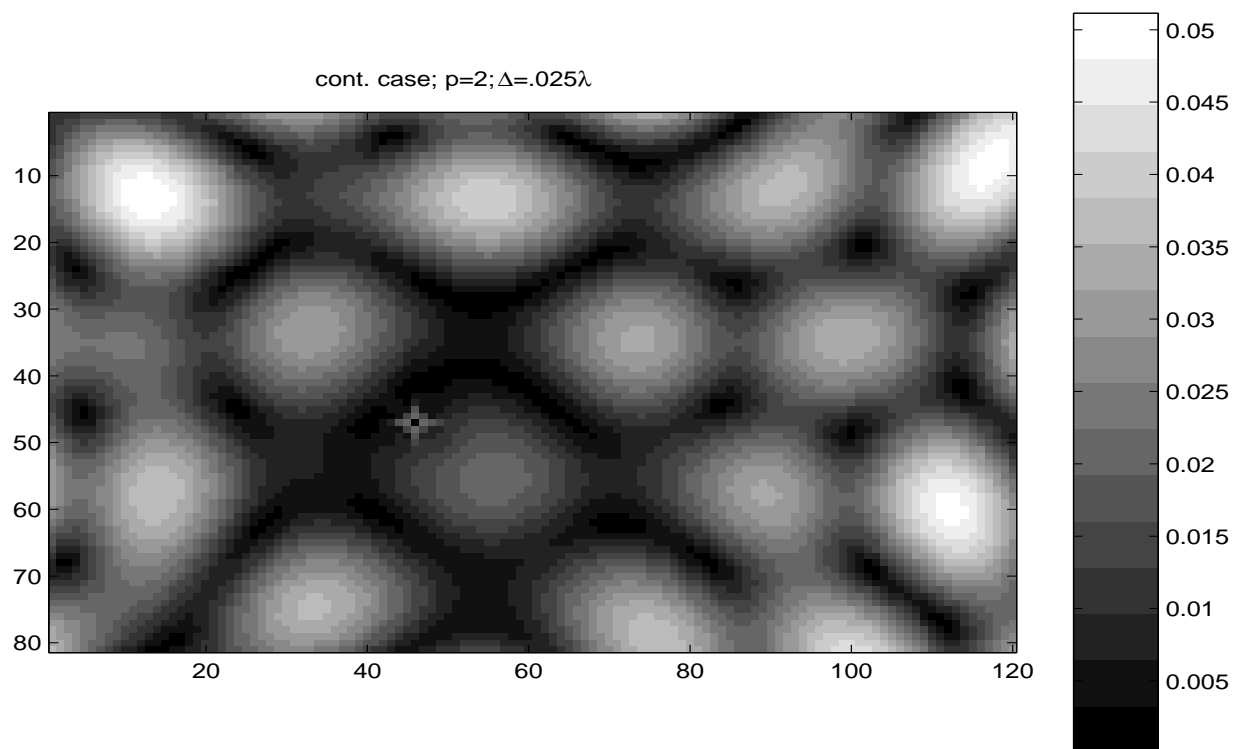


Figure 5:

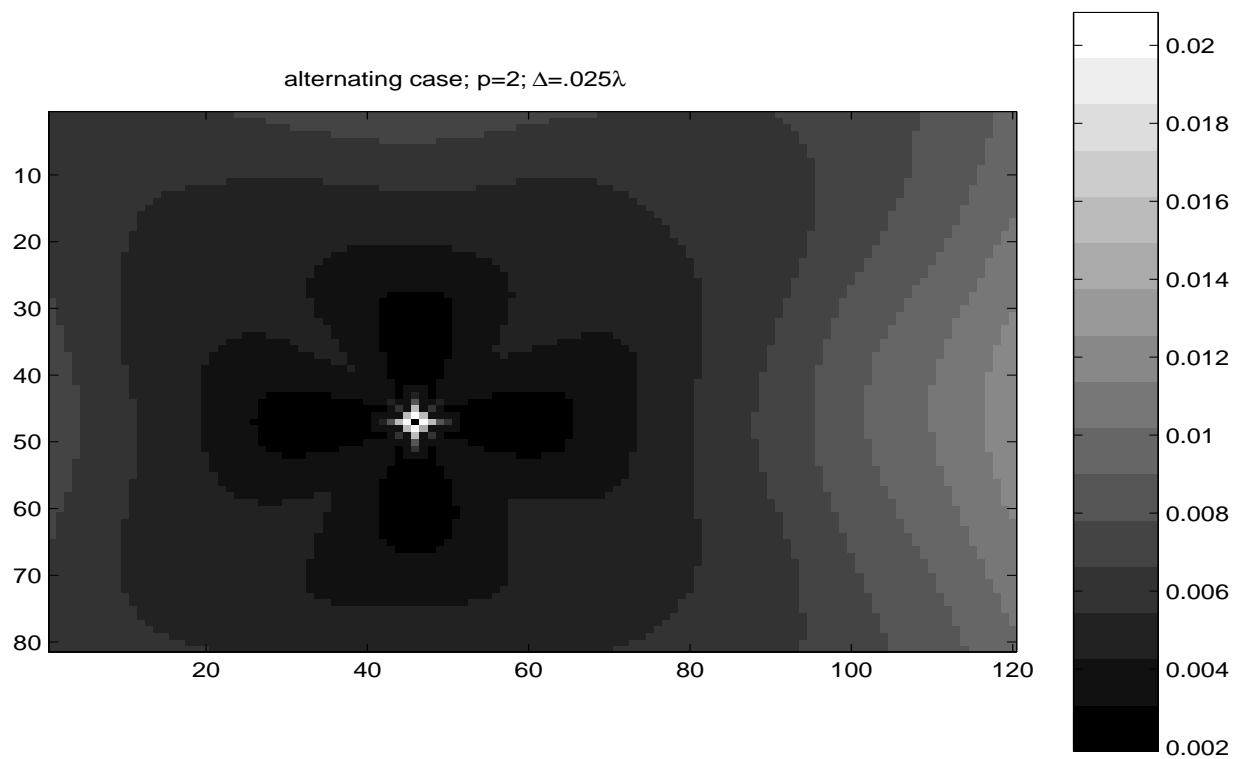


Figure 6:

Multi-mode subsurface sensing and imaging for land mine detection

Carey M. Rappaport, Michael Silevitch, Stephen McKnight, Charles DiMarzio,
Eric Miller, and Harold Raemer

Dept. Elect. and Comp. Eng., Northeastern University, Boston, MA 02115

ABSTRACT

Northeastern University is heading one of three teams involved in five year, \$5,000,000 Multidisciplinary University Research Initiative (MURI) programs in demining, supported by the Army Research Office. The Northeastern effort involves a team of six other universities and small businesses and integrates electromagnetic, acoustic, and optical sensing modalities. We are designing sensors to measure many of the physical characteristics of the ground with and without mines, and developing carefully-tailored signal processing algorithms that specifically take the sensor and the physics into account.

Keywords: GPR, mine detection, focused array, reflector antenna.

INTRODUCTION

The best mine detectors, used in both US Army countermine operations and by humanitarian demining efforts that can afford them, are based on fifty-year-old technology comparable to metal detectors used by beach-combers. There is tremendous potential for improving the reliability of mine detection using state-of-the-art sensors, computer modeling, and signal processing, combined with the phenomenal recent advances in computer power. The overall goal of humanitarian demining is to use these tools to help people around the world clear hazardous land at a cost that even developing countries can afford.

The difficulty with finding buried mines is that there is no single type of “target”, nor surrounding background. In addition, the soil surrounding the mine often strongly absorbs electromagnetic radiation and may also have buried rocks, moisture pockets, tree roots, and bits of metal scrap (especially in a former battlefield). A rough soil surface may also randomly scatter much of the transmitted sensing signal and can be a significant source of detection noise. Conventional radar is relatively good at detecting metallic objects, both in the air and underground, but is much less effective at detecting nonmetallic targets, such as stealth aircraft or plastic mines. In addition, objects too close to the ground surface pose a problem in discrimination of the target from the ground surface itself. Thus, large metallic 30cm diameter anti-tank mines are relatively easy to find, but the 8cm anti-personnel mines with as little as half a gram total metal content, buried just 2 to 10cm under ground are quite difficult to detect. The need to find all sorts of targets, at depths varying from just below the surface to ten centimeters deep, in all terrain, from desert to rice paddy to rain forest, make mine detection a challenging problem.

Figure 1 illustrates four typical mine positions and approaches using conventional sensing modalities to detect them. Although there are newer and more exotic detection technologies being considered for mine detection, cost considerations preclude widespread acceptance for humanitarian demining. One promising novel detection modality which may become economically feasible is Nuclear Quadrupole Resonance sensing of the chemical nature of bulk explosives [1]

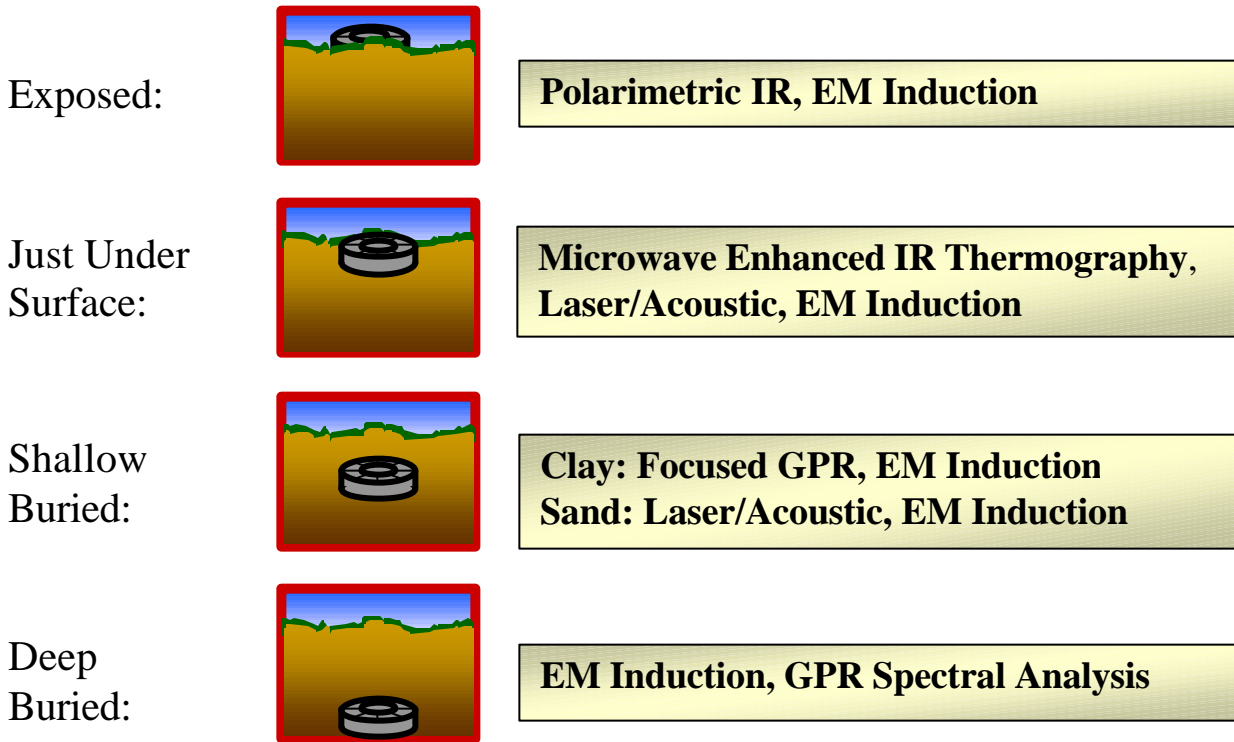


Figure 1.: Typical mine burial scenarios and appropriate detection modalities.

The sensing modalities alone are not enough to solve the mine detection problem. In addition, there must exist powerful physics-based and statistical signal processing algorithms to reconstruct the target features based on observed measurements. These algorithms in turn require computational models that allow for hypothesis testing, iterative or recursive estimation, and inverse scattering. Finally, since it is unlikely that a single detection modality will find every mine, it is important to combine the information from several sensing modalities with appropriate fusion techniques. The Northeastern University MURI strategic plan, showing the various overlapping efforts is shown in Figure 2.

The four main sensing modalities are ground penetrating radar (GPR), electromagnetic induction (EMI), infrared (IR), and acoustic detection. The processing approaches include parameter-based inverse scattering, recursive estimation, diffraction tomography, batch statistical detection, and

matched filtering. The stars with letter correspond to individual projects, abbreviated on the right. In the following sections, we will discuss a selection of these topics.

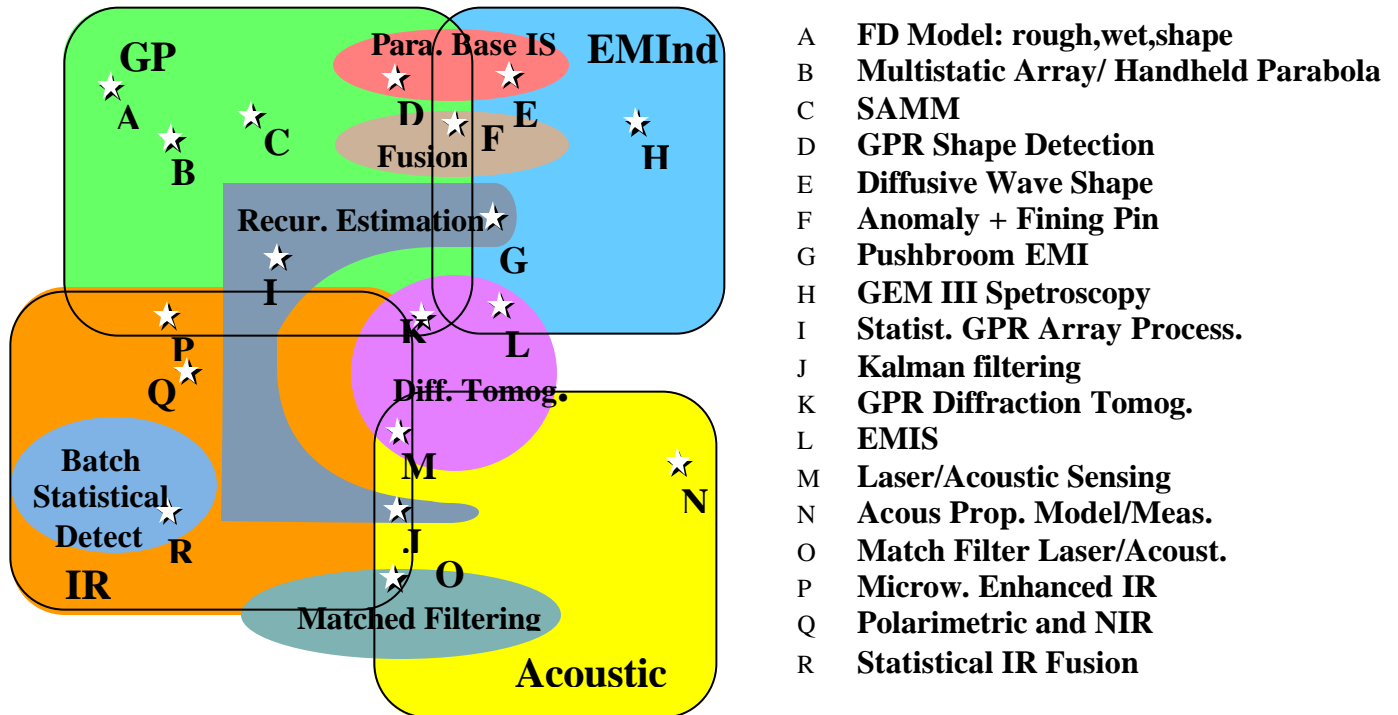


Figure 2: Northeastern University MURI project integration.

ELECTROMAGNETIC INDUCTION

Electromagnetic induction sensing of metallic objects is the standard subsurface detection modality. Metal detectors have been used since before World War II, and are quite effective in alarming on metal case mines and unexploded ordnance (UXO), such as shells and bombs (which are usually metallic). This technology can also sense the small metal firing pin almost always present in plastic and wooden mines [2]. However, when the EMI sensor sensitivity is set to detect such small metal targets, all sorts of man-made metallic and naturally occurring conductive clutter--such as nails, spent bullets, and bits of rust--is detected, and the false alarm rate becomes unacceptable. One approach to better discriminate targets from clutter and reduce the false alarm rate is to use wideband electromagnetic induction spectroscopy (EMIS) [3]. By examining the characteristics of the metal scattered signal from 30Hz to 24 kHz, and using a physical model of the target induced dipole moment, unique features of the mines in question can be distinguished [4].

GROUND PENETRATING RADAR

Ground penetrating radar is effective at identifying dielectric contrasts in the surrounding environment. If the dielectric contrast is large enough, the size of the mine target large enough, and the distance from the ground surface deep enough, it is possible to uniquely detect mines. However, distinguishing a nonmetallic mine from a dry or sandy soil background is quite challenging, because the dielectric constants of TNT and dry soil are practically the same. Also, for the case of antipersonnel mines, the target volume is small and it may only be buried a few centimeters below the surface. Removing the strong ground surface reflection clutter without overly suppressing the mine signal is a topic of active research [5].

Sophisticated electromagnetic modeling illustrates some of the problematic issues. Figure 3 shows two images created using the three-dimensional finite difference frequency domain (FDFD) method [6, 7]. Both images show the total E_x field at two planes above, on and below a random rough ground surface due to a Hertzian x-directed dipole above the ground. The left image shows the response when a rectangular nonmetallic mine target is buried 5cm down, while the right image results from a circular cylindrical mine with the same volume buried at the same place. Only by careful observation can the differences be observed. Clearly, sharp corners and smooth surfaces cannot be readily detected. In practice, direct comparison between the field configurations for known target cases is not possible. Even if the ground roughness could be measured and stored, computationally modeling the effects of the clutter cannot be accomplished in real time. The images of Figure 3 require approximately 50 hours of CPU time on a multiprocessor Compaq Alpha computer.

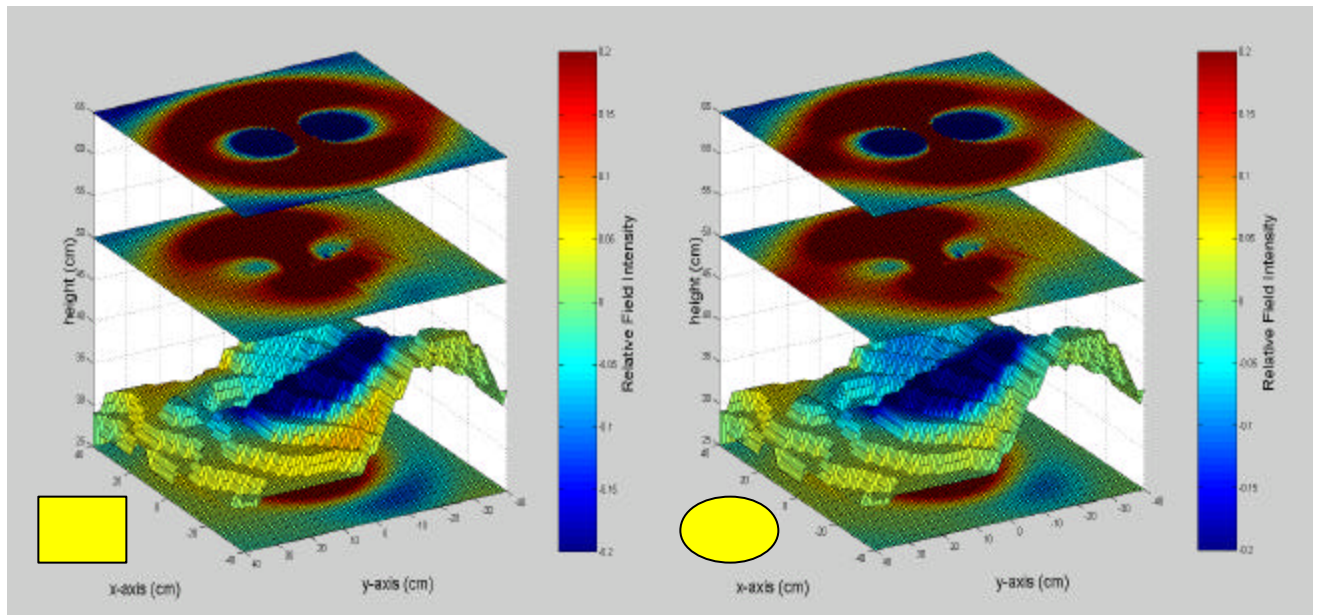


Figure 3: 3-D FDFD computational model of total E_x due to a dipole radiating at 960 MHz for a buried rectangular box target (left) and a circular cylindrical target (right) in clay loam soil with rough surface.

One of our major research projects is to improve the detection capability of ground penetrating radar by reducing clutter by simultaneously using multiple and receivers to focus waves on points under the ground surface. By “looking” at the mine from different angles at the same time, it is possible to emphasize the features of the mine, and minimize the clutter from the rough ground surface and uneven intervening soil. In addition, if the transmitted wave is planar and forward directed, the reflection from the ground surface tends to continue forward, away from the sensor. As such, it does not have as great an effect confusing the target signal. This effect has been confirmed both with computational models [8, 9], and in hardware, with an offset parabolic transmitting antenna [10]. Figure 4 depicts hardware fabricated by Lawrence Livermore National Labs. Test results for a nonmetallic mine buried in dry sand 1 inch under rough surface indicate an easily detectable anomaly with the size and approximate shape of the mine.

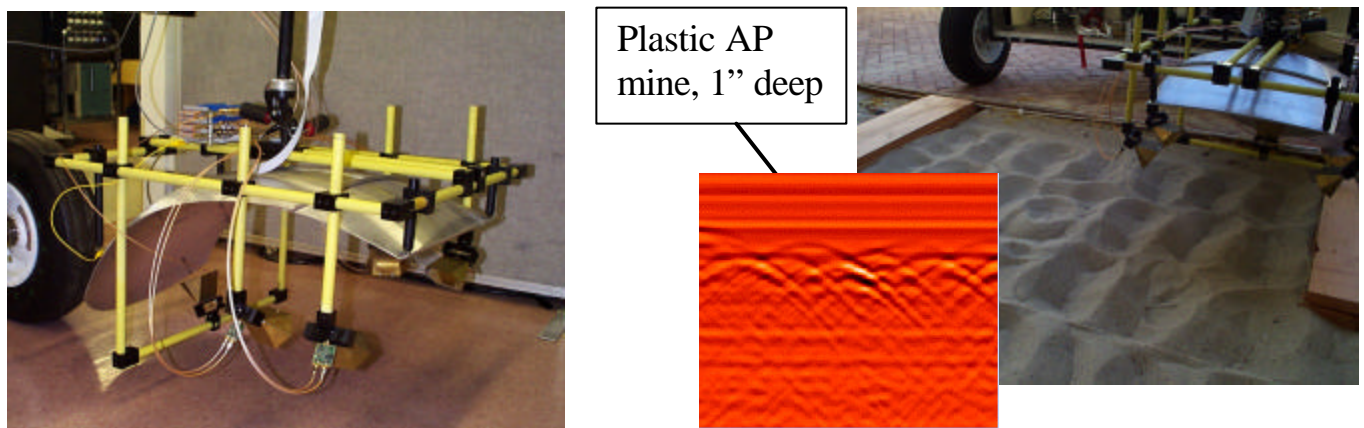


Figure 4: Offset parabolic reflector transmitter and multiple element receiver GPR handheld mine detector, with processed detection results for nonmetallic mine buried in rough sand

ACOUSTIC SENSING

Another exciting research area is the novel use of sound waves to detect acoustic echoes from buried mines. In deserts where the sandy soil is usually quite dry, non-metallic mines look electromagnetically very much like the background, and so are hard to pick out. To solve this problem, Northeastern University’s Demining team is using sound rather than radar or electromagnetic induction to find solid mine-like objects embedded in loose soil. The acoustic mismatch between granular soil and any metallic or non-metallic solid object is very high, giving a strong target-to-medium contrast.

One major problem with ground penetrating sonar is coupling through the ground surface. While medical ultrasound devices can be pressed firmly onto a patient's skin, direct ground contact in a minefield is not feasible. We have developed an alternative: using a laser beam to quickly heat the ground, causing rapid expansion, which in turn generates a sonic shock wave [11]. The scattered signals are detected either with a microphone placed very close to the ground, or interferometrically, using another laser beam to detect the ground motion.

This geometry, as well as sample detection results is shown in Figure 5. The mine simulant target is a bisected hockey puck, positioned slightly inclined to the ground surface. The surrounding material in this experiment was dry sand. A CO₂ TEA laser generated 150 mJ pulses several times a second, resulting in a string of acoustic pulses generated at the sand surface. A microphone was fixed relative to the laser, and the two were moved in a serpentine pattern relative to the ground. The measured signals were first filtered to remove as much of the ground surface reflection as possible, then the recorded signals were displayed as a function of position on the ground (bottom figure) [12]. The shape of the target is clearly distinguishable. It is also possible, by observing the arrival time of the target signal at the particular point, to determine the depth from the ground surface to the top of the target.

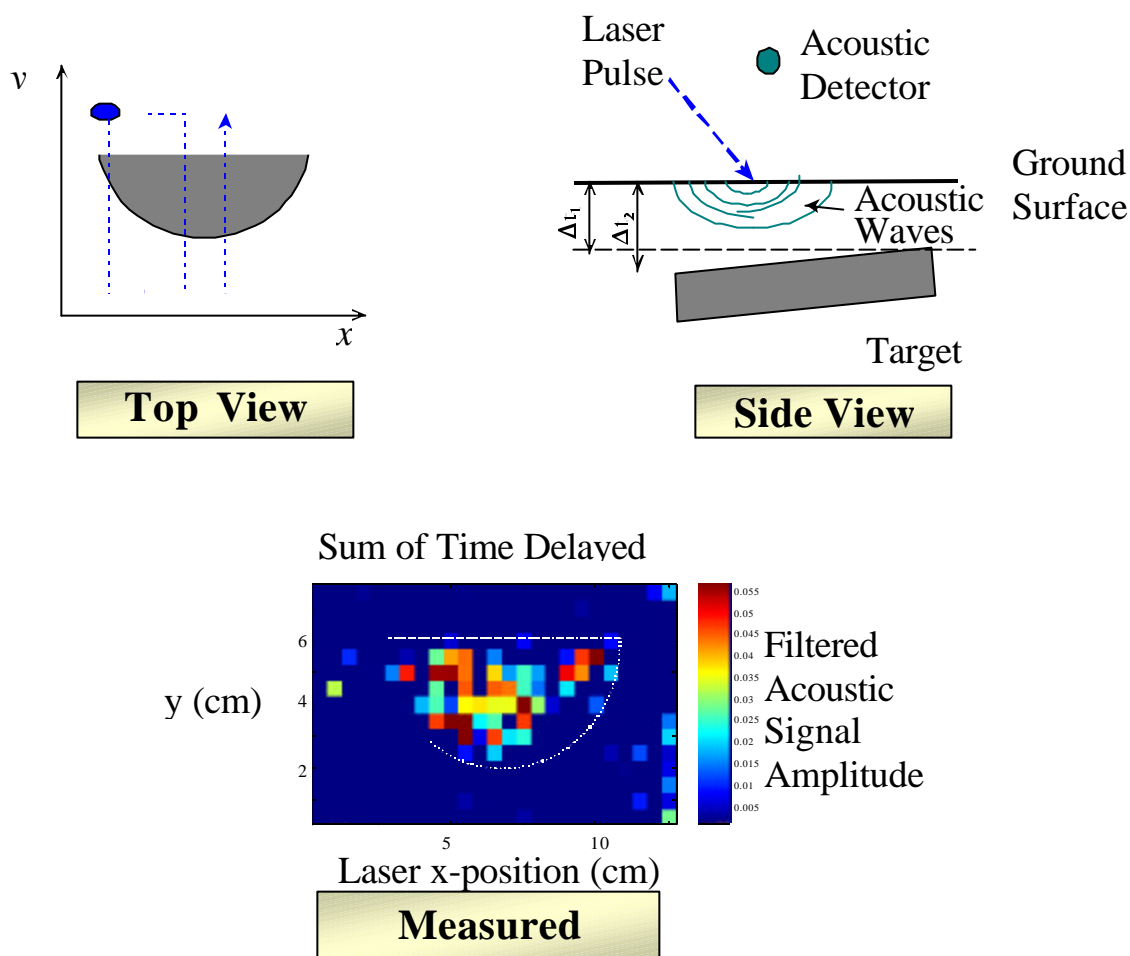


Figure 5: Laser induced acoustic sensing, geometry (top), and measured results (bottom).

INFRARED SENSING

Infrared detection is another sensing modality that has been used to find mines. Both polarimetric effects, which can distinguish exposed man-made objects [13], and thermal effects can be exploited. By measuring the subtle temperature changes from solar heating between soil with and without a mine, it is sometimes possible to find shallowly buried mines. Northeastern University's team has developed a novel means of using IR thermography by simulating the sun with microwave heating [14]. Our heater points down, toward the ground, deeply heating a small patch of soil. Meanwhile, we take IR pictures of the soil patch and look for anomalous heating patterns. Instead of being limited to sunrise and sunset, when the solar heating changes are most noticeable, our system can be used all day. Figure 6 shows this concept of heating the ground with microwaves. The right image indicates the temperature differences on the sand surface due to non-uniform heating of the soil above the targets. The shapes of the targets are clearly visible.

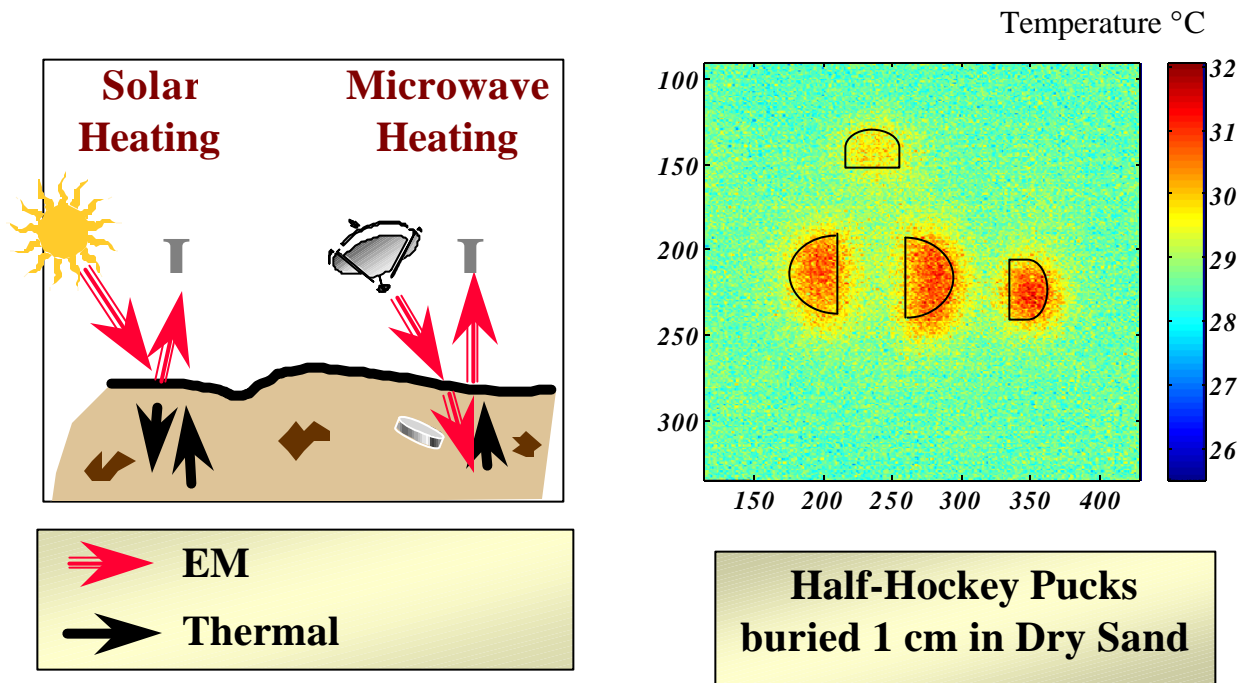


Figure 6: Microwave enhanced IR thermography concept (left) and measured results (right).

It is also possible to use multiple microwave frequencies for heating the ground. Since the surface heating pattern is relatively insensitive to frequency, while the heating at depth depends strongly on frequency, taking the difference from two heatings will better show the presence of buried anomalies [14].

CONCLUSIONS

In designing each of our sensors, our emphasis is to use as much of the available information as possible: use many different probing signals, with differing frequencies; use the widest possible view, with many inexpensive sensors coupled together; use the information gained from mine-free terrain and from confirmed terrain with mines to give improved detection confidence; use the fact that mines are stationary, while the detectors can move, and above all, keep the advanced technology simple in principle, so that it can detect mines in dirty, or frozen, or hard-to-traverse terrain.

ACKNOWLEDGEMENT

This work was supported by The Army Research Office Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013.

REFERENCES

- (1) Hibbs, A., *et. al.* "Field tests results of a nuclear quaterupole resonance landmine detection system," *SPIE Aerosense Conference*, Orlando, FL, April 2000.
- (2) A. Witten, I.J. Won, and S. Norton, "Subsurface imaging with broadband electromagnetic induction," *Inverse problems*, Vol. 13, pp. 1621-1639, 1997.
- (3) I.J. Won, D. Keiswetter, and E. Novikova, "Electromagnetic induction spectroscopy," *Jour. Environmental and Engineering Geophysics*, Vol. 3, Issue 1, pp. 27-40, 1998
- (4) E. Miller, W. C. Karl and S. Norton, "On the detection of buried mines from multifrequency, inductive measurements," *SPIE AeroSense Symposium, Detection Technologies for Mines and Minelike Targets*, Orlando, Florida, April 1997.
- (5) El-Shenawee, M. and Rappaport, C., "Quantifying the Effects of Different Rough Surface Statistics for Mine Detection Using the FDTD Technique," *SPIE Aerosense Conference*, Orlando, FL, April 2000.
- (6) C. Rappaport, M. Kilmer, and E. Miller, "Accuracy Considerations in Using the PML ABC with FDFD Helmholtz Equation Computation", (Invited Paper) accepted for publication in *International Journal of Numerical Modeling*, 1999.
- (7) C. Rappaport and S. Winton, "Using the PML ABC for Air/Soil Wave Interaction Modeling in the Time and Frequency Domains," accepted for publication in *IEEE Transactions on Antennas and Propagation*, 1999.
- (8) C. Rappaport, S. Wu, and S. Winton, "FDTD Wave Propagation Modeling in Dispersive Soil Using a Single Pole Conductivity Model," *IEEE Transactions on Magnetics*, vol.35, May 1999, pp. 1542-1545.
- (9) Jin, D. and Rappaport, C., "FDTD Simulation of a Parabolic Antenna GPR transmitter," accepted for publication in *Progress in Electromagnetics Research Symposium*, July 2000.
- (10) Rappaport, C., Azevedo, S., Rosenbury, T., Gough, J, and Jin, D., "Handheld Forward-Looking Focused Array Mine Detection with Plane Wave Excitation," *SPIE Aerosense Conference*, Orlando, FL, April 2000.

- (11) S. McKnight, C. DiMarzio, and D. Hogenboom and G. Sauermann, "Laser-induced acoustic detection of buried objects," *SPIE's 12th Annual International Symposium on Aerospace /Defense Sensing, Simulation, and Controls*, April 13-17 1998, pp .841-847.
- (12) S. McKnight, W. Li, and C. DiMarzio, "Imaging of buried objects using laser-induced acoustic detection," *SPIE's 13th Annual International Symposium on Aerospace /Defense Sensing, Simulation, and Controls*, April 13-17 1999, pp .231-238.
- (13) H. Scott, S. Jones and Iannarilli, F.I. 1999. "Imaging Infrared Polarimetry: Initial Results and Potential in Detection of Scatterable Mines and Surface Disturbances." *Proceedings of SPIE Vol. 3710, Detection Remediation Technologies for Mines and Minelike Targets IV*. 5-9 April 1999, Orlando FL
- (14) Charles A. DiMarzio, Li Wen, Carey M. Rappaport, Gerhard Sauermann, Herman E. Scott, "Microwave--Enhanced Infrared Thermography," *SPIE AeroSense Conference*, Orlando, FL, April 1999.
- (15) Oktar, T., Rappaport, C., and DiMarzio, C. "The Effects of Surface Roughness on Microwave Heating of Soil for Detection of Buried Land Mines," accepted for publication in *SPIE Aerosense Conference*, Orlando, FL, April 2000.

Multi-mode subsurface sensing and imaging for land mine detection

Carey M. Rappaport, Michael Silevitch, Stephen McKnight, Charles DiMarzio,
Eric Miller, and Harold Raemer

Dept. Elect. and Comp. Eng., Northeastern University, Boston, MA 02115

ABSTRACT

Northeastern University is heading one of three teams involved in five year, \$5,000,000 Multidisciplinary University Research Initiative (MURI) programs in demining, supported by the Army Research Office. The Northeastern effort involves a team of six other universities and small businesses and integrates electromagnetic, acoustic, and optical sensing modalities. We are designing sensors to measure many of the physical characteristics of the ground with and without mines, and developing carefully-tailored signal processing algorithms that specifically take the sensor and the physics into account.

Keywords: GPR, mine detection, focused array, reflector antenna.

INTRODUCTION

The best mine detectors, used in both US Army countermine operations and by humanitarian demining efforts that can afford them, are based on fifty-year-old technology comparable to metal detectors used by beach-combers. There is tremendous potential for improving the reliability of mine detection using state-of-the-art sensors, computer modeling, and signal processing, combined with the phenomenal recent advances in computer power. The overall goal of humanitarian demining is to use these tools to help people around the world clear hazardous land at a cost that even developing countries can afford.

The difficulty with finding buried mines is that there is no single type of “target”, nor surrounding background. In addition, the soil surrounding the mine often strongly absorbs electromagnetic radiation and may also have buried rocks, moisture pockets, tree roots, and bits of metal scrap (especially in a former battlefield). A rough soil surface may also randomly scatter much of the transmitted sensing signal and can be a significant source of detection noise. Conventional radar is relatively good at detecting metallic objects, both in the air and underground, but is much less effective at detecting nonmetallic targets, such as stealth aircraft or plastic mines. In addition, objects too close to the ground surface pose a problem in discrimination of the target from the ground surface itself. Thus, large metallic 30cm diameter anti-tank mines are relatively easy to find, but the 8cm anti-personnel mines with as little as half a gram total metal content, buried just 2 to 10cm under ground are quite difficult to detect. The need to find all sorts of targets, at depths varying from just below the surface to ten centimeters deep, in all terrain, from desert to rice paddy to rain forest, make mine detection a challenging problem.

Figure 1 illustrates four typical mine positions and approaches using conventional sensing modalities to detect them. Although there are newer and more exotic detection technologies being considered for mine detection, cost considerations preclude widespread acceptance for humanitarian demining. One promising novel detection modality which may become economically feasible is Nuclear Quadrupole Resonance sensing of the chemical nature of bulk explosives [1]

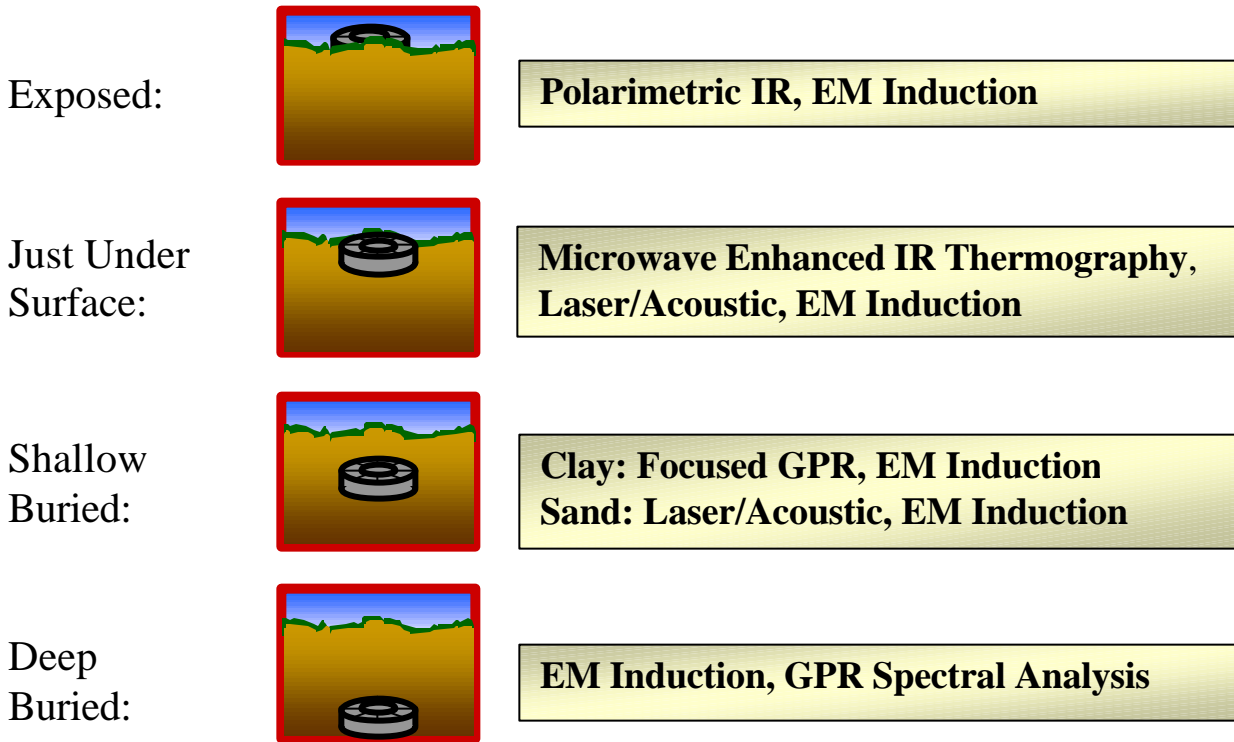


Figure 1.: Typical mine burial scenarios and appropriate detection modalities.

The sensing modalities alone are not enough to solve the mine detection problem. In addition, there must exist powerful physics-based and statistical signal processing algorithms to reconstruct the target features based on observed measurements. These algorithms in turn require computational models that allow for hypothesis testing, iterative or recursive estimation, and inverse scattering. Finally, since it is unlikely that a single detection modality will find every mine, it is important to combine the information from several sensing modalities with appropriate fusion techniques. The Northeastern University MURI strategic plan, showing the various overlapping efforts is shown in Figure 2.

The four main sensing modalities are ground penetrating radar (GPR), electromagnetic induction (EMI), infrared (IR), and acoustic detection. The processing approaches include parameter-based inverse scattering, recursive estimation, diffraction tomography, batch statistical detection, and

matched filtering. The stars with letter correspond to individual projects, abbreviated on the right. In the following sections, we will discuss a selection of these topics.

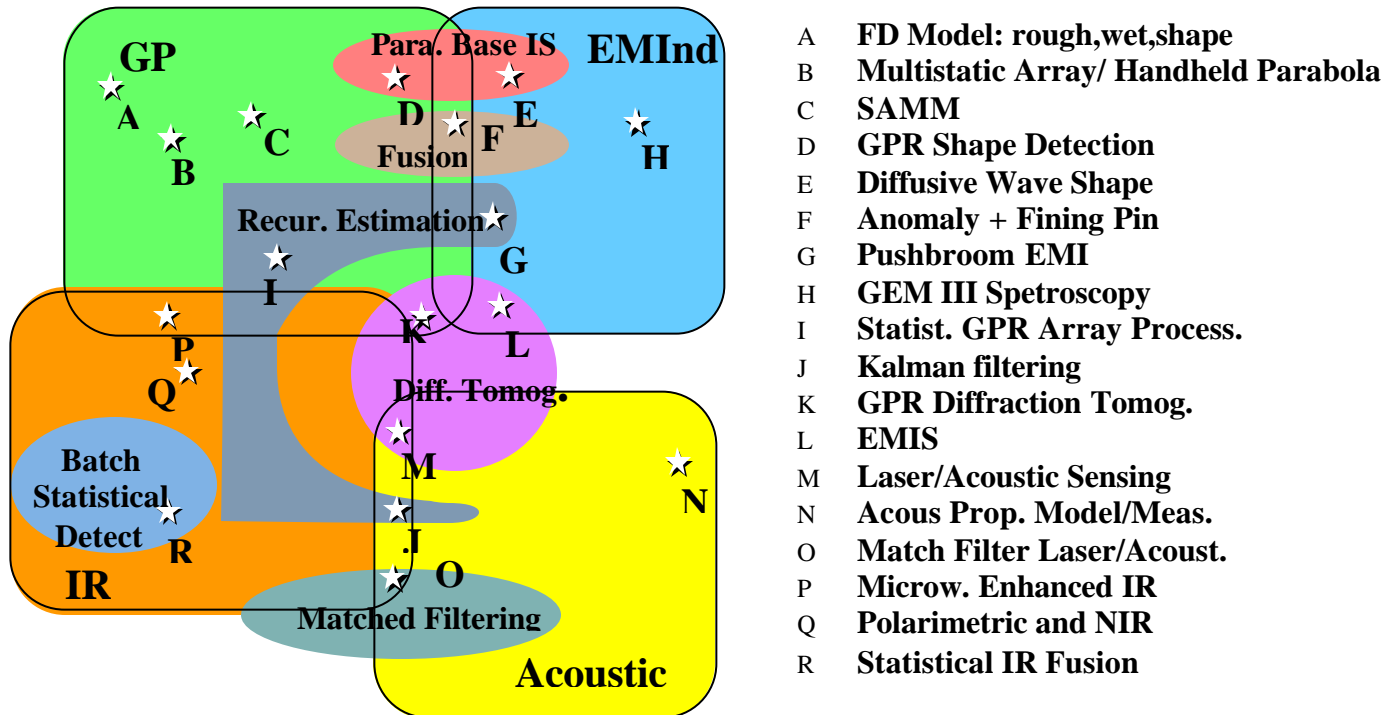


Figure 2: Northeastern University MURI project integration.

ELECTROMAGNETIC INDUCTION

Electromagnetic induction sensing of metallic objects is the standard subsurface detection modality. Metal detectors have been used since before World War II, and are quite effective in alarming on metal case mines and unexploded ordnance (UXO), such as shells and bombs (which are usually metallic). This technology can also sense the small metal firing pin almost always present in plastic and wooden mines [2]. However, when the EMI sensor sensitivity is set to detect such small metal targets, all sorts of man-made metallic and naturally occurring conductive clutter--such as nails, spent bullets, and bits of rust--is detected, and the false alarm rate becomes unacceptable. One approach to better discriminate targets from clutter and reduce the false alarm rate is to use wideband electromagnetic induction spectroscopy (EMIS) [3]. By examining the characteristics of the metal scattered signal from 30Hz to 24 kHz, and using a physical model of the target induced dipole moment, unique features of the mines in question can be distinguished [4].

GROUND PENETRATING RADAR

Ground penetrating radar is effective at identifying dielectric contrasts in the surrounding environment. If the dielectric contrast is large enough, the size of the mine target large enough, and the distance from the ground surface deep enough, it is possible to uniquely detect mines. However, distinguishing a nonmetallic mine from a dry or sandy soil background is quite challenging, because the dielectric constants of TNT and dry soil are practically the same. Also, for the case of antipersonnel mines, the target volume is small and it may only be buried a few centimeters below the surface. Removing the strong ground surface reflection clutter without overly suppressing the mine signal is a topic of active research [5].

Sophisticated electromagnetic modeling illustrates some of the problematic issues. Figure 3 shows two images created using the three-dimensional finite difference frequency domain (FDFD) method [6, 7]. Both images show the total E_x field at two planes above, on and below a random rough ground surface due to a Hertzian x-directed dipole above the ground. The left image shows the response when a rectangular nonmetallic mine target is buried 5cm down, while the right image results from a circular cylindrical mine with the same volume buried at the same place. Only by careful observation can the differences be observed. Clearly, sharp corners and smooth surfaces cannot be readily detected. In practice, direct comparison between the field configurations for known target cases is not possible. Even if the ground roughness could be measured and stored, computationally modeling the effects of the clutter cannot be accomplished in real time. The images of Figure 3 require approximately 50 hours of CPU time on a multiprocessor Compaq Alpha computer.

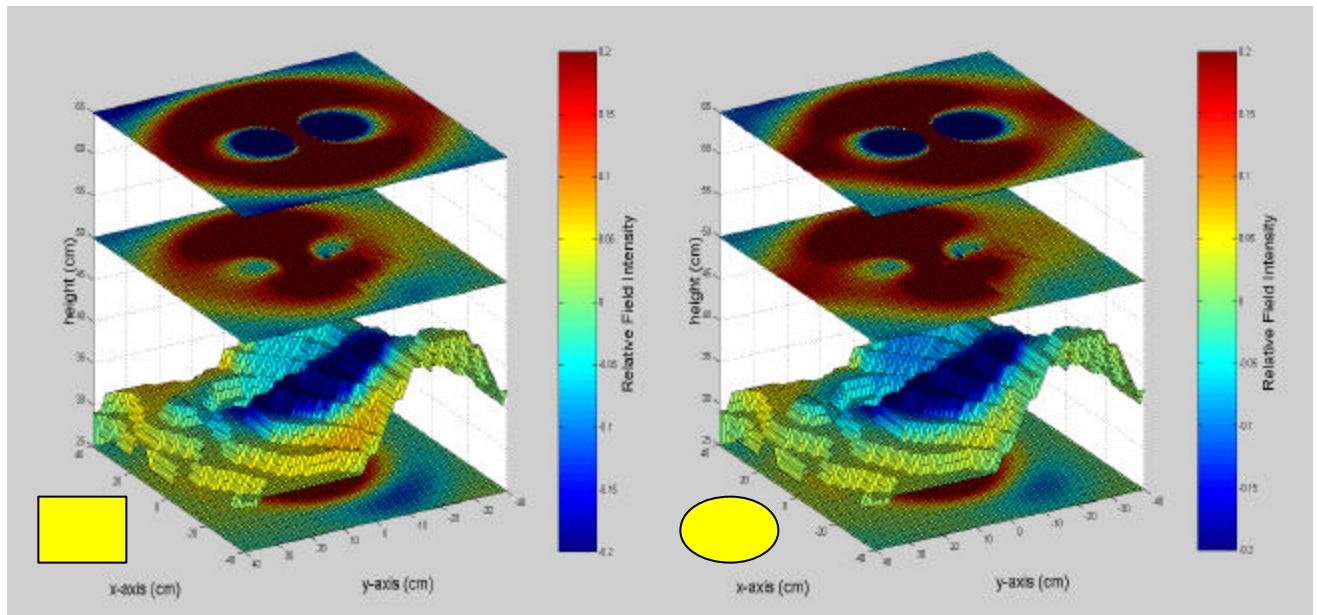


Figure 3: 3-D FDFD computational model of total E_x due to a dipole radiating at 960 MHz for a buried rectangular box target (left) and a circular cylindrical target (right) in clay loam soil with rough surface.

One of our major research projects is to improve the detection capability of ground penetrating radar by reducing clutter by simultaneously using multiple and receivers to focus waves on points under the ground surface. By “looking” at the mine from different angles at the same time, it is possible to emphasize the features of the mine, and minimize the clutter from the rough ground surface and uneven intervening soil. In addition, if the transmitted wave is planar and forward directed, the reflection from the ground surface tends to continue forward, away from the sensor. As such, it does not have as great an effect confusing the target signal. This effect has been confirmed both with computational models [8, 9], and in hardware, with an offset parabolic transmitting antenna [10]. Figure 4 depicts hardware fabricated by Lawrence Livermore National Labs. Test results for a nonmetallic mine buried in dry sand 1 inch under rough surface indicate an easily detectable anomaly with the size and approximate shape of the mine.

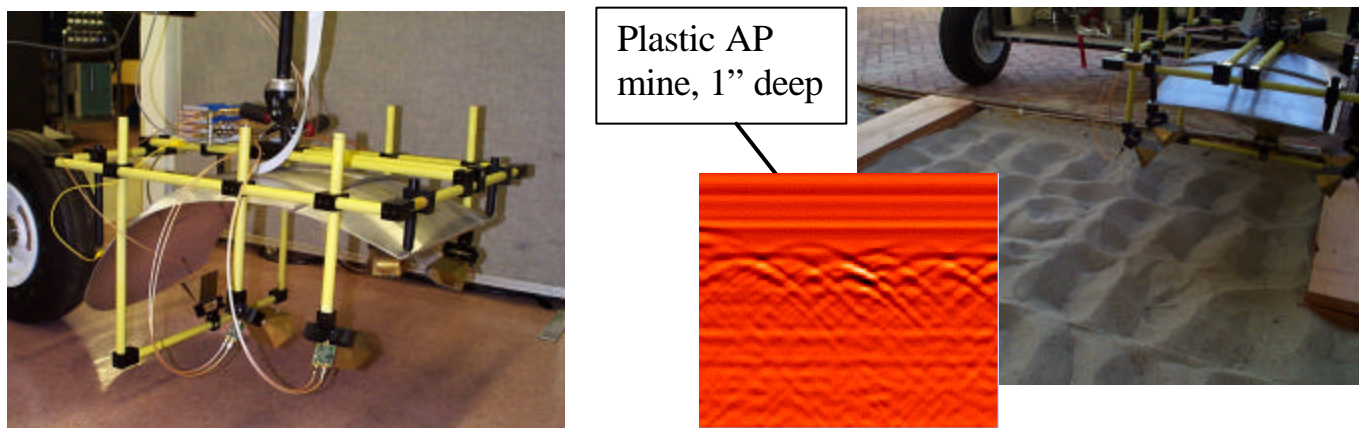


Figure 4: Offset parabolic reflector transmitter and multiple element receiver GPR handheld mine detector, with processed detection results for nonmetallic mine buried in rough sand

ACOUSTIC SENSING

Another exciting research area is the novel use of sound waves to detect acoustic echoes from buried mines. In deserts where the sandy soil is usually quite dry, non-metallic mines look electromagnetically very much like the background, and so are hard to pick out. To solve this problem, Northeastern University’s Demining team is using sound rather than radar or electromagnetic induction to find solid mine-like objects embedded in loose soil. The acoustic mismatch between granular soil and any metallic or non-metallic solid object is very high, giving a strong target-to-medium contrast.

One major problem with ground penetrating sonar is coupling through the ground surface. While medical ultrasound devices can be pressed firmly onto a patient's skin, direct ground contact in a minefield is not feasible. We have developed an alternative: using a laser beam to quickly heat the ground, causing rapid expansion, which in turn generates a sonic shock wave [11]. The scattered signals are detected either with a microphone placed very close to the ground, or interferometrically, using another laser beam to detect the ground motion.

This geometry, as well as sample detection results is shown in Figure 5. The mine simulant target is a bisected hockey puck, positioned slightly inclined to the ground surface. The surrounding material in this experiment was dry sand. A CO₂ TEA laser generated 150 mJ pulses several times a second, resulting in a string of acoustic pulses generated at the sand surface. A microphone was fixed relative to the laser, and the two were moved in a serpentine pattern relative to the ground. The measured signals were first filtered to remove as much of the ground surface reflection as possible, then the recorded signals were displayed as a function of position on the ground (bottom figure) [12]. The shape of the target is clearly distinguishable. It is also possible, by observing the arrival time of the target signal at the particular point, to determine the depth from the ground surface to the top of the target.

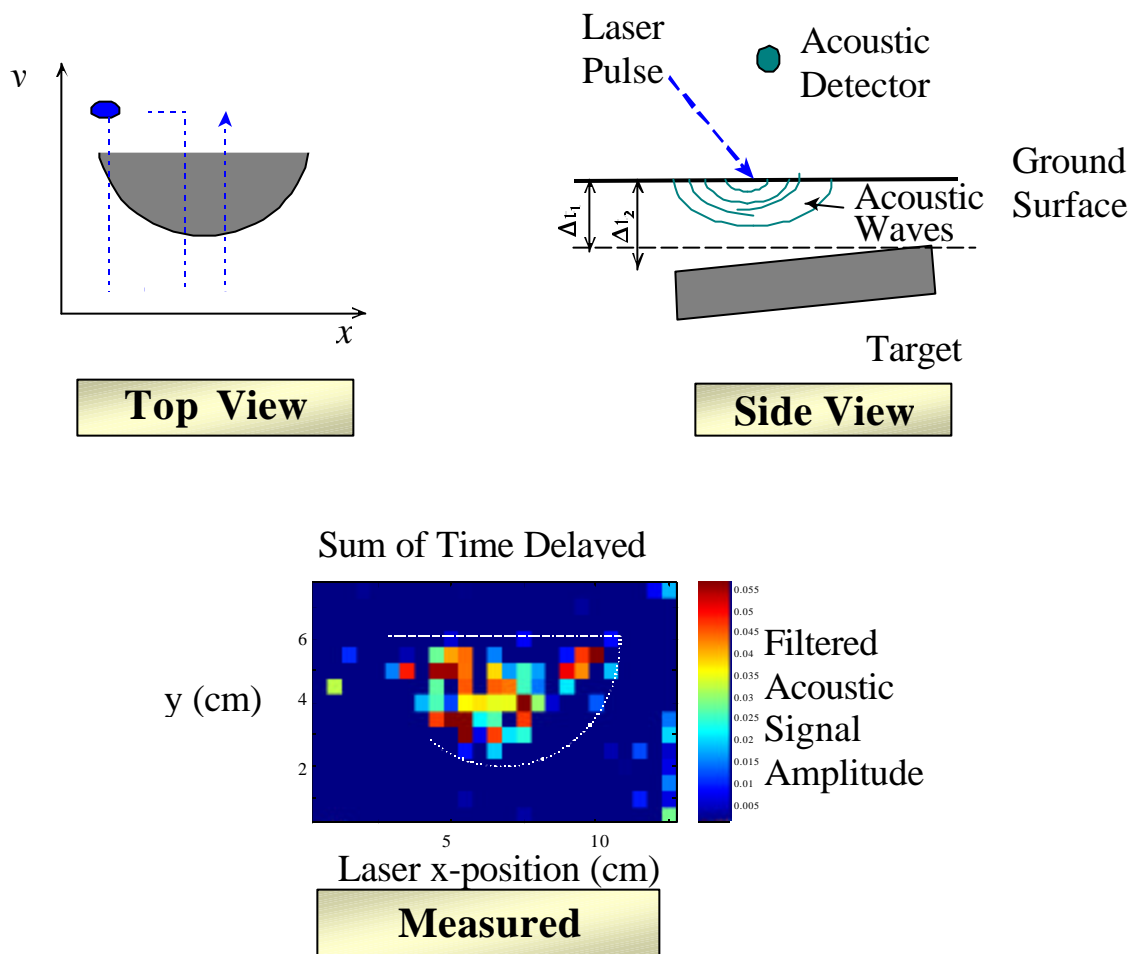


Figure 5: Laser induced acoustic sensing, geometry (top), and measured results (bottom).

INFRARED SENSING

Infrared detection is another sensing modality that has been used to find mines. Both polarimetric effects, which can distinguish exposed man-made objects [13], and thermal effects can be exploited. By measuring the subtle temperature changes from solar heating between soil with and without a mine, it is sometimes possible to find shallowly buried mines. Northeastern University's team has developed a novel means of using IR thermography by simulating the sun with microwave heating [14]. Our heater points down, toward the ground, deeply heating a small patch of soil. Meanwhile, we take IR pictures of the soil patch and look for anomalous heating patterns. Instead of being limited to sunrise and sunset, when the solar heating changes are most noticeable, our system can be used all day. Figure 6 shows this concept of heating the ground with microwaves. The right image indicates the temperature differences on the sand surface due to non-uniform heating of the soil above the targets. The shapes of the targets are clearly visible.

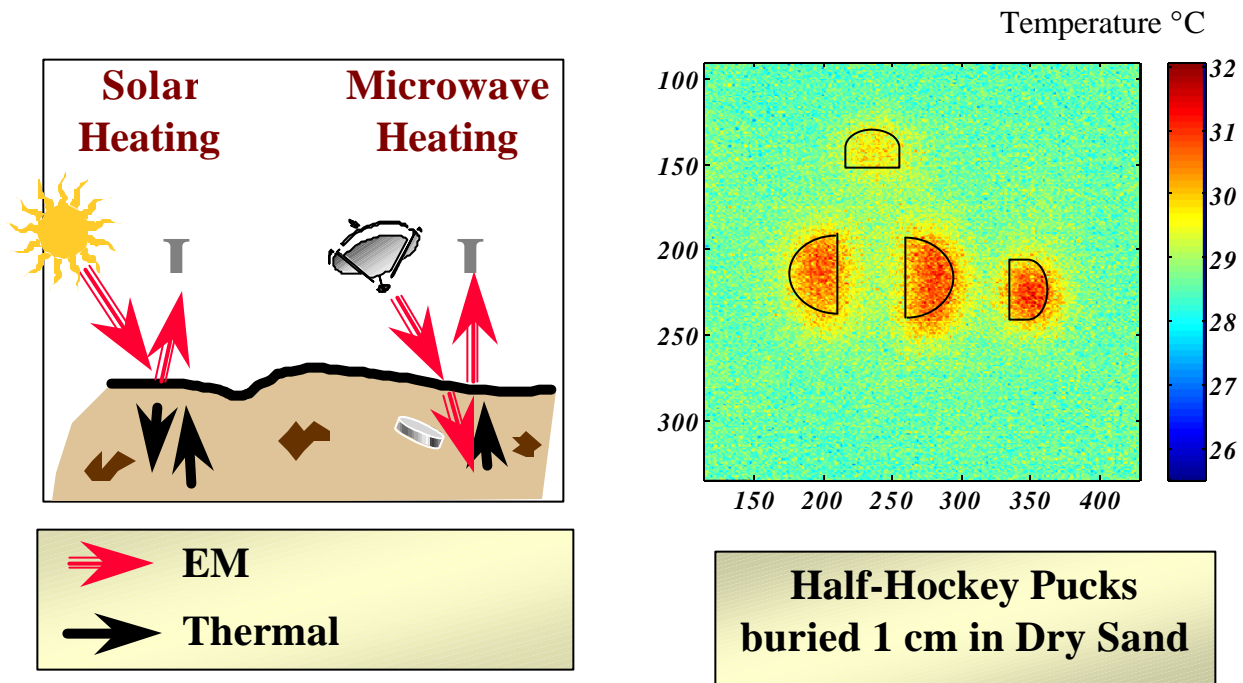


Figure 6: Microwave enhanced IR thermography concept (left) and measured results (right).

It is also possible to use multiple microwave frequencies for heating the ground. Since the surface heating pattern is relatively insensitive to frequency, while the heating at depth depends strongly on frequency, taking the difference from two heatings will better show the presence of buried anomalies [14].

CONCLUSIONS

In designing each of our sensors, our emphasis is to use as much of the available information as possible: use many different probing signals, with differing frequencies; use the widest possible view, with many inexpensive sensors coupled together; use the information gained from mine-free terrain and from confirmed terrain with mines to give improved detection confidence; use the fact that mines are stationary, while the detectors can move, and above all, keep the advanced technology simple in principle, so that it can detect mines in dirty, or frozen, or hard-to-traverse terrain.

ACKNOWLEDGEMENT

This work was supported by The Army Research Office Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013.

REFERENCES

- (1) Hibbs, A., *et. al.* "Field tests results of a nuclear quaterupole resonance landmine detection system," *SPIE Aerosense Conference*, Orlando, FL, April 2000.
- (2) A. Witten, I.J. Won, and S. Norton, "Subsurface imaging with broadband electromagnetic induction," *Inverse problems*, Vol. 13, pp. 1621-1639, 1997.
- (3) I.J. Won, D. Keiswetter, and E. Novikova, "Electromagnetic induction spectroscopy," *Jour. Environmental and Engineering Geophysics*, Vol. 3, Issue 1, pp. 27-40, 1998
- (4) E. Miller, W. C. Karl and S. Norton, "On the detection of buried mines from multifrequency, inductive measurements," *SPIE AeroSense Symposium, Detection Technologies for Mines and Minelike Targets*, Orlando, Florida, April 1997.
- (5) El-Shenawee, M. and Rappaport, C., "Quantifying the Effects of Different Rough Surface Statistics for Mine Detection Using the FDTD Technique," *SPIE Aerosense Conference*, Orlando, FL, April 2000.
- (6) C. Rappaport, M. Kilmer, and E. Miller, "Accuracy Considerations in Using the PML ABC with FDFD Helmholtz Equation Computation", (Invited Paper) accepted for publication in *International Journal of Numerical Modeling*, 1999.
- (7) C. Rappaport and S. Winton, "Using the PML ABC for Air/Soil Wave Interaction Modeling in the Time and Frequency Domains," accepted for publication in *IEEE Transactions on Antennas and Propagation*, 1999.
- (8) C. Rappaport, S. Wu, and S. Winton, "FDTD Wave Propagation Modeling in Dispersive Soil Using a Single Pole Conductivity Model," *IEEE Transactions on Magnetics*, vol.35, May 1999, pp. 1542-1545.
- (9) Jin, D. and Rappaport, C., "FDTD Simulation of a Parabolic Antenna GPR transmitter," accepted for publication in *Progress in Electromagnetics Research Symposium*, July 2000.
- (10) Rappaport, C., Azevedo, S., Rosenbury, T., Gough, J, and Jin, D., "Handheld Forward-Looking Focused Array Mine Detection with Plane Wave Excitation," *SPIE Aerosense Conference*, Orlando, FL, April 2000.

- (11) S. McKnight, C. DiMarzio, and D. Hogenboom and G. Sauermann, "Laser-induced acoustic detection of buried objects," *SPIE's 12th Annual International Symposium on Aerospace /Defense Sensing, Simulation, and Controls*, April 13-17 1998, pp .841-847.
- (12) S. McKnight, W. Li, and C. DiMarzio, "Imaging of buried objects using laser-induced acoustic detection," *SPIE's 13th Annual International Symposium on Aerospace /Defense Sensing, Simulation, and Controls*, April 13-17 1999, pp .231-238.
- (13) H. Scott, S. Jones and Iannarilli, F.I. 1999. "Imaging Infrared Polarimetry: Initial Results and Potential in Detection of Scatterable Mines and Surface Disturbances." *Proceedings of SPIE Vol. 3710, Detection Remediation Technologies for Mines and Minelike Targets IV*. 5-9 April 1999, Orlando FL
- (14) Charles A. DiMarzio, Li Wen, Carey M. Rappaport, Gerhard Sauermann, Herman E. Scott, "Microwave--Enhanced Infrared Thermography," *SPIE AeroSense Conference*, Orlando, FL, April 1999.
- (15) Oktar, T., Rappaport, C., and DiMarzio, C. "The Effects of Surface Roughness on Microwave Heating of Soil for Detection of Buried Land Mines," accepted for publication in *SPIE Aerosense Conference*, Orlando, FL, April 2000.

Semi-analytic mode matching for detecting nonmetallic mines buried in realistic soils

Ann W. Morgenthaler and Carey M. Rappaport
Center for Electromagnetics Research
Northeastern University, Boston, MA 02115

ABSTRACT

The ultra-wideband detection of plastic land mines buried in lossy, dielectric soils is simulated using a new semi-analytic mode matching (SAMM) algorithm. Here, we apply SAMM to the 3D canonical problem of finding the nonspecular reflection of an obliquely-incident plane wave on a lossy dielectric half-space containing a small, shallowly-buried dielectric sphere. However, SAMM can also be extended to smoothly-varying convex-shaped mines buried under modestly rough ground. In the SAMM algorithm, the frequency-dependent scattered fields are constructed from moderately low-order modal superpositions of spherical waves, each satisfying the Helmholtz equation in its respective material (air, ground, or mine). By least squares fitting, mode coefficients are found which optimally match all boundary conditions at designated points along the boundary surfaces. Spherical wave expansions are chosen at multiple coordinate centers so that small numbers of modes are needed to give convergent results.

The inverse discrete Fourier transform (IDFT) of a wideband GPR reflected signal yields a time domain signature dependent on mine and soil characteristics as well as the burial depth and radius of the mine in a particularly simple way: scattering times are dominated by analytically-computed Mie scattering from a spherical mine in an infinite soil background. Specifically, two or more characteristic time peaks are observed in SAMM simulations, suggesting that the ultra-wideband spectral radar response may yield particular advantages not exploited by currently-employed detection systems.

Keywords: GPR, mine detection, dielectric target imaging, modal analysis, ultra-wideband radar

1. INTRODUCTION

The challenging problem of detecting small, buried, nonmetallic objects (e.g., antipersonnel land mines) by GPR in lossy, dispersive, rough soil has been studied extensively,¹⁻³ and fully three-dimensional numerical methods (e.g., FDFD, FDTD,⁴ and moment method^{5,6}) tend to be computationally expensive and time-consuming ways to analyze real problems. Additionally, much of the radar literature analyzes far-field scattering of large, metallic objects in air,⁷⁻⁹ although antipersonnel land mine detection is usually performed in the near field; i.e., the radar transmitter and/or receiver are often held within a few wavelengths of the mine and/or ground surface. Therefore, many of the established far-field techniques are not applicable for detecting buried mines. Analytic simulations^{10,11} (e.g. integral equations, T-matrices) can often be difficult to implement for problems with half-space geometries. We therefore seek a hybrid method which is both physically-based and is applicable to a wide

range of geometries, with special emphasis on near-field scattering: the SAMM algorithm is one such approach.

In other papers^{12,13} we compare SAMM with FDFD simulations of a 2D buried cylinder under a ground plane, and find excellent agreement between the two methods, even when the ground surface is randomly rough, the ground material is modeled by a frequency-dependent complex dielectric constant, and the mine is not perfectly spherical.

Because SAMM is relatively fast and applies over a large frequency range, ultra-wideband frequency sweeps of the scattered fields can be simulated, enabling time-domain analysis of 3D buried mine scattering via the IDFT. Characteristic time peaks can be readily identified (even in fairly lossy ground media) which have a simple physical basis; a corresponding FDTD analysis would need to finesse the problem of dispersive media.¹⁴

2. SEMI-ANALYTIC MODE MATCHING (SAMM)

The geometry of the canonical buried mine problem is given in Fig. 1, where a sphere of complex dielectric ϵ_m and radius R is buried d meters below the surface of the ground; the perpendicular distance from the center of the mine to the surface is $\ell = R + d$. The ground has complex dielectric ϵ_g and extends infinitely far below the mine; the boundary between the ground and air is an infinite flat plane.

Since the mine is spherical, we describe the waves in each region by the electric and magnetic Hertz potentials¹⁵ Π^e and Π^m :

$$\mathbf{H} = \nabla \times (\mathbf{r}\Pi^e)/\eta; \quad \mathbf{E} = \nabla \times (\mathbf{r}\Pi^m) \quad (1)$$

which satisfy the scalar wave equation $(\nabla^2 + k^2)\Pi^{e,m} = 0$ where ∇^2 is the spherical Laplacian. The dispersion relation $k^2 = \omega^2\mu_0\epsilon$ must be obeyed in each of the three regions (air, ground, and mine). The Hertz potentials Π^e and Π^m therefore have the modal structure

$$\Pi^{e,m}(\mathbf{r},k) = \lim_{N_{\max} \rightarrow \infty} \sum_{n=1}^{N_{\max}} \lim_{M_{\max} \rightarrow n} \sum_{m=-\min(n,M_{\max})}^{\min(n,M_{\max})} C_{nm}^{e,m} f_n(kr) P_n^m(\cos\theta) e^{im\phi} \quad (2)$$

where the mode coefficients $C_{nm}^{e,m}$ have units of electric field and we assume a time-harmonic dependence $\exp(-i\omega t)$ with angular frequency ω throughout this paper. The spherical Bessel function $f_n(kr)$ has order n and possibly complex argument kr ; $P_n^m(\cos\theta)$ is the associated Legendre polynomial of order n and degree m .^{16,17} For the SAMM algorithm to succeed, it is crucial to expand the spherical modes about a proper choice of coordinate centers so that the fields converge with increasing mode integers n and m . Otherwise, the scattering fields cannot be constructed with a finite number of modes. In the discussion which follows, we shall consider only incident TE plane waves, as the TM case is easily found by duality.

We begin the SAMM algorithm by finding the reflected and transmitted plane waves that would exist were no mine buried below the ground. For oblique incidence, the usual textbook formulas result.¹⁸ The plane wave transmitted into the ground serves as the driving term for Mie scattering, which can be thought of as the first reflection from the mine before any rescattering from the air/ground interface has occurred. The transmitted plane wave must therefore be decomposed into spherical modes originating at the mine center with coefficients $C_{nm}^{e,m}$ given in Ref. [13]. Because the spherical mine has a boundary which falls along the contour $r = R$ in spherical coordinates, it is straightforward to match boundary conditions analytically on the spherical surface with the spherical modes, resulting in textbook Mie scattering fields.^{15,19} These Mie modes now drive further rescattering from both the air/ground interface and mine, described by spherical mode expansions about both the mine and its image in air; these rescattering modes, or r-modes, must be truncated to be useful in numerical simulations. That is, the radial mode index n is not allowed to exceed a specified N_{\max} while the angular mode index m is limited to the smaller of n and a specified M_{\max} . In the case of a normally-incident plane wave (which

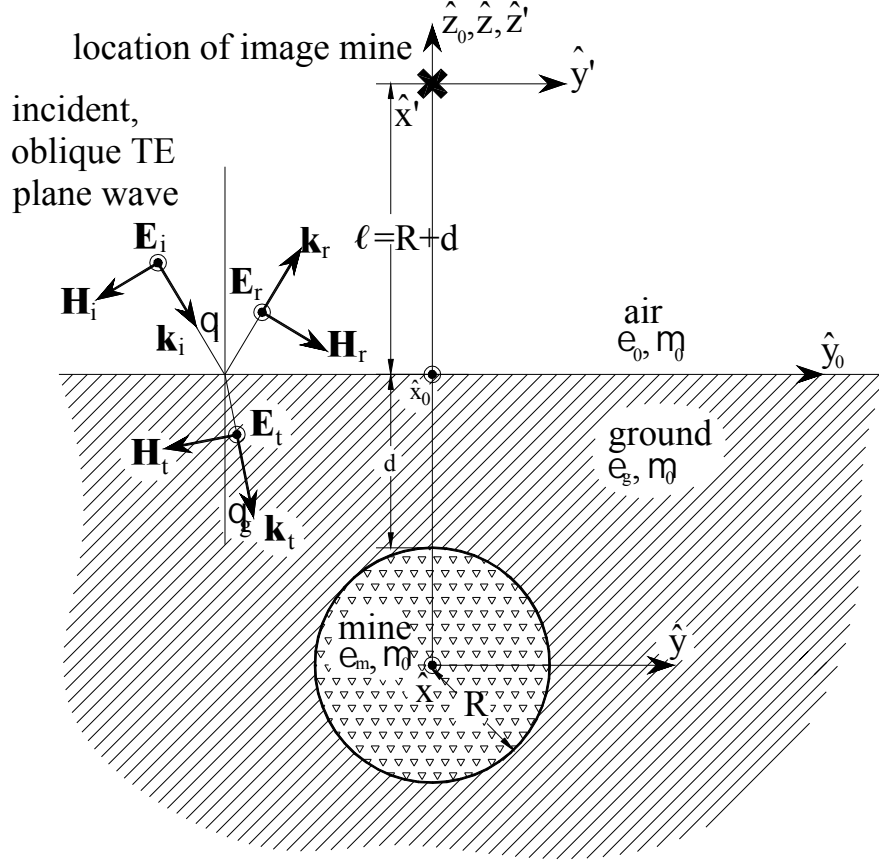


Figure 1: Geometry of the buried land mine. A spherical mine with radius R and complex permittivity ϵ_m is buried a depth d below an infinite planar surface. Air (ϵ_0) extends infinitely far above and ground (ϵ_g) infinitely far below the ground plane. The coordinate systems C_0 , C , and C' are centered at the plane, mine and image mine, respectively, where the image mine is centered a distance $\ell = R + d$ above the plane.

can be generated by a dipole infinitely far above the ground plane), we can show that $M_{\max} = 1$, meaning that the field components have dipole-like $\sin \phi$ or $\cos \phi$ dependence.

In Fig. 2, we show the error caused by truncating the spherical mode expansion of a normally-incident plane wave as a function of N_{\max} for three values of $k_g R$ at the highest frequency in the simulation (8 GHz). Although it may be necessary to use 30-50 spherical modes to compose the plane wave accurately, the more computationally difficult rescattering can often be truncated after many fewer modes (10-20). In essence, rescattering is a local event, limited to a relatively small region of space, and constructed from modes centered near the observation point in the vicinity of the mine. In contrast, constructing an infinite plane wave from local spherical modes requires many more of them since the plane wave has infinite extent. Keeping large numbers of plane wave decomposition modes and their corresponding Mie scattering modes is simple computationally, since these mode coefficients are computed analytically from exact Mie solutions. Unlike the Mie modes, the r-modes are not decoupled because the planar boundary is not a natural contour in spherical coordinates: a single excitation coefficient leads to an infinite number of coupled spherical modes. Fortunately, *with appropriate choices for the scattering centers of these additional modes*, only a relatively few need be kept.

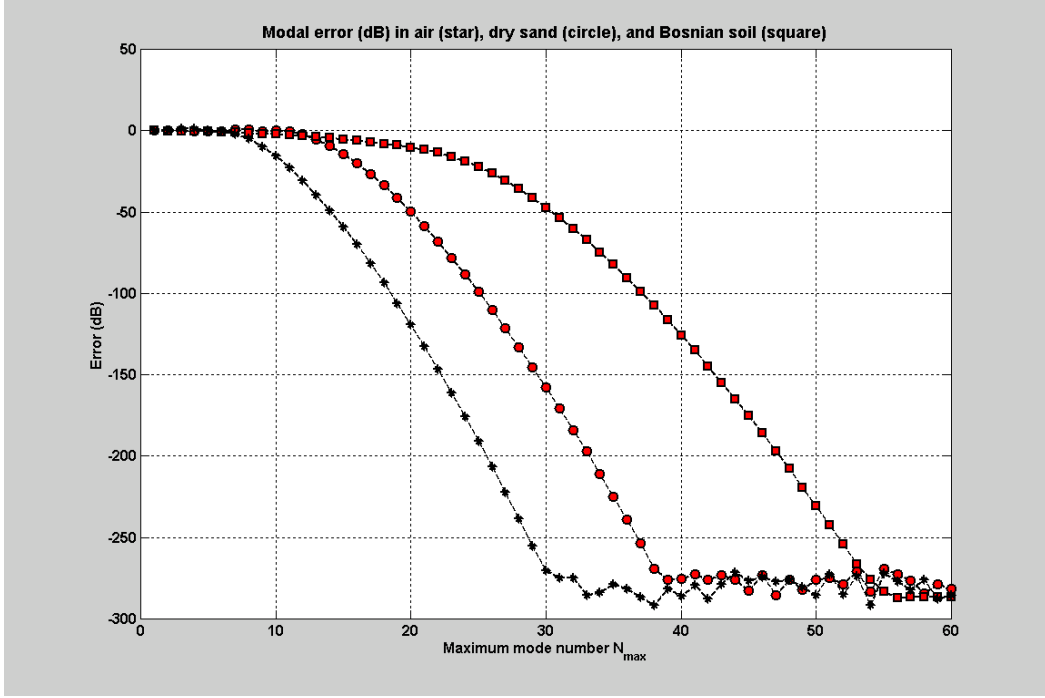


Figure 2: Error (dB) in decomposing a normally-incident TE plane wave in air [$\epsilon_g = \epsilon_0$], dry sand²⁰ [$\epsilon_g = 2.55(1+i0.01)\epsilon_0$], and Bosnian soil from the Alicia test site²¹ [$\epsilon_g = 8.66(1+i0.2)\epsilon_0$ at 8 GHz]. The discrepancy between the transmitted plane wave and its modal decomposition is measured on the surface of a 5-cm spherical mine. We observe that larger error results for a given truncation level N_{\max} in the soil with the larger dielectric constant.

Because each mode couples only to itself, the Mie coefficients can be computed analytically to any order in n and m . We expect that the plane wave plus the Mie scattering will account for much of the buried mine scattering. (For $\epsilon_g = \epsilon_i$, the Mie fields are the exact solution.) However, re-scattering from the planar boundary will lead to additional fields which we shall expand in spherical modes centered at both the mine ($z_0 = -\ell$) and the image of the mine ($z_0 = \ell$). Just as the simple problem of an antenna over a perfectly-conducting plane is solved by creating an image below that plane, we expect the image modes to model the infinite nature of the ground plane. Using the image scattering center for a modal expansion will allow convergence (and hence truncation) of the r-modes. We observe that modes will tend to converge only if all the centers from which scattered waves appear to originate are included; selection of the appropriate scattering centers can be more an art than a science, particularly in complex cases of irregularly shaped mines and rough surfaces. We may test the final solution in two ways: (1) by demonstrating convergence of the field to a stable value as more modes are chosen, and (2) by comparing field results to simulations of the identical geometry by established methods like FDFD or the moment method. We shall find that for the problem of a sphere buried under a flat surface, the actual scattering center plus image center is a reasonable way to implement SAMM. For a rough surface, these scattering centers will prove insufficient and additional ones will need to be chosen.

Once the scattering modes have been identified, we must ensure that the fields arising from the Mie and r-modes obey all boundary conditions (BCs) at the boundary surfaces. Were we satisfying BCs analytically, it would be sufficient to match the tangential components of \mathbf{E} and \mathbf{H} : four BCs would be needed at each boundary for a unique solution.¹⁵ Because we do not have a truly analytic solution (we are matching the field at discrete points), we shall enforce continuity of both tangential components of \mathbf{E} and \mathbf{H} as well as the normal components of \mathbf{D} and \mathbf{B} . We then use least squares analysis¹⁷ to find the mode coefficients that best fit the six BCs at each of N_{surface}

sampled surface points. We require all six (non-independent) BCs to be met at each point on each relevant surface, since minimizing the error in the four necessary BCs could lead to large errors in the remaining two, a common issue in least squares matrix inversion. There will be problems for which expansion in finite numbers of spherical modes will be quite difficult: targets which are not “fairly” spherical, mines with sharp protrusions, rough surfaces with steep slopes etc. are probably not amenable to modal expansion techniques. Although sharp metallic target corners, for which fields can approach infinity, are also poorly modelled with SAMM, the plastic antipersonnel mines for which SAMM simulations are successful present the greater detection challenge and are sufficient reason to utilize the new algorithm.

For an infinite planar surface, we shall not be able to allocate points in such a way as to completely cover the boundary. Placing a large fraction of the available points too far away from the region of the plane immediately over the mine results in inaccuracy near the mine, since the far points adjust the global behavior of unimportant regions of space and insufficient numbers of points are left to describe the near field. Conversely, placing points over too small a region of the plane results in modes which are highly degenerate; if the observation point is located above the plane (rather than on it), a wider planar region must also be covered. By comparing SAMM to standard simulations, we can get a sense of how best to allocate the surface points.

In summary, we implement SAMM by: (1) Choosing the appropriate scattering centers, about which modes will be expanded, (2) Creating a finite series of spherical modes with the appropriate values for wave numbers k and wave impedances η so that these modes automatically satisfy the Helmholtz equation in air, ground, and target material, (3) Allocating points on all boundaries and finding the mode coefficients which best fit—in the least squares sense—all boundary conditions at these points. We choose the spherical Bessel functions properly, so that radiation conditions and singularity issues are addressed: fields that fall within the origin of a modal coordinate system and fields which represent plane waves are specified by spherical Bessel functions of the first kind $j_n(kr)$, while scattering fields are described by outwardly-propagating spherical Hankel functions $h_n^{(1)}(kr)$, and finite regions which do not include the origin have two linearly independent field solutions.

3. APPLYING SAMM TO THE BURIED SPHERE PROBLEM

Referring to Fig. 3, we shall express the fields within each region of the problem as:

$$\Pi_{\text{air}}^{e,m} = \sum_{n=0}^{N_m} \sum_{m=-M_m}^{M_m} r_{nm}^{e,m} h_n^{(1)}(k_i r) P_n^m(\cos \theta) e^{im\phi} \quad (3)$$

$$\begin{aligned} \Pi_{\text{ground}}^{e,m} = & \sum_{n=0}^{N_{\max}} \sum_{m=-M_{\max}}^{M_{\max}} \left(C_{nm}^{e,m} j_n(k_g r) + A_{nm}^{e,m} h_n^{(1)}(k_g r) \right) P_n^m(\cos \theta) e^{im\phi} + \\ & \sum_{n=0}^{N_m} \sum_{m=-M_m}^{M_m} q_{nm}^{e,m} h_n^{(1)}(k_g r) P_n^m(\cos \theta) e^{im\phi} + \sum_{n=0}^{N_m} \sum_{m=-M_m}^{M_m} q_{nm}'^{e,m} h_n^{(1)}(k_g r') P_n^m(\cos \theta') e^{im\phi'} \end{aligned} \quad (4)$$

$$\Pi_{\text{mine}}^{e,m} = \sum_{n=0}^{N_{\max}} \sum_{m=-M_{\max}}^{M_{\max}} T_{nm}^{e,m} j_n(k_m r) P_n^m(\cos \theta) e^{im\phi} + \quad (5)$$

$$\sum_{n=0}^{N_m} \sum_{m=-M_m}^{M_m} t_{nm}^{e,m} j_n(k_m r) P_n^m(\cos \theta) e^{im\phi} + t_{nm}'^{e,m} h_n^{(1)}(k_m r') P_n^m(\cos \theta') e^{im\phi'} \quad (6)$$

The wave prefixed by the coefficient $C_{nm}^{e,m}$ is the transmitted plane wave, while the modes multiplied by $A_{nm}^{e,m}$ and $T_{nm}^{e,m}$ are the Mie coefficients of the first scattering from the mine; these Mie modes are truncated at N_{\max} and M_{\max} . The r-modes, truncated at $N_m \leq N_{\max}$ and $M_m \leq M_{\max}$, have coefficients $r_{nm}^{e,m}$, $q_{nm}^{e,m}$, $q_{nm}'^{e,m}$, $t_{nm}^{e,m}$, and $t_{nm}'^{e,m}$. Note that there is no part of the Hertz potential in air arising from the image scattering center; the image fields are valid only *below* the ground plane. In the case of a normal plane wave excitation, $M_m = M_{\max} = 1$

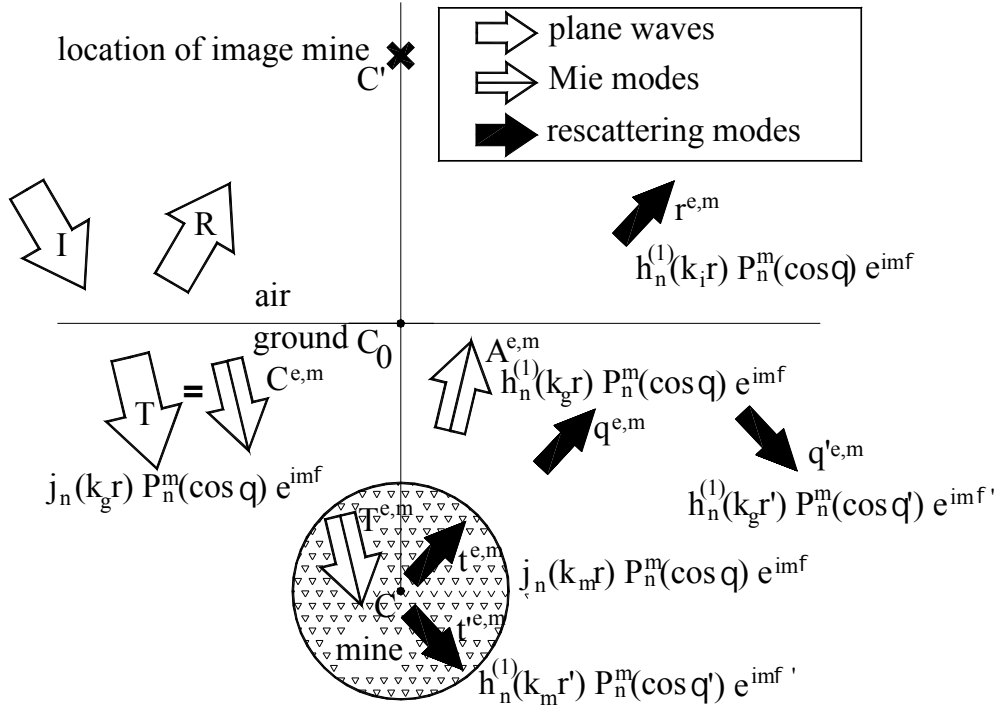


Figure 3: Schematic of plane waves, Mie modes, and spherical scattering modes (r-modes) in the SAMM algorithm. The incident plane wave (I) is reflected (R) and transmitted (T). This latter wave is decomposed into spherical modes, which lead to Mie scattering modes ($A^{e,m}$, $T^{e,m}$). These Mie modes reflect and refract at the ground plane, leading to r-modes ($r^{e,m}$, $q^{e,m}$, $q'^{e,m}$, $t^{e,m}$, and $t'^{e,m}$) valid in their respective domains as shown. The unprimed spherical modes originate at the center of the mine; the primed spherical modes originate at the image mine and the mode indices nm have been suppressed for clarity.

so each of the ten r-mode families will require $2N_n$ coefficients, leading to $20N_n$ unknown mode coefficients. For oblique incidence (or for rough surface scattering of any type of plane wave), $|m|$ is not limited to one and approximately $20N_m M_m$ coefficients will need to be fitted. Clearly, SAMM is tractable only if N_m and M_m are limited to relatively small values. Further analysis in this paper is limited to the case of normally-incident plane waves.

With the mode coefficients determined by inverting an overconstrained matrix equation having approximately $6N_{\text{surface}}$ equations and $20N_m$ unknowns so as to minimize the least squares error, the electric and magnetic fields can be reconstructed. Fig. 4 shows the transverse field component E_x plotted in both the y - z plane (at $x_0 = 0$), and x - y plane (at $z_0 = 0$). There is quite a striking difference between filling the ground with air, dry sand and Bosnian soil, with air-filled ground returning the strongest signal of the three. Determining mine signatures using air as the surrounding medium is clearly doomed to failure. Calculation of the r-modes for the single frequency in Fig. 4 took about 6 minutes in Matlab 5.3 on a Pentium desktop computer. The linear matrix has 4818×240 elements and is densely filled. A spherical Bessel function and a Legendre polynomial must be computed for each element, which is the major computational burden in the simulation. Although $N_m = 12$ in this simulation, $N_{\text{max}} = 25$ so additional special functions must be calculated for the Mie scattering which forms the right hand side of the overconstrained matrix equation. Using a compiled language would greatly increase computational speed.

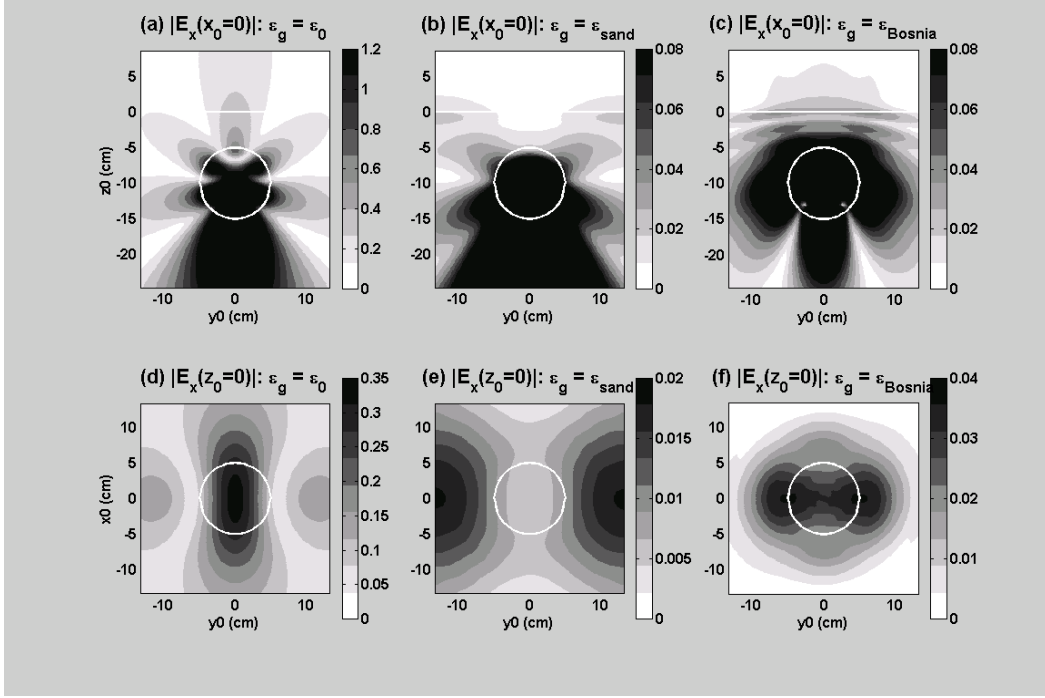


Figure 4: Magnitude of scattered electric field $|E_x|$ at (a)–(c) $x_0 = 0$ (y - z slice) and (d)–(f) $z_0 = 0$ (x - y slice). A 3 GHz TE plane wave is normally incident on a half-space with ground having the electrical characteristics of (a),(d) air [$\epsilon_g = \epsilon_0$] (b),(e) dry sand²⁰ [$\epsilon_g = 2.55(1 + i0.01)\epsilon_0$], and (c),(f) Bosnian soil from the test site Alicia²¹ [$\epsilon_g = 8.82(1 + i0.13)\epsilon_0$ at 3 GHz]. In all cases, the mine is a 5-cm sphere filled with TNT²⁰ [$\epsilon_m = 2.9(1 + i0.001)\epsilon_0$] buried 5 cm below the ground surface. The mine and plane, where appropriate, are outlined in white on each plot. Because the gray scale is nonlinear to enable small amplitude features to be seen, note that the maximum field magnitudes are: (a) 4.5, (b) 0.80, (c) 0.80, (d) 0.33, (e) 0.018, and (f) 0.036 (normalized to the plane wave transmission coefficient).

4. MINE DETECTION USING SAMM

The electric field is plotted in Fig. 4 at a single frequency, but one of the powerful features of the SAMM algorithm is that it applies over a very wide frequency band. In Fig. 5, the electric field component E_x is plotted as a function of frequency at the observation point $x_0 = y_0 = z_0 = 0$, which is directly over the mine. Once again, the mine is a 5-cm radius TNT sphere buried 5 cm below the planar surface, the ground is either air, dry sand, or Bosnian soil, the excitation is a normally-incident TE plane wave, and the frequency sweep runs between 0.5 and 8 GHz in increments of 0.025 GHz. All components of electric and magnetic fields are matched at the planar and mine boundary surfaces at 401 and 258 points, respectively, and the planar data points are placed within a circle of radius 20 cm centered at the observation point. Because the plane wave is normally incident, the angular modes have only $\exp(\pm i\phi)$ dependence, so $M_{\max} = M_m = 1$. The radial mode cutoffs are $N_{\max} = 30$ (air), $N_{\max} = 40$ (sand), $N_{\max} = 50$ (Bosnian soil) and $N_m = 16$. The resulting frequency sweeps are windowed to eliminate the worst errors, caused by an insufficiently large area over which the planar points are placed at low frequencies, and too few r-modes (N_m too small) at high frequencies; the value of N_m is chosen to be a compromise between frequency range and computational speed. In these simulations, the plane surface points are placed along the intersections of evenly-spaced radial (ρ) and angular (ϕ) contours, the sphere surface points are evenly-spaced in the θ and ϕ directions, and all points are given equal weighting. The optimal locations and weightings for the surface points which minimize field error are still unknown.

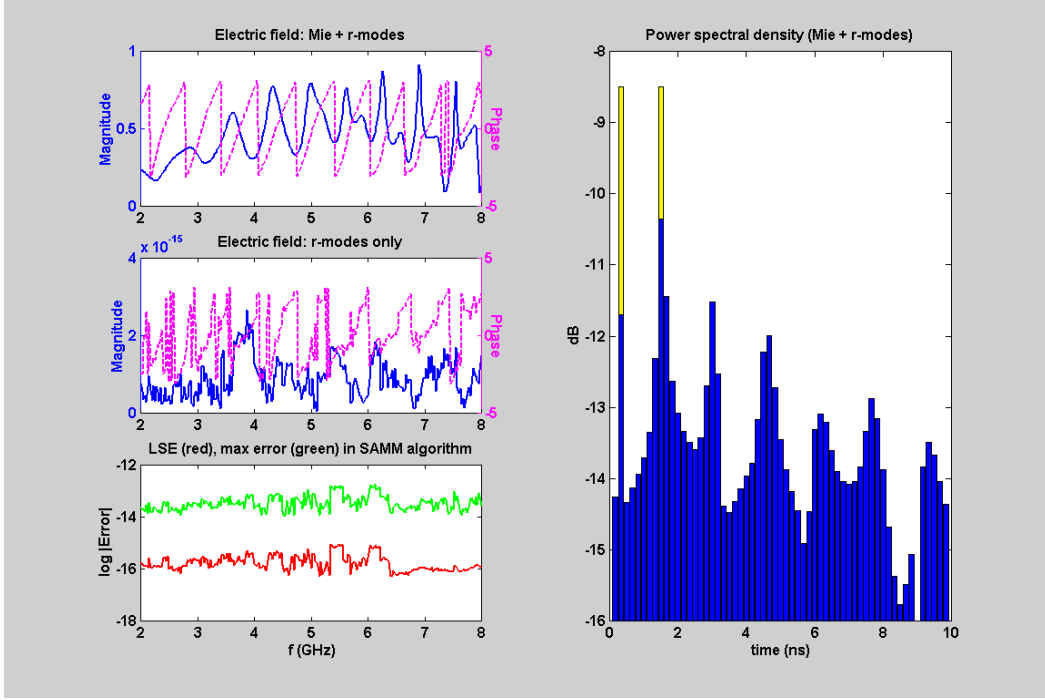


Figure 5: Frequency and time domain IDFT of E_x calculated at the observation point $x_0 = y_0 = z_0 = 0$ for the case where the ground has the electrical characteristics of air (true Mie scattering). The upper left plot is the magnitude and phase (dashed line) of the scattered field E_x over the frequency range 0.5 to 8 GHz taken at intervals of 0.025 GHz. The plot immediately below is the r-mode scattering which is essentially zero because with no discontinuity between ε_i and ε_g , exact Mie scattering results. The bottom left plot shows the least squares error (lower curve) in the SAMM simulation and also the maximum error (upper curve) of any one point along the planar or sphere surface. The IDFT of the total scattered field (Mie plus r-mode) is given on the right. Both characteristic times T_0 and T_1 ($T_0 < T_1$) are clearly observed, as well as additional characteristic times separated by $\Delta t = 4R/c_m + 2R/c_g$ corresponding to multiple reflections within the sphere. Perfect matching of the ground to air is responsible for the additional structure in the IDFT, which tends to disappear in imperfectly-matched, lossy soils (see next two figures).

In each of Figs. 5 to 7, the upper left plot gives the magnitude and phase of the electric field E_x as a function of frequency for the entire scattered field; the middle left plot gives just the r-mode scattering. The lower left plot shows the least square error (LSE) and maximum error over the surface points as a function of frequency; in general, the LSE is about two orders of magnitude less than the maximum error. The right plot is the IDFT of the entire scattered field. Overlaid on this IDFT are the two characteristic times T_0 and T_1 , given by

$$\begin{aligned} T_0 &= 2d/c_g \\ T_1 &= T_0 + 4R/c_m \end{aligned} \quad (7)$$

In all three figures, these two characteristic times are a dramatic feature.

The T_0 and T_1 values correspond to the round trip time of a ray from a point directly above the mine traveling through ground and mine with velocities c_g and c_m , respectively. Because the mine has the well-measured dielectric constant of TNT and a characteristic size and shape (assumed here to be a sphere of radius R), the difference $T_1 - T_0 = 4R/c_m$ is specific to the mine target. Rocks, which have different dielectric constants and sizes, will give different pairs of time peaks. Once these time peaks are identified, the vertical position of the mine can be determined by estimating the ground dielectric constant (and thus the resulting ground propagation

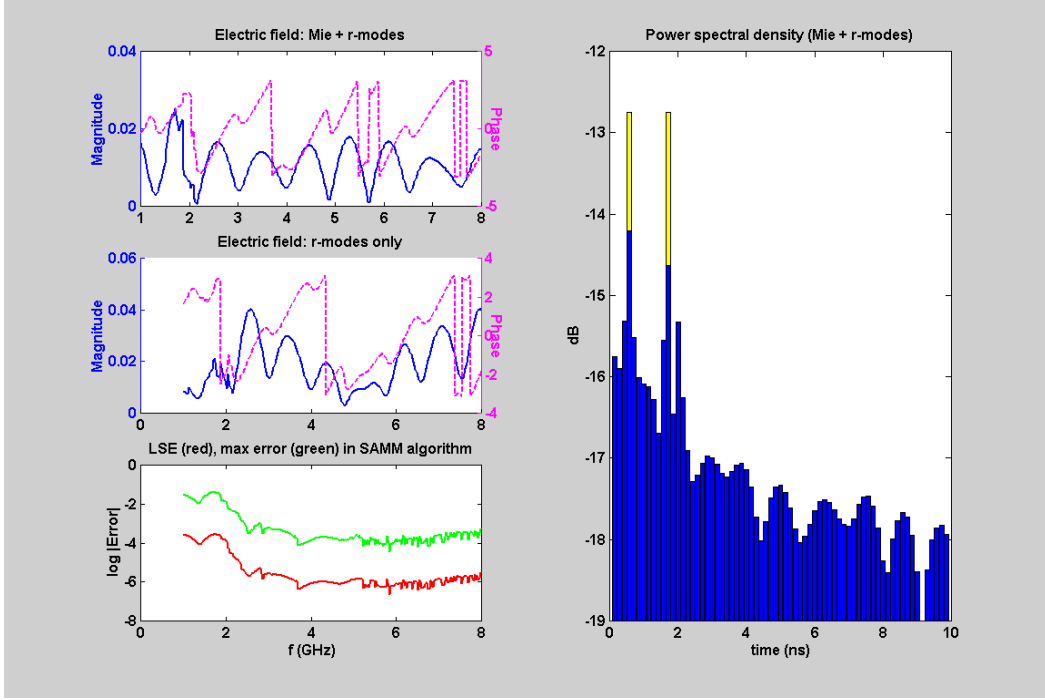


Figure 6: Scattering of a TNT sphere buried in dry sand, with all other parameters identical to those of the previous figure. The frequency is plotted from 1 to 8 GHz in increments of 0.025 GHz and the LSE is significantly worse than in the case of air because the r-modes play a dominant role in the scattering. The IDFT of the total scattering signal possesses the characteristic times T_0 and T_1 as well as an additional time constant corresponding to $T_1 + 2d/c_g$. Although the sand is not very lossy, it is significantly mismatched to the air. No characteristic times greater than 2 ns are observed, unlike the scattering from the mine surrounded by air shown in the previous figure.

velocity), and multiplying by $T_0/2$. Since modal expansions of the target are relatively insensitive to small variations of mine shape and air/ground interface compared with burial depth and dielectric characteristics, the SAMM method efficiently identifies the salient features needed for mine detection.

5. CONCLUSIONS

A new algorithm is developed which finds all field components in the region of interest near a buried dielectric land mine by fitting spherical modes to all interfaces. This SAMM algorithm quickly and efficiently computes the frequency domain fields where they would be measured (on or near the ground surface in the vicinity of the suspected target) using the important physical aspects of the detection problem: size, aspect ratio, orientation, burial depth, and dielectric characteristics of the mine and surrounding soil.

Although a finely-sampled ultra-wideband frequency sweep requires too much computation to be performed in real time on currently available computers, the knowledge gained from studying several typical cases elucidates some of the important feature of wideband mine detection. Inverse Fourier transforming the scattered field frequency response gives a time signature with two characteristic peaks corresponding to burial depth and target size. These particular features are robust and uniquely characterize the target (though not necessarily the surrounding soil) with a single monostatic measurement. With multiple receivers and multiply-incident plane waves, more

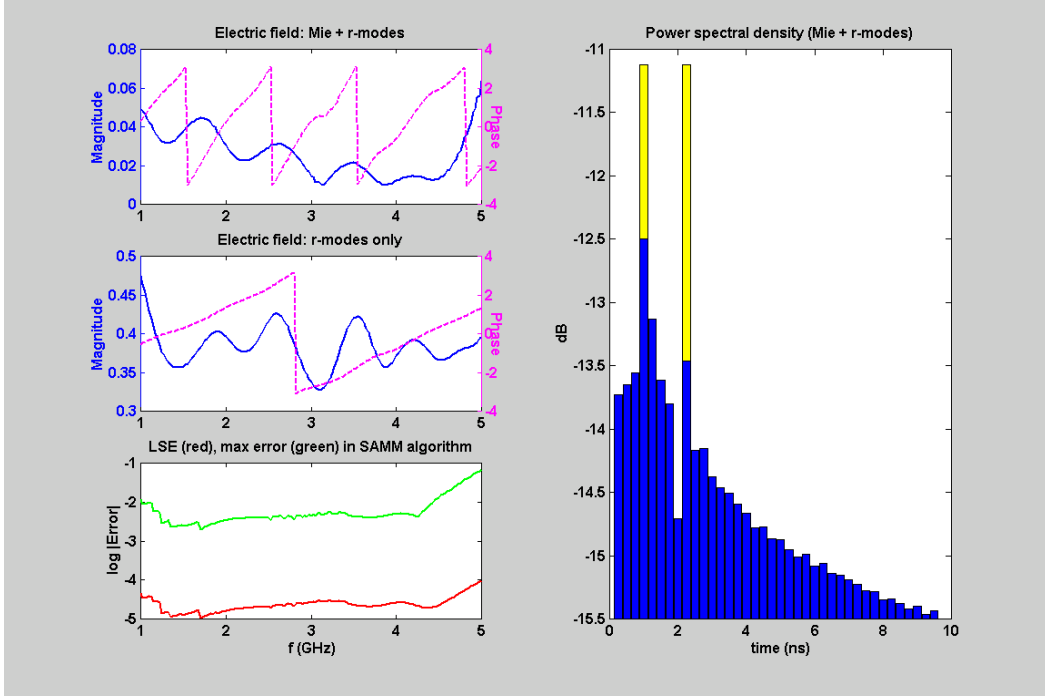


Figure 7: Scattering from a spherical TNT mine buried in Bosnian soil for 1 to 5 GHz in increments of 0.025 GHz, with other parameters equivalent to those of the previous two figures. In this case, the r-modes dominate because of the extreme mismatch between the soil, air, and mine. Errors at low frequency could be improved by making the area over which points are sampled on the plane larger, and errors at high frequency would be improved by using more modes. The IDFT of the total signal has only the two characteristic times T_0 and T_1 ; as in the case of sand, soil mismatch dampens the longer time constants.

independent information would be available to improve both detection and false alarm rates.

In addition, as long as the specular ground reflection is eliminated, these characteristic peaks become apparent regardless of the surrounding soil type, depth of burial, or even relative contrast between mine and ground. The peaks for a TNT mine can be clearly identified even when buried in soil with approximately the same dielectric constant, e.g., sand. The SAMM method is readily extended to oblique plane wave incidence, nonspherical targets, multiple targets and rough ground surfaces by specifying additional modal scattering centers. Work is also progressing on more sophisticated processing of the wideband frequency response and its IDFT beyond the simple identification of the time difference between characteristics peaks.

6. ACKNOWLEDGMENT

This work was supported by The Army Research Office Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013.

7. REFERENCES

- [1] J. Young, M. Poirier, L. Peters, Jr., "A review of current ground-penetrating radar concepts," *IEEE Trans. on Ant. and Propagat. Int'l Symp. Digest*, **3**, pp. 1250–3, 1992.
- [2] C.-C. Chen, "Electromagnetic resonances of immersed dielectric spheres," *IEEE Trans. on Ant. and Propagat.*, **46**:7, pp. 1074–83, 1998.
- [3] C. J. Leat, N. V. Shuley, G. F. Stickley, "Complex image model for ground-penetrating radar antennas," *IEEE Trans. on Ant. and Propagat.*, **46**:10, pp. 1483–7, 1998.
- [4] M. Sadiku, *Numerical Techniques in Electromagnetics*, CRC Press, Boca Raton, pp. 135–204, 1992.
- [5] R. Harrington, *Field Computation by Moment Method*, MacMillan, New York, 1968.
- [6] S. Vitebskiy and L. Carin, "Moment-method modeling of short-pulse scattering from and the resonances of a wire buried inside a lossy, dispersive half-space," *IEEE Trans. on Ant. and Propagat.*, **43**:11, pp. 1303–12, 1995.
- [7] Q. Li, E. J. Rothwell, K.-M. Chen, D. P. Nyquist, "Scattering center analysis of radar targets using fitting scheme and genetic algorithm," *IEEE Trans. on Ant. and Propagat.*, **44**:2, pp. 198–206, 1996.
- [8] C.-C. Chen and L. Peters, Jr. "Buried unexploded ordnance identification via complex natural resonances," *IEEE Trans. on Ant. and Propagat.*, **45**:11, pp. 1645–54, 1997.
- [9] S. Vitebskiy, K. Sturgess, L. Carin, "Short-pulse plane-wave scattering from buried perfectly-conducting bodies of revolution," *IEEE Trans. on Ant. and Propagat.*, **44**:2, pp. 143–51, 1996.
- [10] G. Kristensson and S. Ström, "Electromagnetic scattering from geophysical targets by means of the T-matrix approach: a review of some recent results," *Radio Science*, **17**:5, pp. 903–12, 1982.
- [11] P. Arcioni, M. Bressan, L. Perregrini, "A new boundary integral approach to the determination of the resonant modes of arbitrarily shaped cavities," *IEEE Trans. Microwave Theory and Tech.*, **43**:8, pp. 1848–55, 1995.
- [12] A. W. Morgenthaler and C. M. Rappaport, "Validating the semi-analytic mode matching technique for buried nonmetallic target scattering with two-dimensional FDFD," to be submitted to IGARSS 2000.
- [13] A. W. Morgenthaler and C. M. Rappaport, "Semi-analytic mode matching method for accurate, efficient computation of scattered fields from buried dielectric objects," to be submitted to *IEEE Trans. on Ant. and Propagat.*
- [14] C. M. Rappaport, S. Wu, S. Winton, "FDTD wave propagation in dispersive soil using a single pole conductivity model," *IEEE Trans. on Mag.*, **35**:5, pp. 1542–5, 1999.
- [15] J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill Book Co., Inc., New York, pp. 392–419, 1941.
- [16] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Pub., Inc., New York, 1972.
- [17] W. H. Press, W. T. Vetterling, S. A. Teukolsky, B. P. Flannery, *Numerical Recipes in C, 2nd Ed.*, Cambridge Univ. Press, London, pp. 59–70, 240–54, 496–509, 549–57, 1992.
- [18] D. H. Staelin, A. W. Morgenthaler, J. A. Kong, *Electromagnetic Waves*, Prentice-Hall, New Jersey, pp. 157–160, 1994.
- [19] J. A. Kong, *Electromagnetic Wave Theory*, John Wiley & Sons, New York, pp. 485–89, 1986.
- [20] A. von Hippel, *Dielectric Materials and Applications*, John Wiley & Sons, New York, pp. 314–5, 1953.
- [21] J. Curtis, "Dielectric properties of soils: various sites in Bosnia," *US Army Corp. of Eng., Waterways Experim. Station Data Rep.*, 1996.

Scattering from Dielectric Objects Buried Beneath Random Rough Ground: Validating the Semi-Analytic Mode Matching Algorithm with Two-Dimensional FDFD

Ann W. Morgenthaler and Carey M. Rappaport

Northeastern University, 235 Forsyth Street, Boston, MA 02115

Tel: (617)373-2043, Fax: (617)373-8627, Email: rappaport@neu.edu

Abstract— A 2D finite difference frequency domain (FDFD) algorithm is used to verify new semi-analytic mode matching (SAMM) simulations of scattered fields resulting from plane waves incident on a random rough dielectric half-space containing a buried dielectric target. The SAMM algorithm uses moderately low-order modal superpositions of cylindrical waves, each of which satisfies the 2D-Helmholtz equation in its appropriate region (air, ground, or mine) and then matches all nonzero electric and magnetic field components at each interface by least squares fitting.

For smooth ground, coordinate scattering centers (CSCs) are chosen at the mine center and at its image above the plane to model scattering. For random rough ground, additional CSCs are located within the rough layer. Excellent agreement between 2D-FDFD and the two dimensional version of SAMM is observed, with 2D-SAMM being at least an order of magnitude faster; 3D-SAMM is estimated to be four orders of magnitude faster than 3D-FDFD, with drastically reduced memory requirements.

Keywords— GPR, mine detection, dielectric target imaging, modal analysis.

I. INTRODUCTION

A NEW 3D semi-analytic mode matching (SAMM) algorithm has been described previously [1], which is particularly adept at finding the scattered near fields from buried dielectric objects (e.g. antipersonnel land mines). Here, we validate the 2D version of SAMM for scattering from a 2D buried cylinder by comparing it to a 2D finite difference frequency domain (FDFD) simulation. We also extend SAMM to cases where the ground has a random rough surface, a topic of considerable interest [2,3]. We consider only normally-incident TM plane waves scattered from buried mines with circular cross-section, though oblique incidence and modestly irregular mine shapes are easily handled by SAMM. Three-dimensional rough surface scattering is more computationally difficult both for SAMM and FDFD methods, so we focus here on the simpler 2D geometry from which useful physical insights can also be extracted.

II. SEMI-ANALYTIC MODE MATCHING

For 2D cylindrical geometry having no variation along z , the TM electric and magnetic fields ($H_z = 0$) may be described by the vector Hertz potential $\mathbf{\Pi} = \hat{\mathbf{z}}\Pi_z(\rho, \phi)$ obeying the scalar wave equation $(\nabla^2 + k^2)\Pi_z = 0$, where

∇^2 is the 2D cylindrical Laplacian and k is the wave number in the medium of interest [4]. A modal expansion of Π_z is given by:

$$\Pi_z(\rho, \phi; k) = 1/k^2 \lim_{N_{\max} \rightarrow \infty} \sum_{n=-N_{\max}}^{N_{\max}} C_n F_n(k\rho) \exp(-in\phi) \quad (1)$$

where C_n is the cylindrical mode coefficient (units of electric field) and $F_n(kr)$ is a Bessel function of order n . With judicious use of the Bessel recurrence relations, we can find the Cartesian field components, *each of which must also satisfy the scalar wave equation*, in terms of the cylindrical modes of (1):

$$E_z = \lim_{N_{\max} \rightarrow \infty} \sum_{n=-N_{\max}}^{N_{\max}} C_n F_n(k\rho) e^{-in\phi} \quad (2)$$

$$H_x = \lim_{N_{\max} \rightarrow \infty} \sum_{n=-N_{\max}}^{N_{\max}} \frac{(C_{n-1} + C_{n+1})}{2i\eta} F_n(k\rho) e^{-in\phi} \quad (3)$$

$$H_y = \lim_{N_{\max} \rightarrow \infty} \sum_{n=-N_{\max}}^{N_{\max}} \frac{(C_{n-1} - C_{n+1})}{2i\eta} F_n(k\rho) e^{-in\phi} \quad (4)$$

where η is the wave impedance. To be useful numerically, (2)–(4) must be truncated at $n = N_{\max}$, a finite (small) integer. For an error of -290 dB (about machine accuracy) at 3 GHz, plane wave decomposition in dry sand [5] ($\epsilon_g = 2.55(1 + i0.01)\epsilon_0$) and Bosnian soil from the test site Alicia [6] ($\epsilon_g = 8.82(1 + i0.13)\epsilon_0$) requires 24 and 32 modes, respectively.

Equations (2)–(4) are applied to each region (air, ground, mine) where we expand the modes with respect to convenient coordinate axes. We locate these coordinate scattering centers (CSCs) so modal expansions will converge for small values of n ; in general, we place CSCs everywhere waves appear to originate.

Fig. 1 shows the geometry of a cylindrical mine with circular cross-section of radius R buried a depth d below the nominal surface of a random rough ground plane. Because there is no variation along z , the rough surface contour is actually a series of randomly-spaced, rounded grooves extending infinitely far in the $\pm z$ direction. The figure also indicates schematically how modes are defined in each region. For geometries which include an infinitely-extended ground plane, CSCs are chosen at *both* the mine and its image above the plane, in analogy with

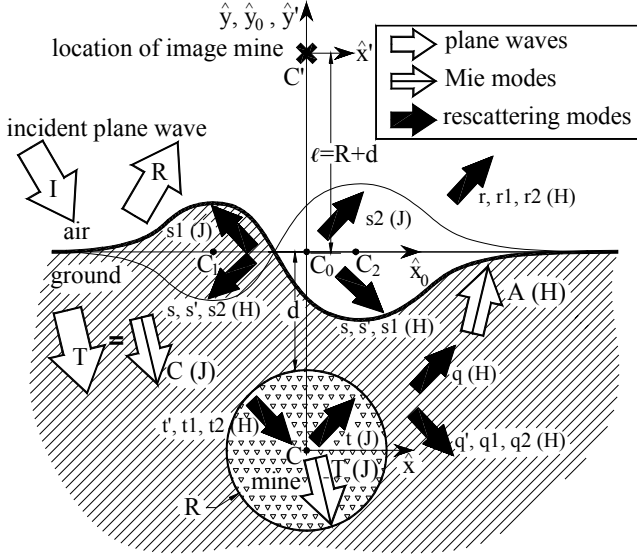


Fig. 1. A dielectric cylinder with radius R is buried d meters below a rough ground surface. This particular randomly-generated surface has a ground-filled bump on the left and an air-filled pit on the right. Modes are given schematically in each of the regions where they apply, and are designated “J” or “H” depending on which Bessel function is necessary. Modes with no sub- or superscripts originate at the mine (C -coordinate system), while the primed modes originate at the image mine (origin at C'). Modes with 1 or 2 as subscripts originate at the CSCs C_1 and C_2 (there may be more surface CSCs if the rough surface has a small correlation length, and none if the ground is smooth). The incident plane wave (I) is reflected (R) and transmitted (T); this latter wave is decomposed into cylindrical modes (C), which lead to Mie scattering modes (A , T) [see text]. These Mie modes reflect and refract at the ground plane, leading to rescattering or r-modes, valid in their respective domains as shown. For the simpler case of a flat planar boundary, we eliminate the scattering centers C_1 and C_2 , and only the r-modes $\{r, q, q', t, t'\}$ remain.

image techniques routinely used for analyzing antennas over perfectly-conducting ground planes. For a smooth air/ground interface, this single image is sufficient to describe the scattering. For a rough surface, the bumps and depressions on the surface will each act as scattering centers, particularly in the case where the size of the defects is comparable to the size of the mine. In this case, we add CSCs at each place on the plane surface where rescattering is expected with the following procedure: (1) create a reflection of the rough surface about the nominal smooth ground surface making linked “bubbles” containing either air or ground, (2) put CSCs along this nominal plane in the center of each defect, and (3) choose modes within and without the “bubbles” which behave appropriately; i.e., use Bessel functions of the first kind $J_n(k\rho)$ for regions which include the CSC or to expand a plane wave, and use outwardly-propagating Hankel functions $H_n^{(1)}(k\rho)$ for regions which do not include the CSC and which must obey radiation conditions far from the scatterer.

Analytic (modal) Mie solutions for a dielectric mine of circular cross-section buried in infinite ground [4] describe unperturbed scattering with the coefficients A_n and T_n shown in Fig. 1. These Mie modes are then perturbed by the actual half-space ground with its random rough surface, leading to rescattering (r-modes) which come from all the multiple interactions between the mine and plane. We truncate r-modes originating at the mine, its image, and the rough surface at N_m , N_m , and N_s , respectively. The coefficients of these r-modes are found by locating N_{surface} points along the air/ground and mine/ground interfaces and reconstructing E_z , H_x , and H_y at each point using (2)–(4). The mode coefficients that best fit the three boundary conditions (one for each nonzero field component) at each of the N_{surface} sampled surface points are found using least squares techniques. We may test the final solution in two ways: (1) by demonstrating convergence of the nonzero fields to stable values as more modes are chosen, and (2) by comparing fields to those generated by established methods like FDFD or the moment method for identical geometries.

III. RESULTS AND CONCLUSIONS

In all simulations, a 3 GHz TM plane wave was normally incident on a 5-cm radius TNT mine having circular cross-section and buried 5 cm below the nominal ground surface of dry sand. For the SAMM algorithm, $N_{\text{max}} = 25$; the 35-cm planar surface and mine surface were both sampled with 600 points. FDFD simulations had a grid spacing of 1/12 cm corresponding to 441×420 grid points. For a smooth surface, the SAMM algorithm performs exceptionally well, exceeding the accuracy of the FDFD simulation after $N_m = 6$, as shown in column 3 of Table I. Column 4 compares the SAMM simulations to the most accurate case shown in the table: SAMM($N_{\text{max}} = 12$). SAMM(6) is about 16 times more accurate than the FDFD simulation, which has about 0.126% average error (attributable to discretization). The SAMM simulation took 3 [SAMM(1)] to 6 [SAMM(12)] min. on a Pentium-based personal computer, while the FDFD simulation took 45 min., almost an order of magnitude longer, and required the huge memory allocation of 0.93 million nonzero sparse matrix elements. By contrast, SAMM(6) required singular value decomposition of a 3600×65 element dense matrix.

The rough surface used for the SAMM and FDFD simulations is given in Fig. 2, and the magnitude of E_z is plotted in Fig. 3 for both the SAMM($N_m = 8$, $N_s = 4$) and FDFD simulations. The average error is 1.66% between methods, lies primarily near the rough surface, and is mainly attributable to SAMM. Table II shows how the average error varies with the mine/image modes number N_m and rough surface mode number N_s . As desired, excellent results are obtained by truncating r-modes at

TABLE I

COMPARISON OF SAMM AND FDFD FOR SMOOTH GROUND SURFACE

The average least squares error (LSE) on the smooth plane is computed for SAMM simulations as a function of r-mode truncation integer N_m . In addition, the SAMM simulations are compared both with a finely-gridded FDFD simulation and with the best SAMM simulation in the table, SAMM(12).

Smooth Surface, Average Error			
N_m	LSE at plane	SAMM(N_m) vs. FDFD	SAMM(N_m) vs. SAMM(12)
1	2.05e-03	0.00554	5.21e-03
2	1.49e-03	0.00379	3.38e-03
4	4.07e-04	0.00131	4.90e-04
6	7.78e-05	0.00127	7.62e-05
8	1.33e-05	0.00126	1.87e-05
10	2.16e-06	0.00126	7.99e-06
12	3.91e-07	0.00126	—

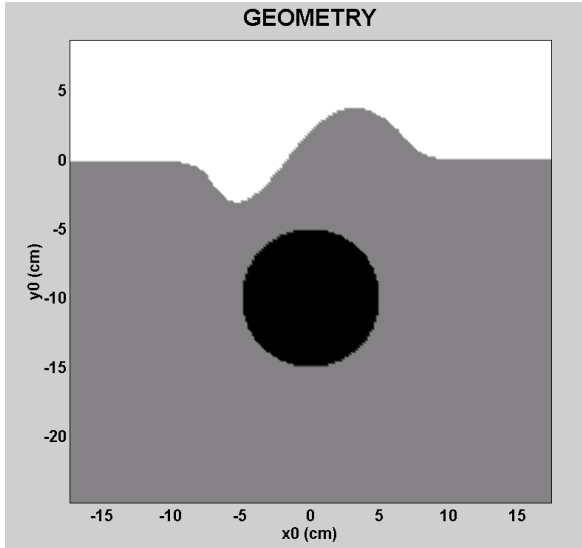


Fig. 2. Geometry of the mine and random rough surface. The rough surface is generated by creating a coarse grid with spacing equal to 7 cm, the correlation length. Each grid point is randomly assigned a height from a normal distribution having a 1 cm rms surface height. Surface points are found by spline interpolation to the coarse surface grid.

many fewer than the $N_{\max} = 25$ modes required for accurate plane wave expansion. Increasing N_m appears to affect the error more sharply than increasing N_s , leading to the conclusion that fewer scattering modes need be kept from CSCs placed at the rough surface. Since these are the modes that add combinatorially, we expect that 3D-SAMM algorithms will prove quite tractable.

The 3D-SAMM would require about 6 times as many modes and 30 times as many surface points, corresponding to 200 times the memory and 1000 times the CPU time of

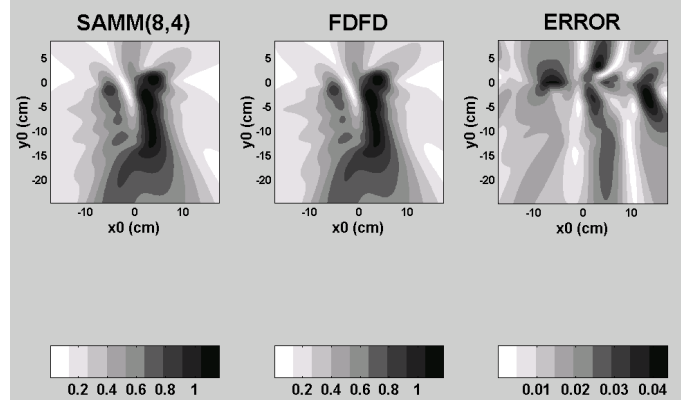


Fig. 3. $|E_z|$ plotted in the x - y plane for both SAMM(8,4) and FDFD using the rough surface of Fig. 2. The error is less than 4% and averages 1.66%.

TABLE II

COMPARISON OF SAMM AND FDFD FOR RANDOM ROUGH SURFACE

SAMM simulations having N_m modes for the CSCs centered at both the mine and image mine and N_s modes for each CSC centered within the rough surface are compared with a finely-gridded FDFD simulation.

Rough Surface, Average Error					
$N_m \backslash N_s$	0	1	2	4	8
1	0.192	0.121	0.0953	0.0830	0.0811
2	0.147	0.0850	0.0797	0.0768	0.0764
4	0.0672	0.0257	0.0232	0.0222	0.0224
8	0.0234	0.0171	0.0169	0.0166	0.0169

2D-SAMM; by contrast, 3D-FDFD needs more than 1000 times the storage and 10^6 times the CPU requirements of 2D-FDFD. Although SAMM is more difficult to program than FDFD and cannot currently model fine-scale roughness, the favorable scaling from 2D to 3D in the SAMM algorithm makes this new method worth pursuing.

REFERENCES

- [1] A. W. Morgenthaler and C. M. Rappaport, "Semi-analytic mode matching for detecting nonmetallic mines buried in realistic soils," *Proc. of SPIE Aerosense 2000*, in press.
- [2] K. O'Neill, R. F. Laskey, Jr., K. D. Paulsen, "Scattering from a metallic object embedded near the randomly rough surface of a lossy dielectric," *IEEE Trans. on Geosci. and Remote Sens.*, **24**(2), pp. 367–76, 1996.
- [3] J. R. Arias-González, M. Nieto-Vesperinas, A. Madrazo, "Morphology-dependent resonances in the scattering of electromagnetic waves from an object buried beneath a plane or random rough surface," *J. Opt. Soc. Am. A*, **16**(12), pp. 2928–34, 1999.
- [4] J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill Book Co., Inc., New York, pp. 392–419, 1941.
- [5] A. von Hippel, *Dielectric Materials and Applications*, John Wiley & Sons, New York, pp. 314–5, 1953.
- [6] J. Curtis, "Dielectric properties of soils: various sites in Bosnia," *US Army Corp. of Eng., Waterways Experim. Station Data Rep.*, 1996.

Mitigating Ground Clutter Effects with Lightweight Artificial Dielectrics

Carey Rappaport, Fred Biehold, and Robert Linnehan

Center for Subsurface Sensing and Imaging Systems

Northeastern University

Boston, MA 02115

Fax: (617) 373-8627

Tel: (617) 373-2043

Email: rappaport@cer.neu.edu

Abstract:

The problem of scattered and transmitted electromagnetic wave distortion by random rough ground surfaces can be reduced by using a lightweight dielectric matching layer. For mine detection applications, it is essential for this layer to be lightweight, low loss, readily conformable, and adaptable to different soil types.

Arrays of metal-coated plastic spheres act as lossless artificial dielectrics with impedance determined by the volume packing fraction. By controlling the thickness of insulator surrounding each sphere, a close-packed array with the dielectric properties of soil can be created inside a compliant rolling bag that will conform to the rough surface of the ground. Since this artificial dielectric is matched to the soil, the ground surface interface is “softened”, without an abrupt transition from soil to air. Signals transmitted and received by GPR antennas immersed in the artificial dielectric within the bag will not be corrupted by ground surface clutter. Alternatively, an artificial dielectric layer on the ground with a planar air interface could be used to ensure that the surface reflection is a constant, well-calibrated signal.

Computational models indicate complete removal of the ground clutter, even with occasional gaps between the artificial dielectric and the ground. Experimental studies with swept-frequency measurements and impulse GPR indicate that using this dielectric layer matching to a rough loamy soil ground surface results in signals that are practically indistinguishable from those of an equivalent layer of the same type of soil.

Keywords: mine detection, artificial dielectrics, ground-penetrating radar, coupling, impedance matching

Introduction

A major challenge of radar detection of near-surface buried objects—such as land mines—is to distinguish the signals scattered from targets as opposed to the ground surface. In the particular case of nonmetallic mines, the contrast between the mine and the surrounding soil is often lower than between the soil and the air above it; and the clutter signal from a rough interface is often much stronger than the target signal from the mine. For more deeply buried targets, it is possible to reduce this clutter by time gating a short duration radar signal. However for targets buried at depths comparable to the height variation of the rough surface, this gating is not feasible. While it is possible to reduce the ground scattering clutter by attempting to characterize it using statistical methods [1-3], or illuminate the ground with an inclined plane wave and measure the backscattered signal [4,5], a more direct way to is to directly eliminate the surface roughness with an artificial matching layer.

Artificial Matching Layer

The purpose of an artificial layer is to fill the voids and lessen the effects of bumps in the ground surface with a medium that electromagnetically resembles the soil. To reduce clutter, it is more important to “soften” and “planarize” the interface than to act as an impedance transformer between the soil and the air. If the effective interface at the top of the layer is planar, and its impedance is known, the reflection of waves from the interface are easy to account for, and their contribution to clutter can be effectively subtracted. For example, this reflection from the layer top could be measured as part of the radar calibration procedure, or calculated numerically. Alternatively, the entire antenna could be immersed in a compliant bag filled with the artificial dielectric material, so that there would be no air interface.

Figure 1 schematically shows how this configuration might look, with a thin, but tough plastic bag filled with a lightweight material that simulates soil. It is obviously important to ensure that the material be light enough to avoid triggering mines, and that the material be loose enough to reconfigure itself as the bag rolls along the ground.

Conducting spheres that are spatially separated have been used as artificial dielectrics (AD) for many years [6]. However, solid metal balls are too heavy for this use, and separating them in a fixed matrix is clearly infeasible. Instead metal-coated plastic balls could be used. These lightweight “beads” are available as decorative craft items. An additional benefit is that these beads are lacquered or plastic coated with an insulating layer that serves to isolate them for one another.

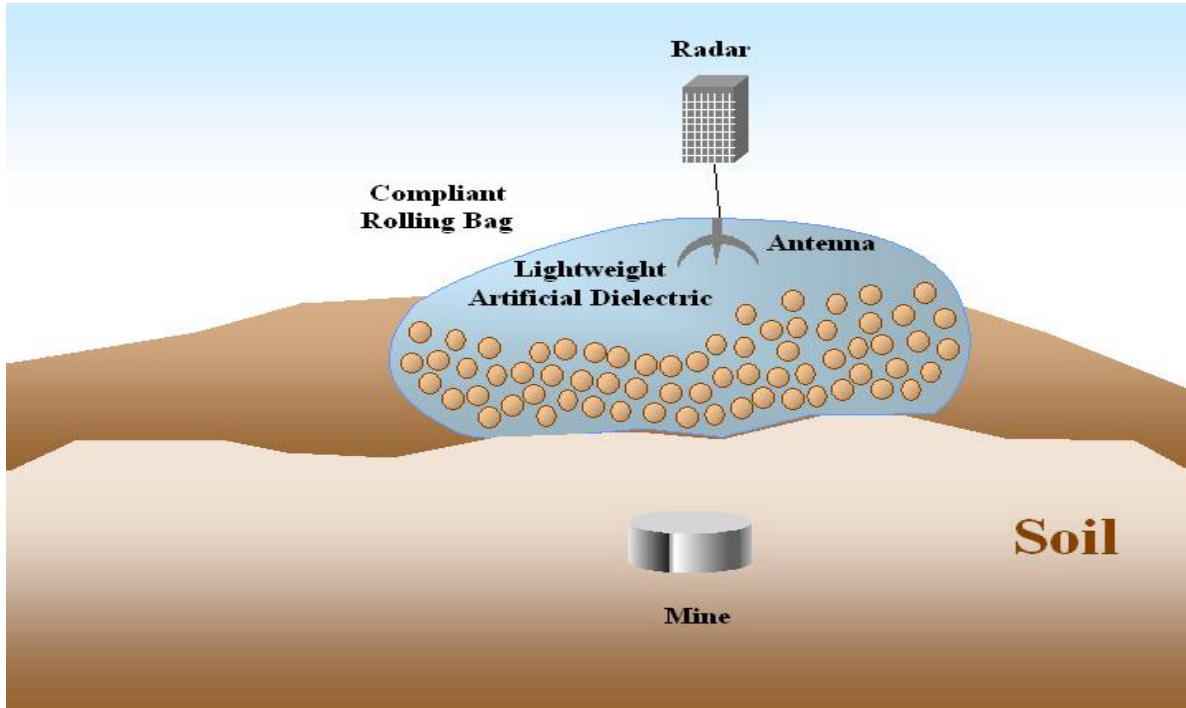


Figure 1: Geometry of the artificial dielectric matching layer, used to reduce the clutter effects of random rough ground for GPR mine detection.

Simulation of Imperfect Void Filling

One concern with using an array of discrete objects in a bag as an artificial layer is the possibility of occasional gaps between the soil and the matching medium. The extent of this imperfect filling can be predicted numerically.

Using the FDTD method for dispersive soil [7], we computed the distorting effects of occasional air voids (depicted in Figure 2). A modulated TM Gaussian pulse, originating at a point 35 cm above the rough surface of Puerto Rican clay loam with a metallic target buried 10 cm below the nominal surface, was modeled with and without the AD.

Figure 3 shows the total field at the same time instances for the three comparative cases. At left, the moderately rough ground surface, with ± 5 cm height variation causes enough

wave scattering to completely obscure the buried target. The incident cylindrical wave and the transmitted wave are visible, as is the lack of field within the mine. Even for this strong scatterer, the ground clutter dominates. The second and third plots show the reduction in clutter when an AD is used. With no air voids (center), the incident field exists solely in the AD and soil. The interface is,

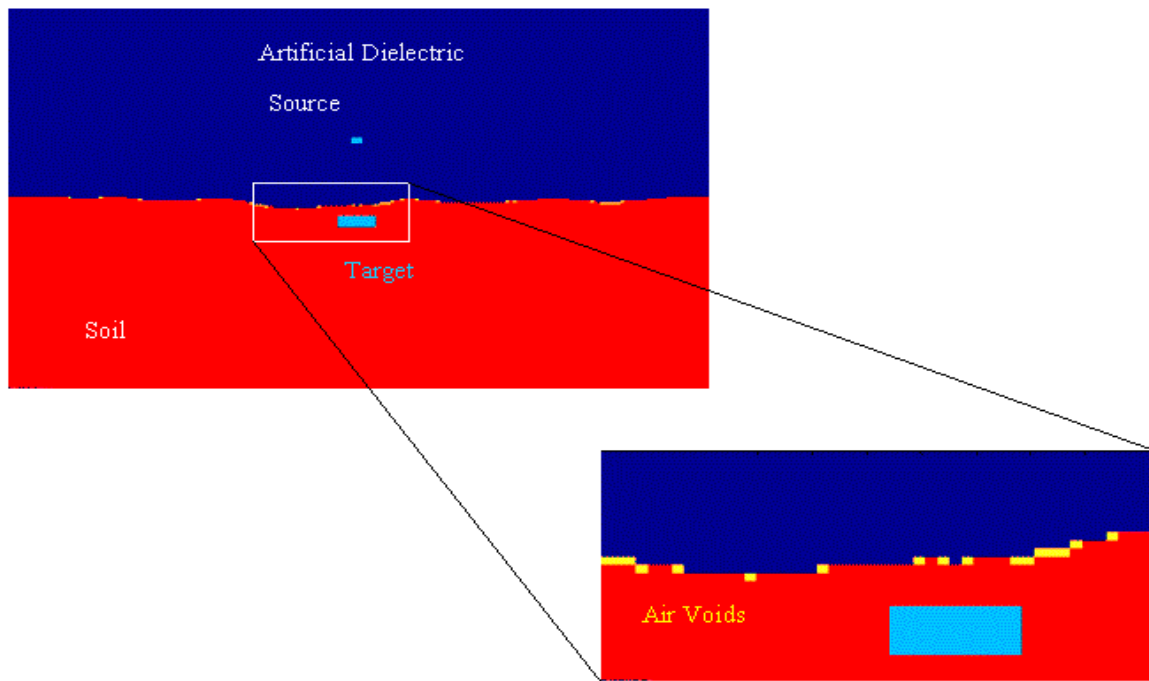


Figure 2: Modeled geometry with air, dispersive clay soil, and occasional air voids

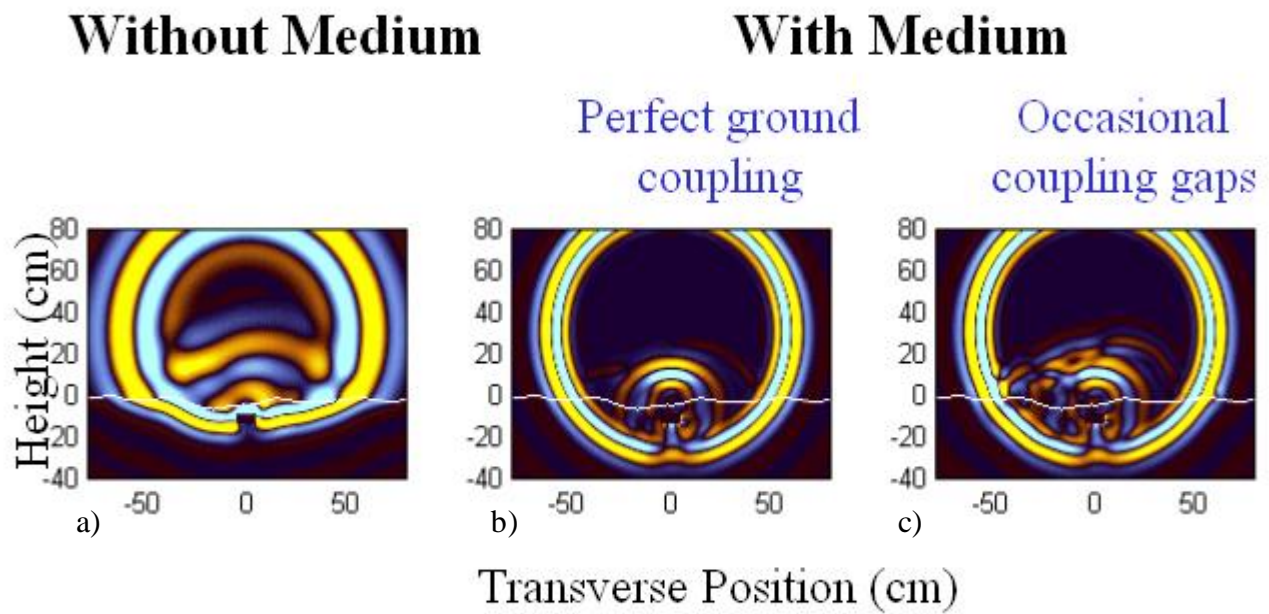


Figure 3: Comparison of modeled field a) without AD matching layer, b) with matching layer perfectly conforming to ground surface, and c) with realistic layer with air gaps

for all practical purposes, undetectable. The wave scattered by the target is clearly visible as the second smaller radius circle centered at the target. Compared to the first case, it is quite straightforward to identify which signal originates at the mine. In the third plot, the gaps of Figure 2 result in some clutter from the imperfectly matched interface. However, these scattered fields are small compared to those of the mine target.

Artificial Dielectric Impedance Calculation

The effective impedance of an array of conducting spheres is dependent on the geometry of the array, and the proximity of the spheres. The sphere separation can be computed in terms of the volume packing fraction of spheres in space. For simple cubic (sc) packing, with a sphere of radius r on each corner of a unit cube, the volume occupied by the eight sphere octants that lie within the cube is $4/3\pi r^3$, which is also the packing fraction. The maximum packing fraction occurs when $r = 1/2$, with a value of about 0.52. For a face centered cubic (fcc) packing, in which there are spheres centered on the cube corners and the centers of the cube faces, the volume (and hence packing fraction) within the six hemispheres and eight octants that lie within the cube is $16/3\pi r^3$. The maximum sphere radius in this case is one-quarter of a face diagonal, or $\sqrt{2}/4$, leading to a maximum packing fraction of 0.74.

The calculated impedance of an array of conducting spheres, as a function of packing fraction for sc and fcc can be approximated by the formulas [6]:

$$\left. \begin{array}{l} \epsilon/\epsilon_0 \\ \mu/\mu_0 \end{array} \right\} = 1 + 3p/A(p)$$

where

$$\begin{aligned} \text{For sc: } A(p) = & -1/R_1 - p + 1.3047R_3p^{10/3} + 0.0723R_5p^{14/3} \\ & - 0.5289R_3^2p^{17/3} + 0.1526R_7p^6 \end{aligned}$$

$$\begin{aligned} \text{For fcc: } A(p) = & -1/R_1 - p + 0.0753R_3p^{10/3} + 0.2420R_5p^{14/3} \\ & + 0.0558R_3^2p^{17/3} + 0.0231R_7p^6 \end{aligned}$$

and

$$R_n = \begin{cases} -1 & \text{for effective dielectric constant} \\ n/(1+n) & \text{for effective magnetic permeability} \end{cases}$$

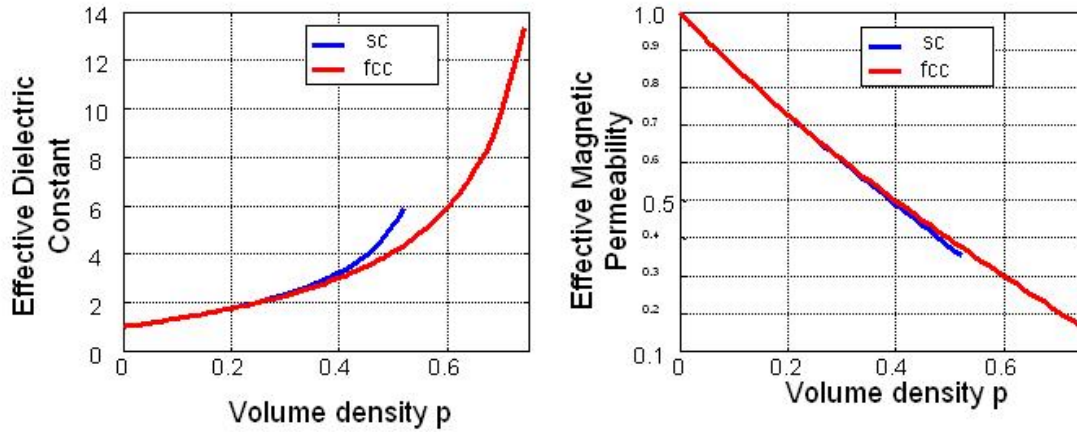


Figure 4: Dependence of effective dielectric constant and magnetic permeability of AD in terms of volume packing fraction p .

These formulas are plotted in Figure 4, showing that the sc packing has slightly greater effective dielectric constant and lower permeability than that of the fcc packing, but that a higher packing fraction is possible with fcc. In practice, we would expect that random distributions of spheres would fall between the two sets of curves. For a dense random packing of thinly insulated spherical beads, we might expect a relative impedance as high as $\sqrt{(0.25/10)}$, or about 0.16 times the impedance of free space. This matches a dielectric constant of 39 in a non-magnetic ($\mu = \mu_0$) medium, much greater than most soils.

As most mine detection systems operate over several frequencies at once, it is important that the AD layer match over a wide frequency band. Figure 5 shows the reflection coefficient as a function of

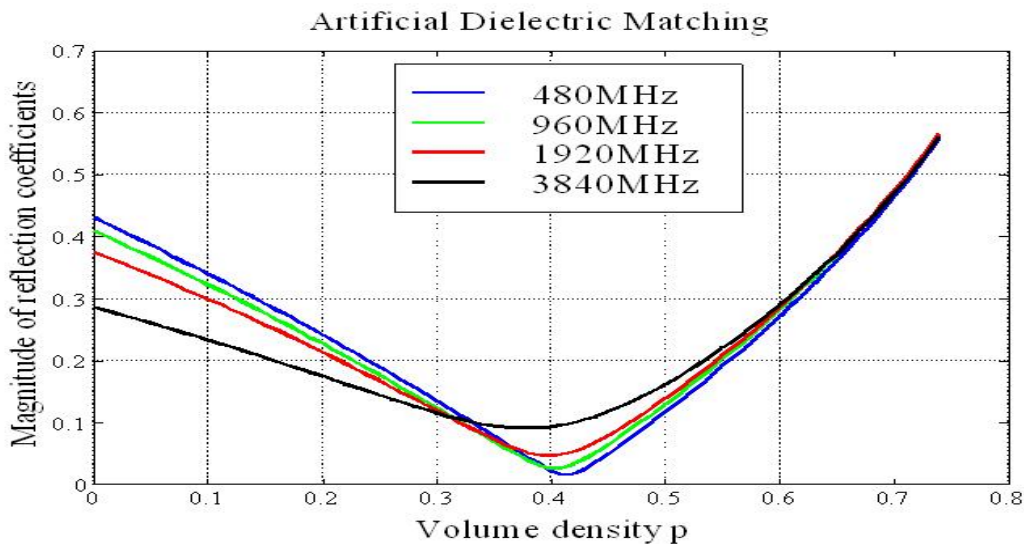


Figure 5: Computed reflection coefficient for AD/soil interface with normal incidence for four discrete frequencies, as a function of packing fraction p . The soil is Puerto Rican clay loam with 10% moisture and 1.6 g/cc density.

packing fraction for various frequencies for a planar interface between AD and frequency-dependent lossy soil. As long as the wavelength is large compared to the diameter of spherical conductors in the AD array, its bulk dielectric constant is independent of frequency. In addition, the AD is lossless, so the reflection coefficient increases as the loss tangent increases. Figure 5 indicates that even for fairly dispersive soil, a good wide-band impedance match is possible.

Considering a typical range of soil dielectric constants from 2.5 (dry sand) to 20 (20% moist clay loam), it is feasible to match the impedance using one of as few as three or 4 well-chosen ADs. Dividing the range into three dielectric regions centered at constants 3.33, 7.3, and 16.0 would introduce a maximum normally incident reflection coefficient of 0.098. For the four-region division with dielectric constants 2.98, 5.22, 9.13, and 16.0, the maximum reflection coefficient is 0.070. The range of dielectric constants covered in each region is shown in Figure 6. The boundaries of each of these regions are determined by equating the ratio of the central value to the boundary value (or its inverse).

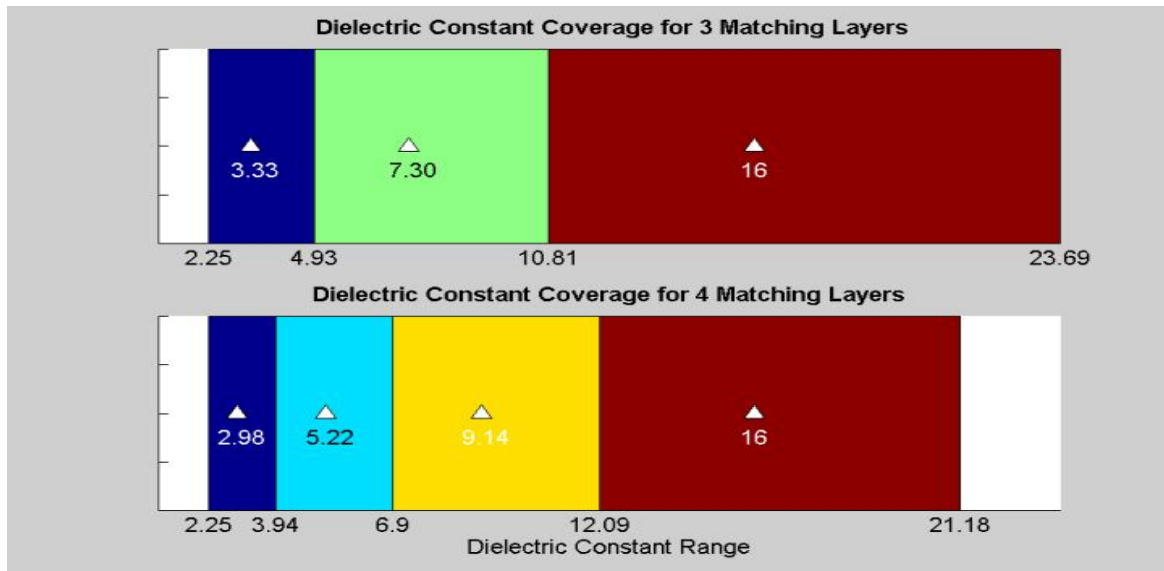


Figure 6: Segmenting the typical range of soil dielectric constants into three or four regions with the same maximum reflection coefficient: 0.098 (top), and 0.070 (bottom).

Experimental Realization of Artificial Dielectric

Using just freely flowing insulated conducting craft beads for a matching layer would produce the maximum effective AD dielectric constant. To match to a particular soil requires increasing the spacing between the beads. Alternatively, it has been found that by introducing small air-filled plastic balls to the beads reduces the volume packing fraction, albeit non-uniformly. As long as the balls are fairly uniformly distributed amongst the beads, and their diameters are small compared to the smallest

wavelength, they can be viewed as an array of lower dielectric constant spheres in a higher dielectric constant background: sort of a reverse artificial dielectric. The net effect is to reduce the overall bulk dielectric constant.

Several sizes of polypropylene balls were used to adjust the packing fraction and effective dielectric constant: 20, 10, and 6 mm diameter. Each AD mixture of beads and balls was tested in a cylindrical cavity by comparing its S-parameters to those of soil. In order to prevent direct interaction with the coaxial probe inserted into the cavity, the bottom portion was filled with paraffin. Although this complicates the direct wideband measurement of the dielectric constant of the material under test, it does provide for direct impedance comparisons across the entire frequency band of interest.

Relatively good performance from 700 MHz to 1.7 GHz was obtained using several mixture combinations. The S_{11} response is presented in Figure 7 for various mixtures compared to 15% moist clay loam. Also shown is the response for dry sand, which is quite different.

The S_{21} measurements for a transmitter/receiver probe pair inserted in a soil-filled tank with the AD or soil filling an intervening region gave similar well-matched results.

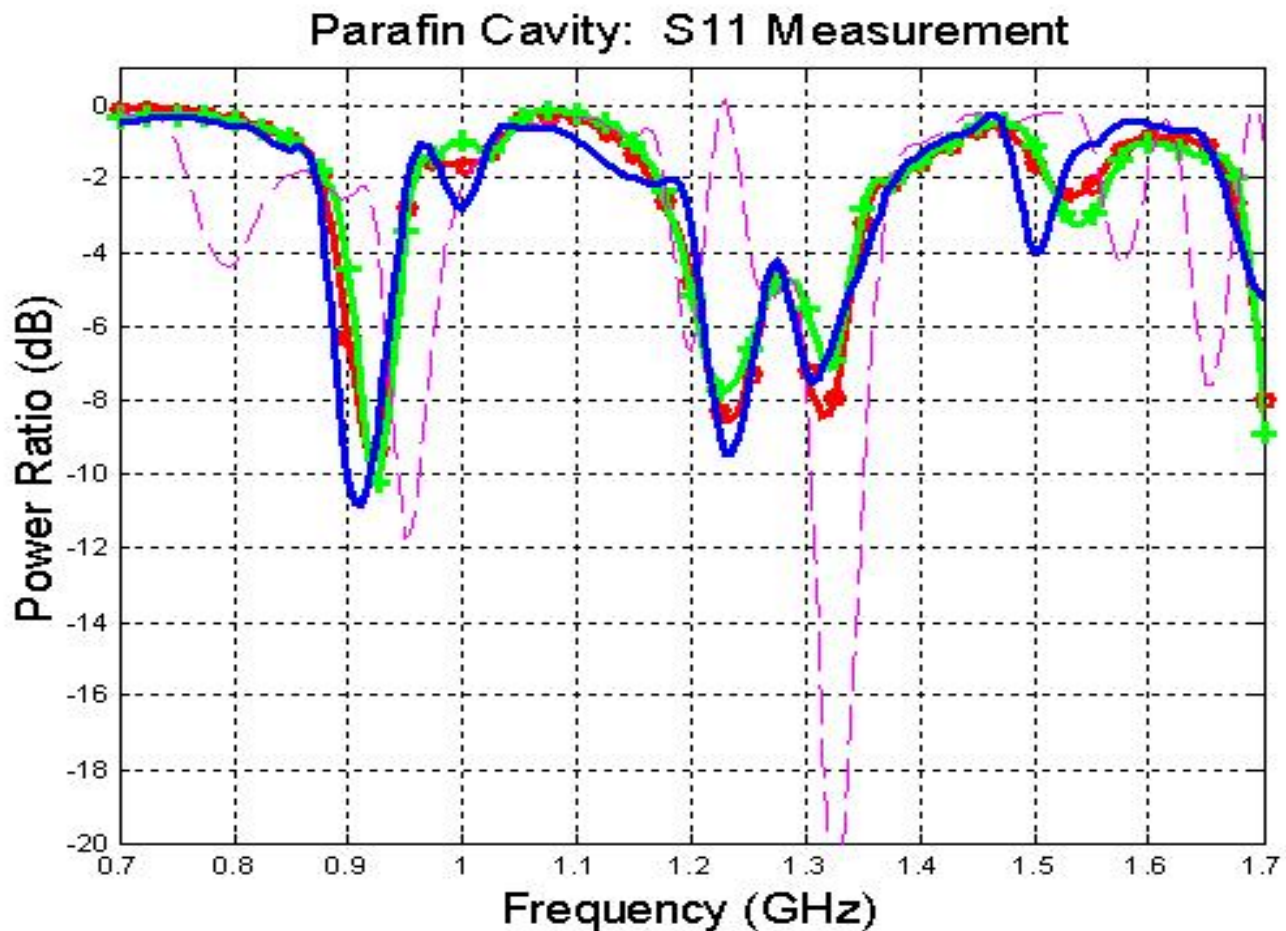


Figure 7: S_{11} measurements of a cylindrical cavity partially filled with: soil (dark solid line), AD with 20 mm balls (o), 10mm balls (+), or sand (--).

Impulse Radar Experiments using AD Matching Layer

With a good impedance match over the frequency range of an impulse radar, the final step is to compare the time domain signals scattered from smooth, roughened, and AD matched rough soil. For this experiment, the impulse radar used was the Geo-Centers, Inc. EFGPR with TEMR antennas [8]. Two antennas—one acting as the transmitter, and the other the receiver—were positioned next to one another roughly 2 meters away from the surface of the tank of sifted clay loam soil. A 900 ps single pulse was fed to the transmitter, and the waves scattered from the soil surface was measured by the receiver for various cases.

First, the soil surface was raked as flat as possible, providing a planar half-space interface. The scattered field was measured. Next, soil was removed down to about 3 inches in several spots (corresponding to a standard deviation of about 1 in. relative to the mean level) and set aside. The rough surface signal was measured. The depressions were next filled with the AD mixture of beads and 20 mm polypropylene balls in the volume ratio of 1.73 to 1, and the top surface smoothed. A photograph of this configuration appears in Figure 8. The field scattered from this smoothed surface was measured, and the AD was twice removed, replaced, and the signal remeasured. Finally, the AD was completely removed (facilitated by using the plastic sheet seen in Figure 8), and the original removed soil was replaced and smoothed, and the scattered signal measured.



Figure 8: Soil test tank, with rough surface “smoothed with AD composed of 6mm insulated conducting beads and 20 mm balls, in a volume ratio of 1.73 to 1.

The measured results for the six test cases are presented in Figure 9. Five of the six traces appear to have almost the same behavior, with the greatest variation at late times, 3.5 ns after the first arrival. The sixth trace, corresponding to the rough soil surface without AD is significantly different, with as much as 0.3 ns delay relative to the flat soil surface. Although this delay may not seem like much, it can cause significant error in the processing of the radar signal. For example, at time $7.98\text{E-}8$, the rough surface clutter is negative, while the flat surface signal is positive with almost the same magnitude. Any clutter reduction using the wrong assumption of the ground surface would introduce 100% more error at this time point!

It is interesting to note that the time shift between the rough and smooth ground signals is not constant, indicating that there are multiple scattering effects occurring with this rough ground surface. This is why attempts at identifying and subtracting the ground reflected signal as a shifted copy of the flat ground signal does not always reduce ground clutter [2].

As is evident from Figure 9, the various AD placements repeatably form a well-matched layer to the rough surface soil. Although the response from the soil with the AD layer is not exactly the same as the original flat surface soil, its variations are comparable to the case of the original soil being replaced and resmoothed.

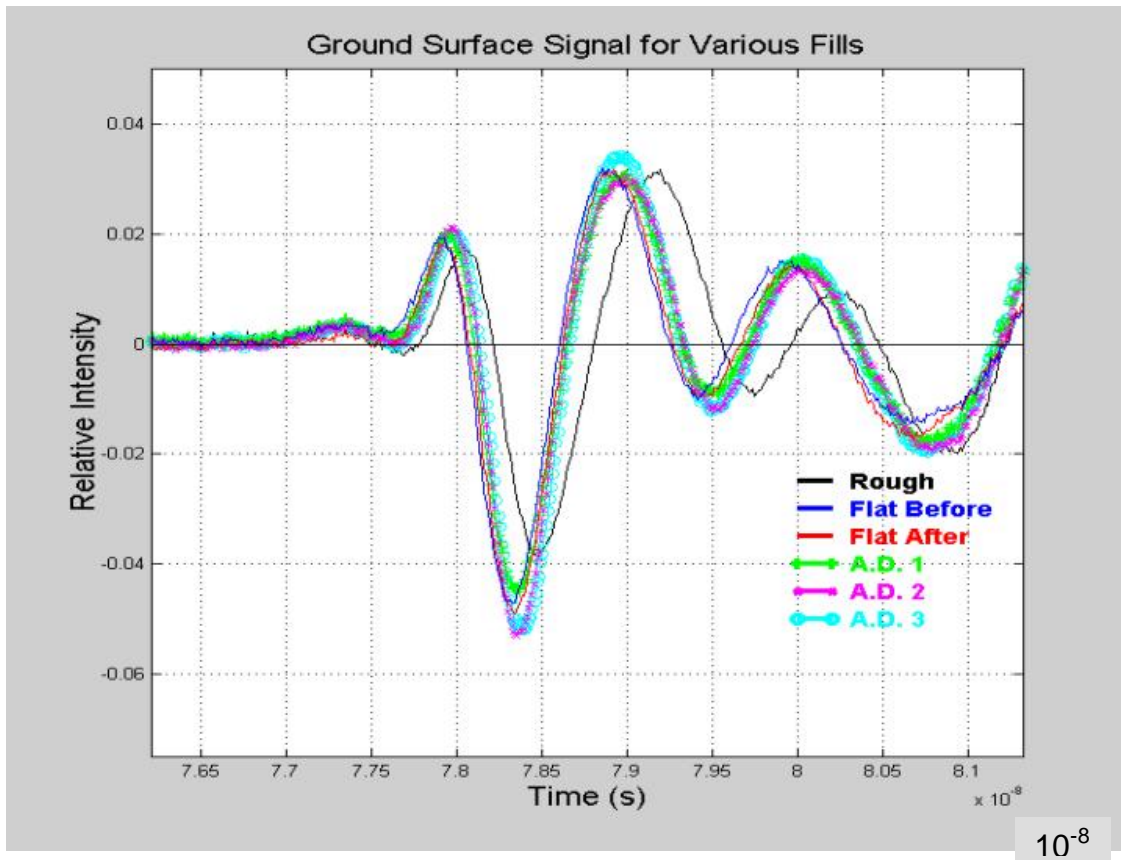


Figure 9: Time domain measurements of soil with flat surface, rough surface, and with added AD composed of beads and 20 mm balls in volume ratio 1.73 to 1.

Conclusions

An artificial dielectric, which maintains the approximate impedance characteristics of soil over a wide frequency range, has been developed and experimentally tested. Being flexible, lightweight, and with adjustable dielectric constant, this AD composed of insulated, metallized craft beads can be used in mine detection applications where it is important for this layer to conform to ground surface variations while exerting minimal pressure on the ground.

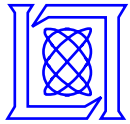
The AD can be used as a rough ground surface smoothing layer with a planar top surface, or as a medium to fill the entire space between the antenna and the ground. As a smoothing layer, the AD can effectively replace the random interface roughness that corrupts the GPR signal with hard to characterize clutter with an unchanging planar interface the effects of which can be easily identified and eliminated. Computational modeling has shown that even with imperfect filling of rough surfaces, the clutter reduction is significant.

The effective impedance of the artificial dielectric is easily controlled by adjusting the volume ratio of beads to non-conducting polypropylene balls. Only three or four types of bead/ball AD mixtures are needed to match the entire range of most soil dielectric constants with reflection coefficients less than 10%.

Experiments with fielded impulse GPRs indicate that the AD layer matches to rough soil surfaces about as well as the same soil does.

References

1. Yang, B. and Rappaport, C., "Response of Realistic Soil for GPR Applications with Two Dimensional FDTD," accepted for publication in *IEEE Trans. on Geoscience and Remote Sensing*, June 2000.
2. El-Shenawee, M. and Rappaport, C., "Quantifying the Effects of Different Rough Surface Statistics for Mine Detection Using the FDTD Technique," *SPIE Aerosense Conference*, Orlando, FL, April 2000, pp. 966--975.
3. Rappaport, C., El-Shenawee, M. and Zhan, H., "Suppressing GPR Signal Degradation from Randomly Rough Ground Surfaces," in review for publication in *IEEE Trans. on Geoscience and Remote Sensing*
4. Rappaport, C., Stephen Azevedo, Tom Rosenbury, and Jamie Gough, and Arnold Dean, "Detecting Antipersonnel Mines with a Handheld Parabolic Reflector Transmitter / Multistatic Receiver Impulse GPR," *2000 UXO/Countermines Forum*, May 2000.
5. Rappaport, C., Azevedo, S., Rosenbury, T., and Gough, J., "Handheld Forward-Looking Focused Array Mine Detection with Plane Wave Excitation," *SPIE Aerosense Conference*, Orlando, FL, April 2000, pp. 1118--1126.
6. Lam, J., *Journal of Applied Physics*, 66(8), October 1989.
7. Rappaport, C., Wu, S., and Winton, S., "FDTD Wave Propagation Modeling in Dispersive Soil Using a Single Pole Conductivity Model," *IEEE Transactions on Magnetics*, vol. 35, May 1999, pp. 1542--1545.
8. Sahin, A., Rappaport, C., and Dean, A., "Design Considerations for Short Time Pulse TEMR Antennas using Finite Difference Time Domain Algorithm," *SPIE Aerosense Conference*, Orlando, FL, April 1998, pp. 784-793.



A Time Domain Signature Investigation for the GPR Detection of Plastic Land Mines Buried in Soils

Audrey J. Dumanian^{1, 2}

Carey M. Rappaport¹

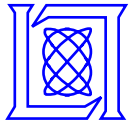
Ann W. Morgenthaler¹

¹Northeastern University

²MIT Lincoln Laboratory

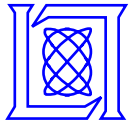
IEEE URSI B - Rough Surfaces and Random Media

July 10, 2001



Outline

- ➔ • **Background**
- **Frequency Domain Results**
- **Time Domain Results**
- **Summary**



Land Mine Detection Problems

- No single type of “target”
 - 647 types of mines (i.e. size, shape, material - plastic)
- Ground is different from country to country
 - ground isn't flat:
 - rough soil surface may also randomly scatter much of the transmitted sensing signal and be a significant source of detection noise ----> this is a MAJOR problem**
 - soil type is never known (inhomogeneous soil, moisture content, etc.)
 - it may also have buried rocks, moisture pockets, tree roots, and bits of scrap metal especially if a former battlefield
- Burial depth of mine in ground varies
 - exposed
 - just under surface
 - shallow burial
 - deep burial

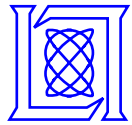
} Radar not effective



Humanitarian Demining Mission

- **Finding anti-personnel land mines:**
 - small “targets”: ~ 10 cm diameter
 - non-metallic: ~ .5 gram metal firing pin
 - not deeply buried < 15 cm deep
- **Ground-Penetrating Radar (GPR) ----> one approach for shallow penetration**
 - distinguish mine based on the shape of anomaly in soil
 - time domain or frequency domain
 - inexpensive to build
 - very susceptible to natural clutter

For GPR applications, the “target” is characterized not only by the inherent properties of the buried object and soil but also by the effects of the air-ground interface



Finite Difference Frequency Domain Technique

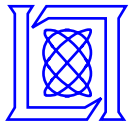
FDFD

- **Computational tool for simulating the scattered wave distribution**
 - A direct solution of Maxwell's curl equations in the frequency domain
- **Permits accurate representation at each frequency**
 - random rough surface (ground roughness)
 - various target shapes
- **Appropriate for near-field scattering**



Outline

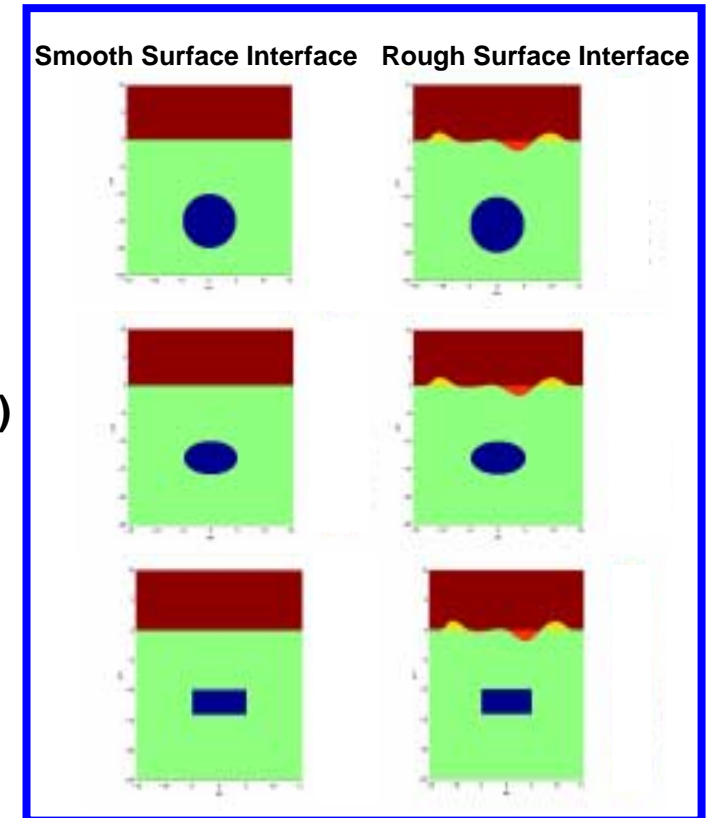
- Background
- ➔ • Frequency Domain Results
- Time Domain Results
- Summary



2-D FDFD Land Mine Simulation Cases

Parameters Varied (36 Cases Total)

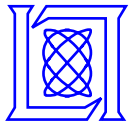
- **TNT Mine:** $E_r = 2.9(1+0.001i)$
 - Circular (5 cm radius)
 - Elliptical (5 cm by 3 cm)
 - Rectangular (10 cm by 4 cm)
- **Ground Surface**
 - Smooth
 - Rough (avg. roughness = 1 cm, corr. length = 4 cm)
- **Ground Type**
 - **Dry Sand:** $E_r = 2.55(1+0.01i)$ at .5 GHz
 - **Bosnian Soil:** $E_r = 8.9(1+0.13i)$ at .5 GHz
- **Burial Depth of Mine in Ground**
 - 2.5 cm, 5 cm, and 10 cm
- **Frequency Bandwidth:** .5 to 5 GHz, 100 MHz steps
- **Time Domain Considerations:**
 - 341 receivers above mine (transverse position)
 - receiver observation heights: at interface, 2.5 cm and 5 cm above interface,



Note: Each FDFD simulation took 12 hours to run on an SGI Octane (output file = 100 MB)

MIT Lincoln Laboratory

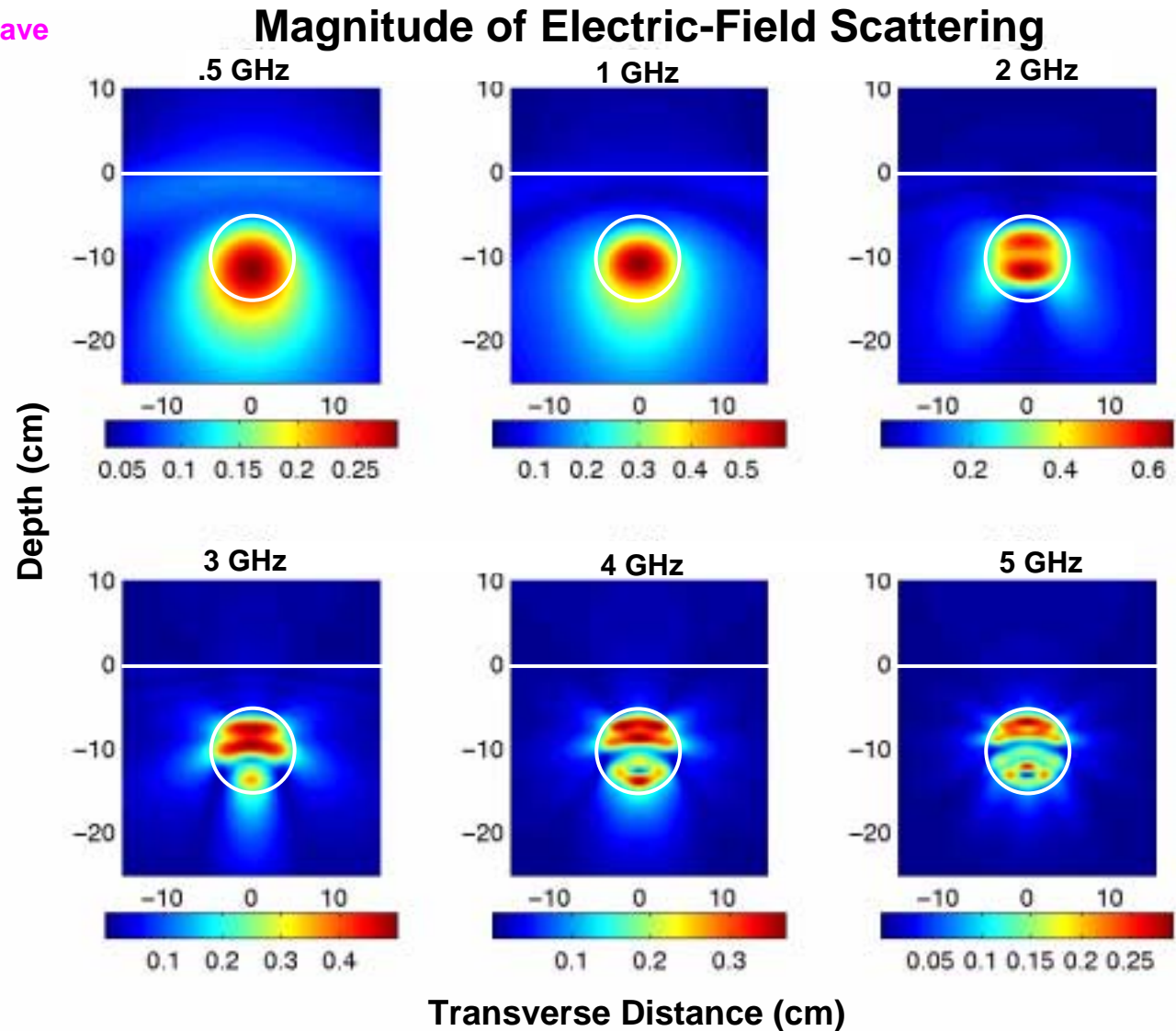
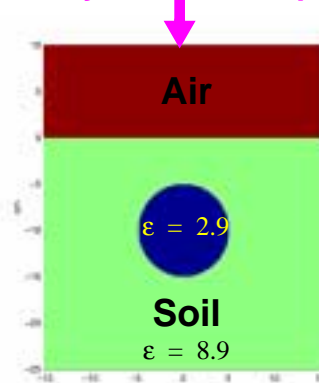
This viewgraph is under construction



Circular TNT Mine 5 cm Deep in **Bosnian Soil**

FDFD Simulation Results for Smooth Surface Interface

normally incident TM plane wave

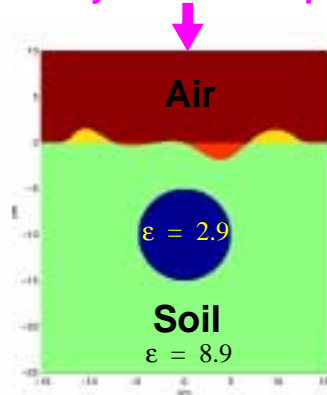




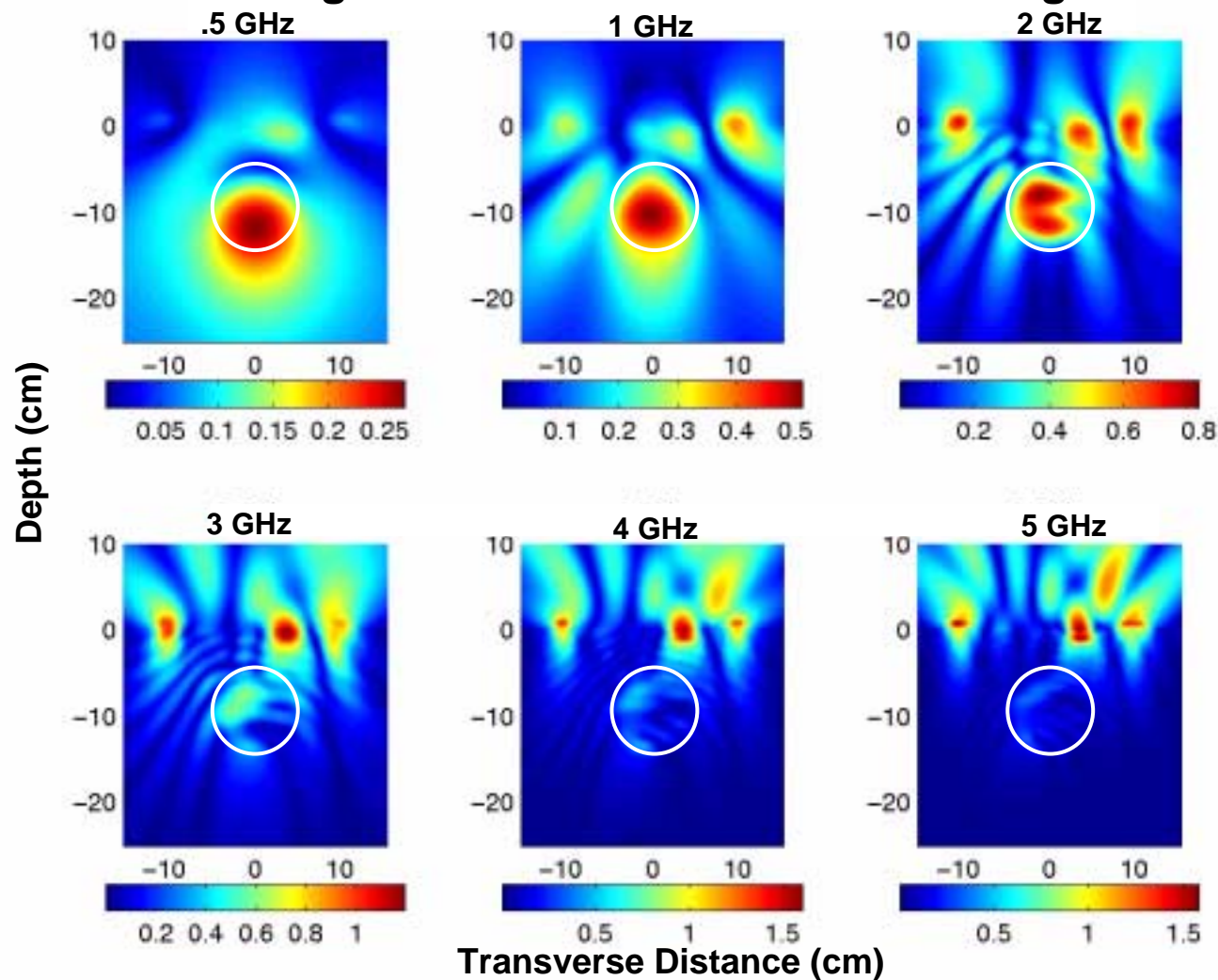
Circular TNT Mine 5 cm Deep in **Bosnian Soil**

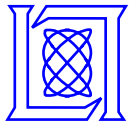
FDFD Simulation Results for Rough Surface Interface

normally incident TM plane wave



Magnitude of Electric-Field Scattering

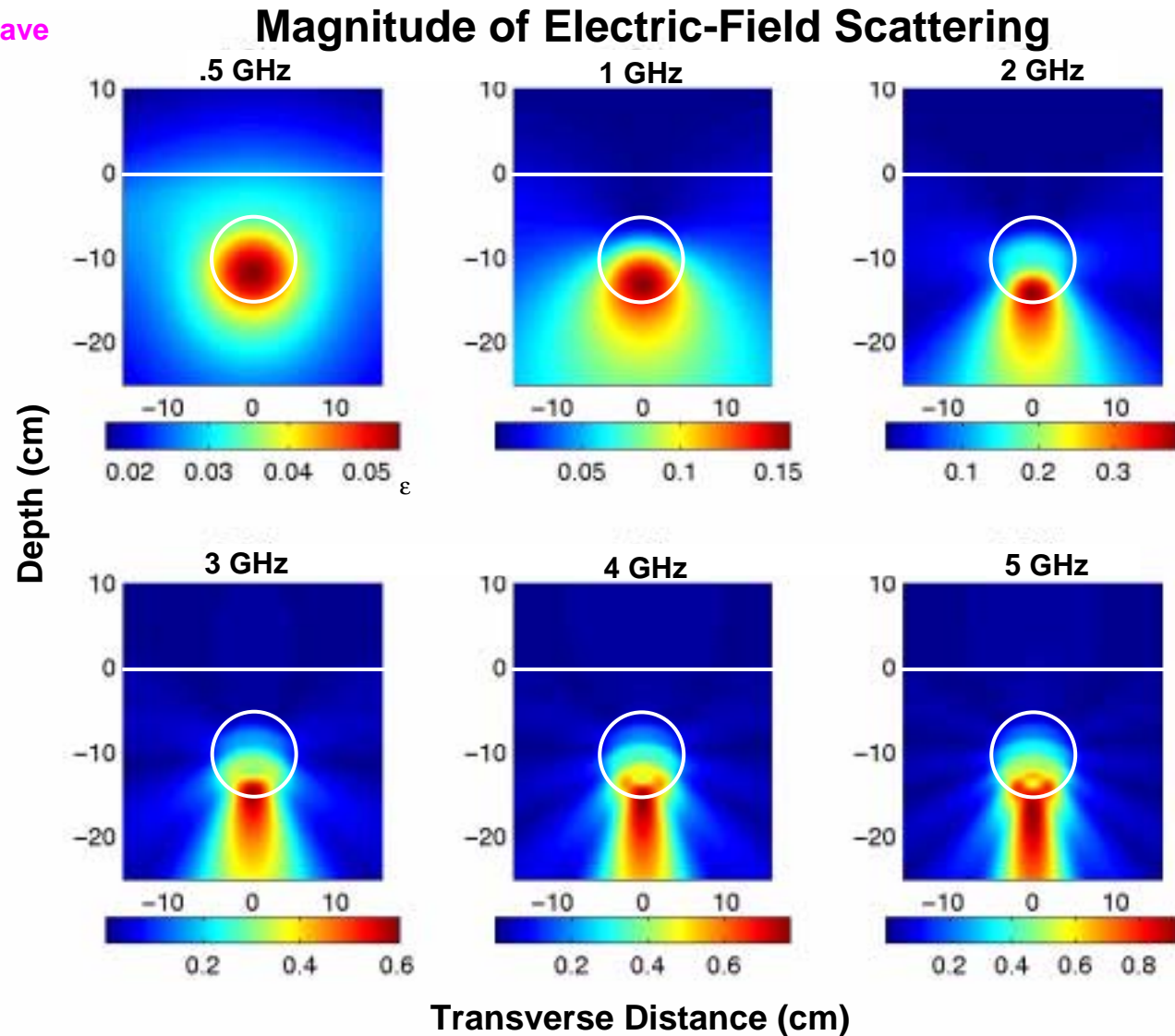
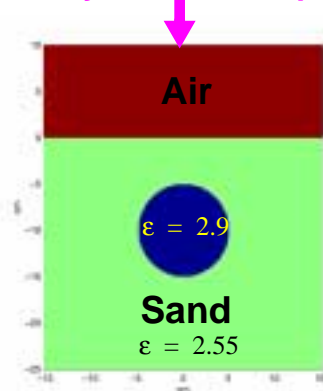


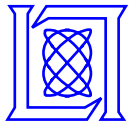


Circular TNT Mine 5 cm Deep in Dry Sand

FDFD Simulation Results for Smooth Surface Interface

normally incident TM plane wave

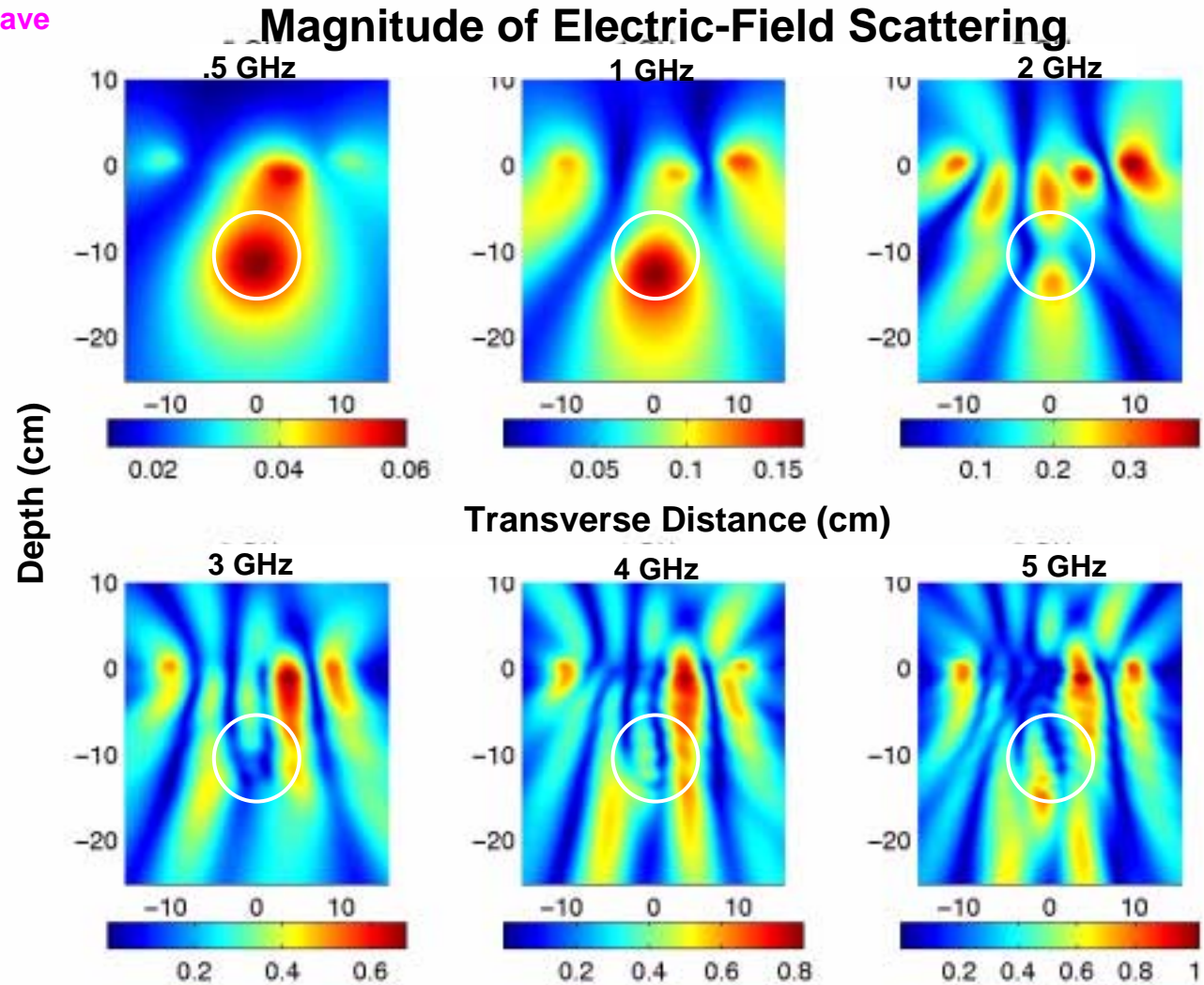
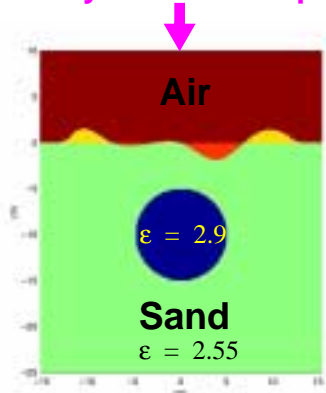




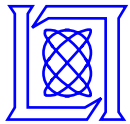
Circular TNT Mine 5 cm Deep in Dry Sand

FDFD Simulation Results for Rough Surface Interface

normally incident TM plane wave



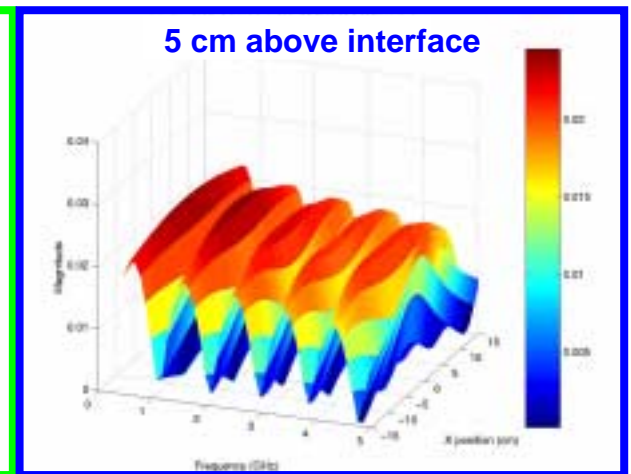
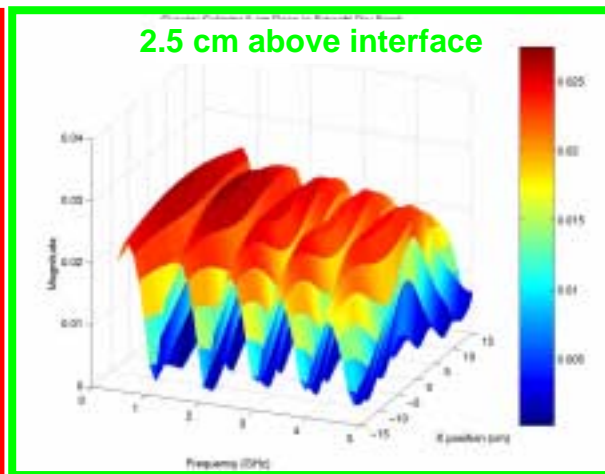
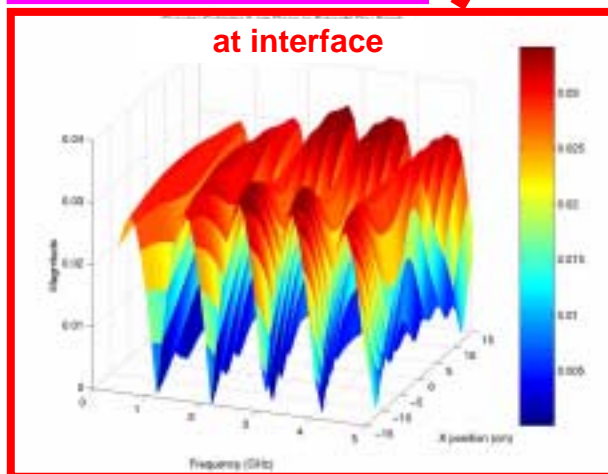
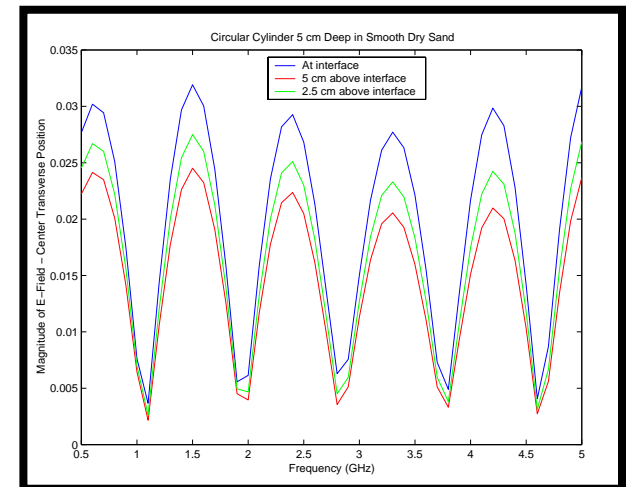
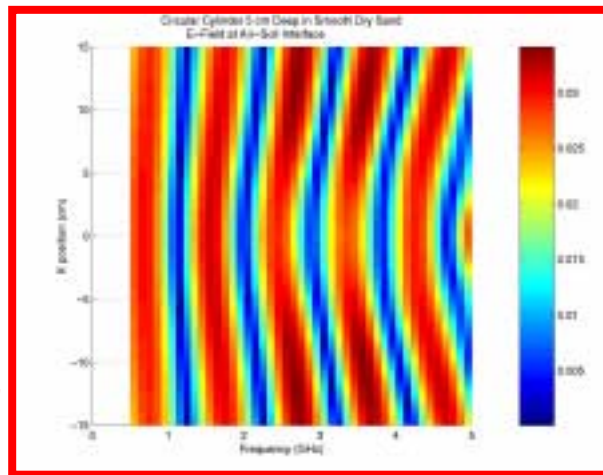
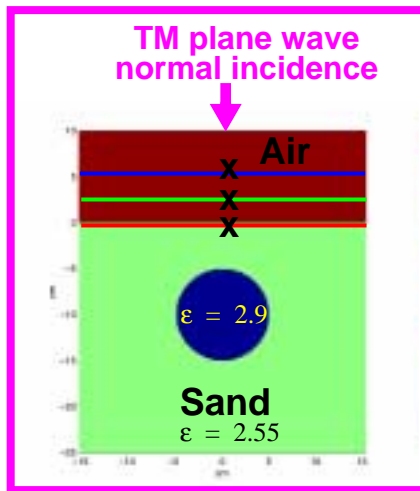
At high frequencies the rough surface scattering dominates

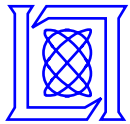


Circular TNT Mine 5 cm Deep in Dry Sand

FDFD Simulation Results for Smooth Surface Interface

E-Field vs. Frequency vs. Transverse Position at fixed depth heights:
At **interface**, **2.5 cm** and **5 cm** above air-soil interface



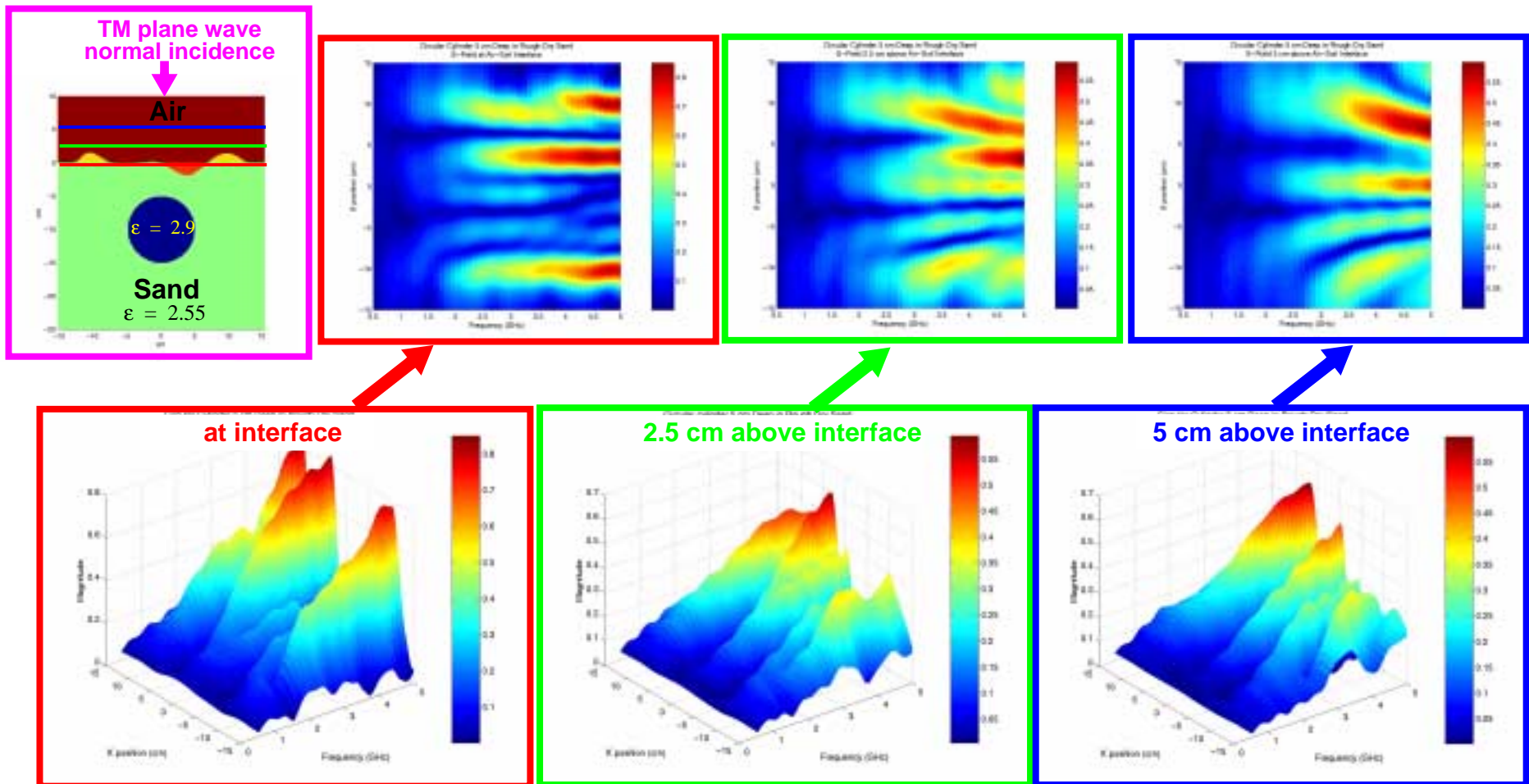


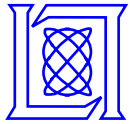
Circular TNT Mine 5 cm Deep in Dry Sand

FDFD Simulation Results for Rough Surface Interface

E-Field vs. Frequency vs. Transverse Position at fixed depth heights:

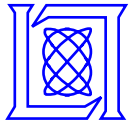
At **interface**, **2.5 cm** and **5 cm** above air-soil interface





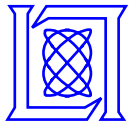
Outline

- Background
- Frequency Domain Results
- • Time Domain Results
- Summary



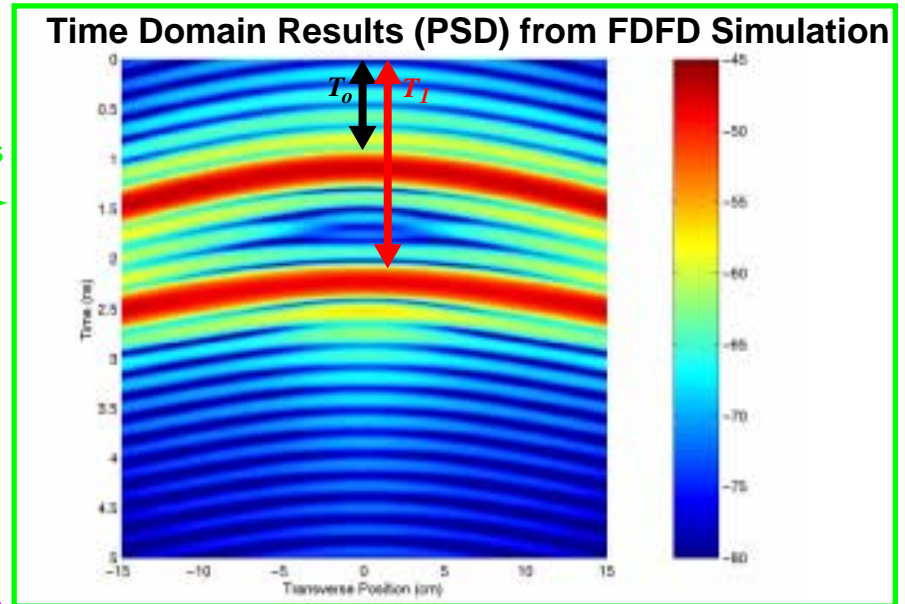
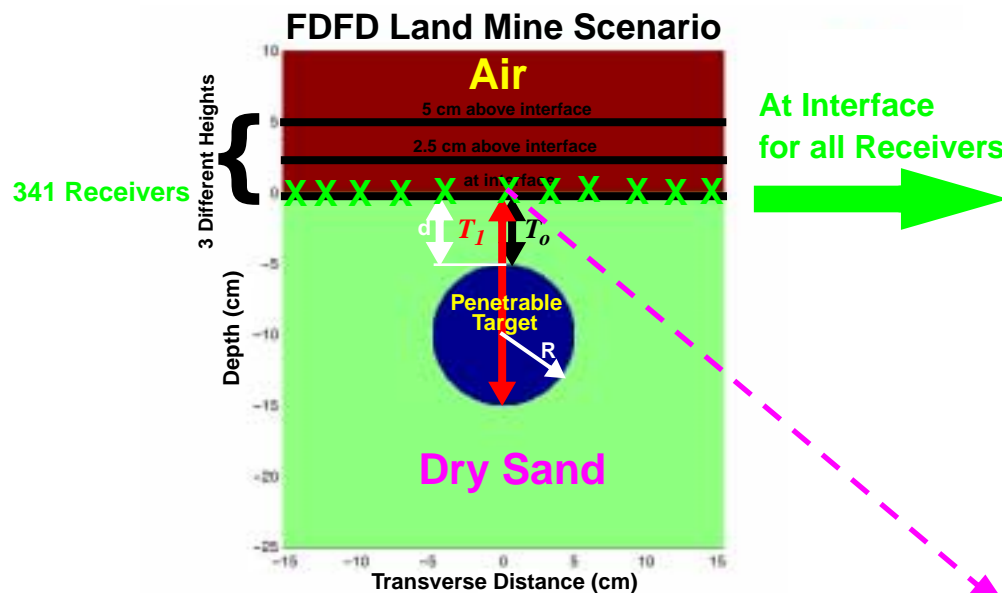
Time Domain Signature Investigation

- The power spectral density of the reflected signal yields a time domain signature dependent of the soil characteristics as well as the burial depth and radius of the mine
- Two or more characteristic time peaks (T_0 and T_1) are observed suggesting that the ultra-wideband spectral radar response may yield particular advantages not exploited by currently employed detection systems
- The ultra wideband timing characteristic behavior is examined for three different heights at/above interface (i.e., at interface, 2.5 cm and 5 cm above interface) over the entire transverse distance and time
- This study determines the feasibility of detecting mine-like targets based on stepped-frequency GPR time signatures



Characteristic Times for Circular Target in Lossy Media

Land Mine (5 cm radius) Buried 5 cm Deep in Dry Sand with Smooth Air-Ground Interface

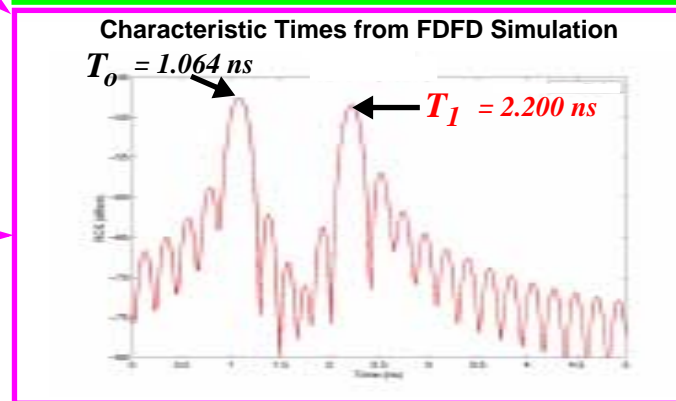


Characteristic Time Equations at Interface directly above mine

$$T_o = \frac{2d}{c_g}$$

$$T_1 = \left(\frac{2d}{c_g} + \frac{4R_{av}}{c_m} \right)$$

d = burial depth
 c_g = average soil wave velocity
 c_m = mine wave velocity
 R = average mine radius

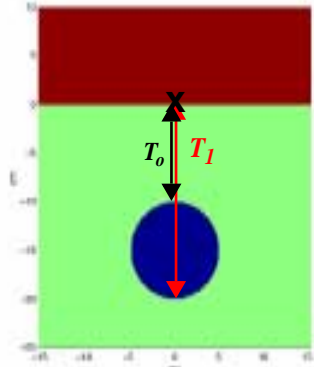


Two-way travel times, T_o and T_1 , are timing peaks that directly correspond to the burial depth and target size if the dielectric constant of the land mine and soil are known a priori

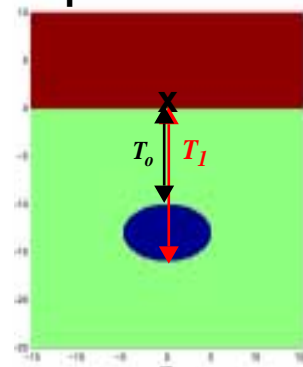


Time Domain Scattering Results for Land Mines Buried 10 cm Deep in **Dry Sand**: Smooth Air-Ground Interface

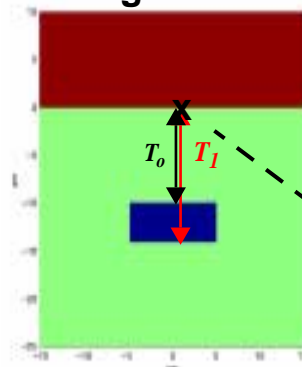
Circular Land Mine



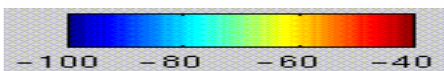
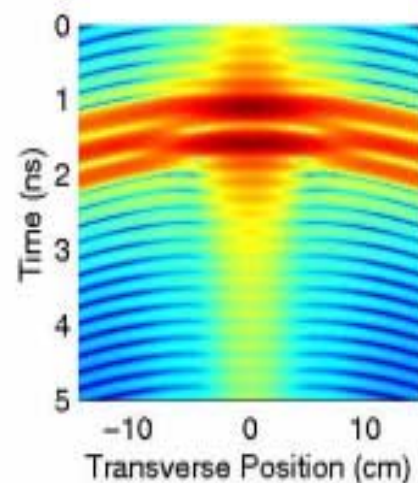
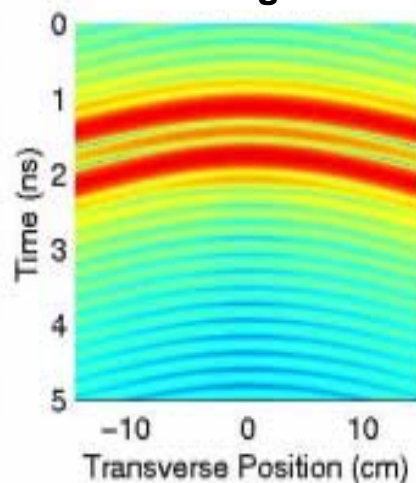
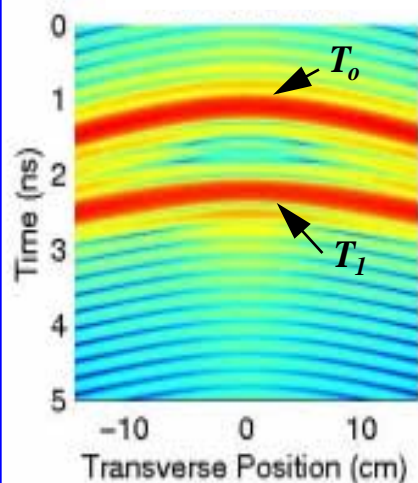
Elliptical Land Mine



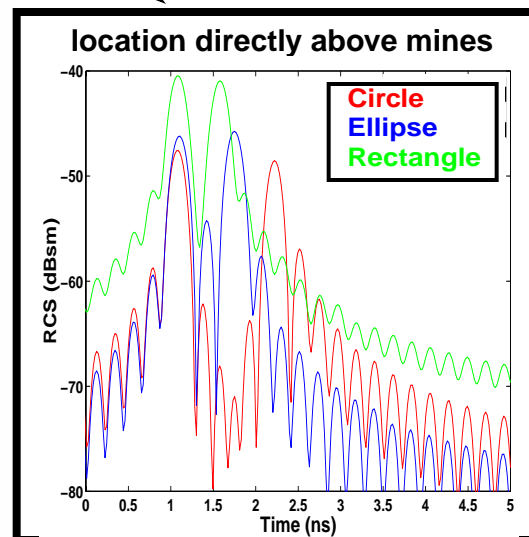
Rectangular Land Mine

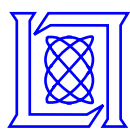


PSD Results using Welch's Method



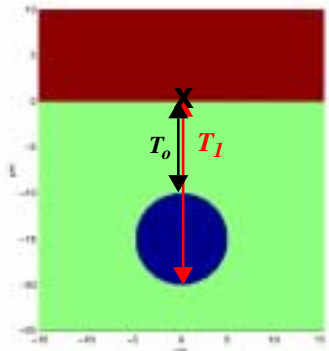
At Interface



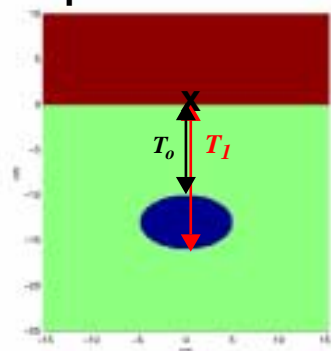


Time Domain Scattering Results for Land Mines Buried 10 cm Deep in **Bosnian Soil**: Smooth Air-Ground Interface

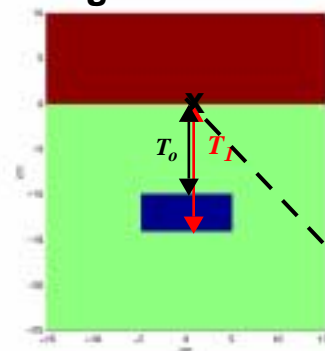
Circular Land Mine



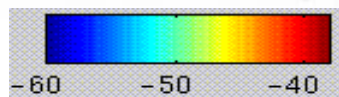
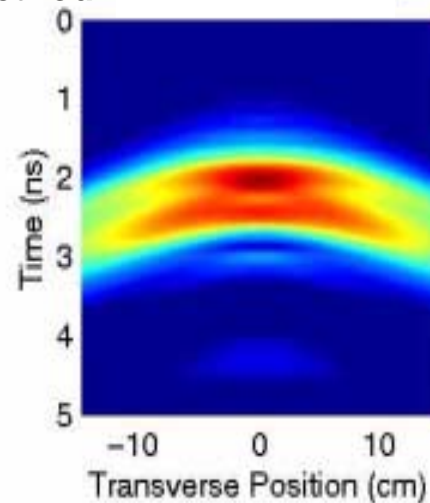
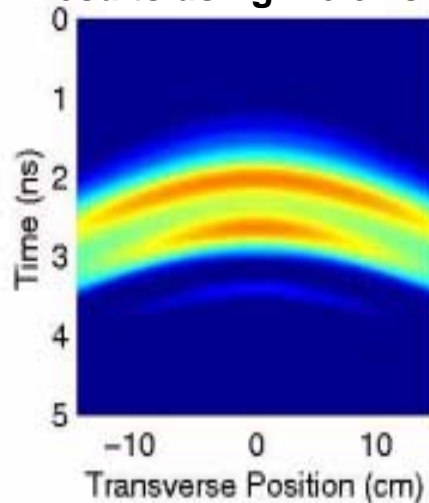
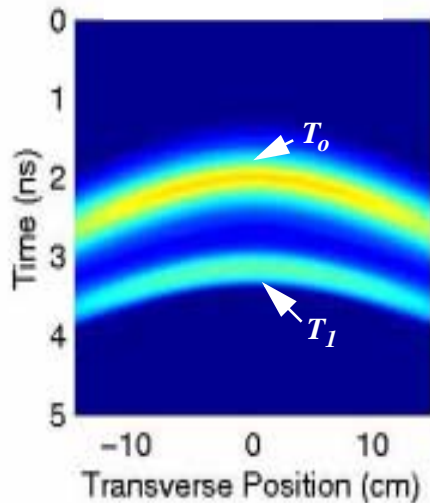
Elliptical Land Mine



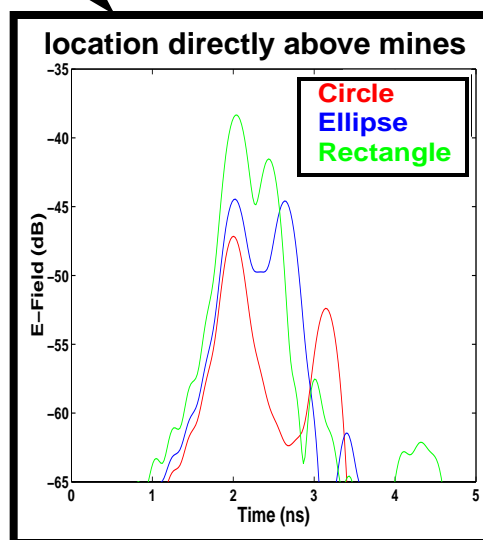
Rectangular Land Mine

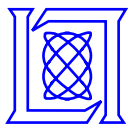


PSD Results using Welch's Method



At Interface





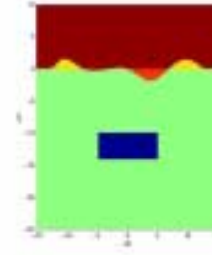
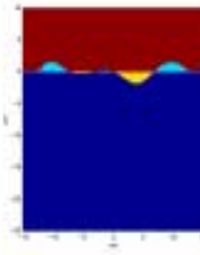
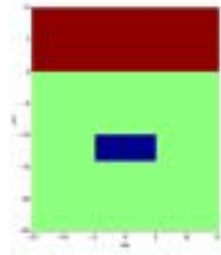
Time Domain Scattering Results for Rectangular Land Mine Buried 10 cm Deep in Dry Sand

Smooth Surface Interface versus Rough Surface Interface

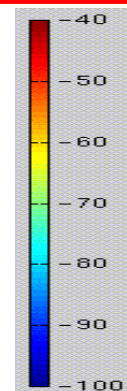
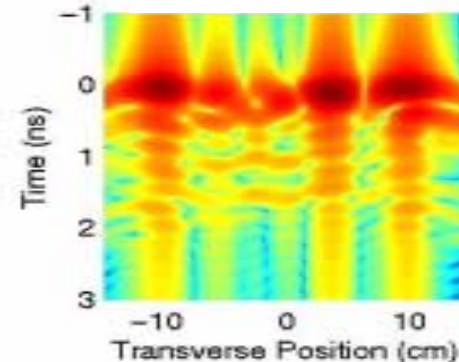
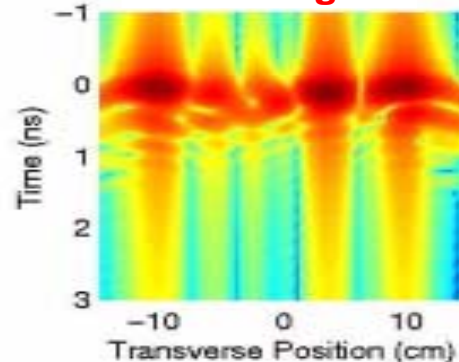
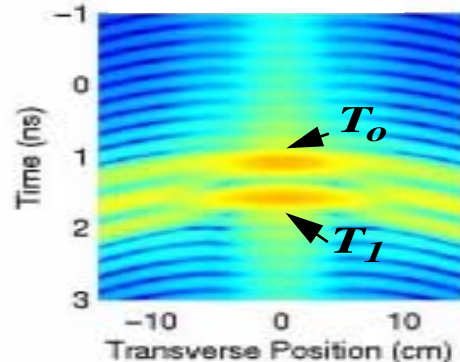
Smooth Surface Interface w/Mine

Rough Surface Interface only

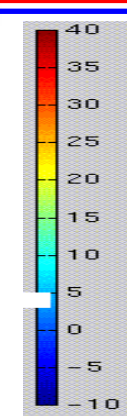
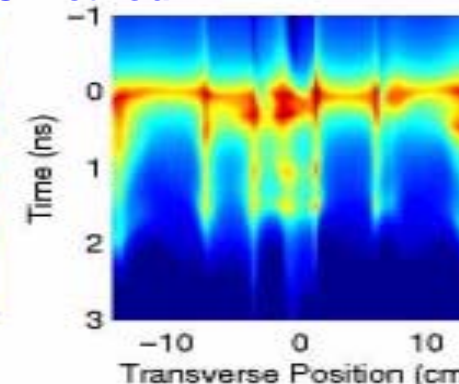
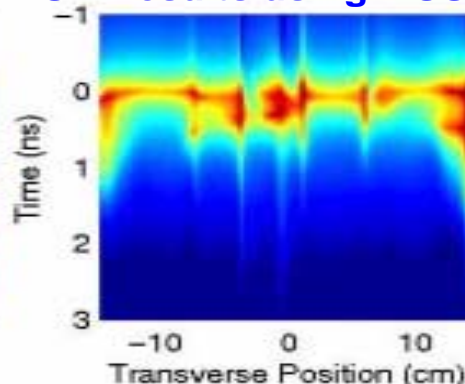
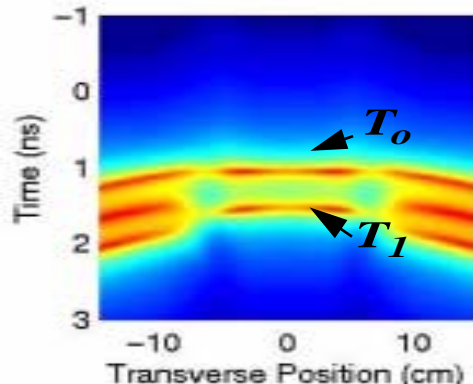
Rough Surface Interface w/Mine



PSD Results using Welch's Method



PSD Results using MUSIC Method

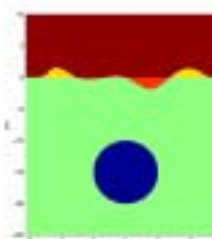
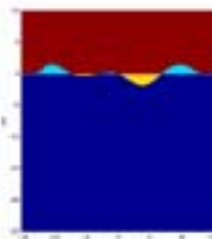
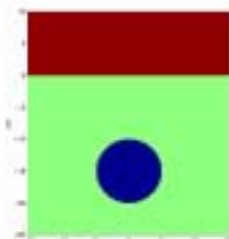




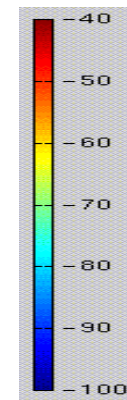
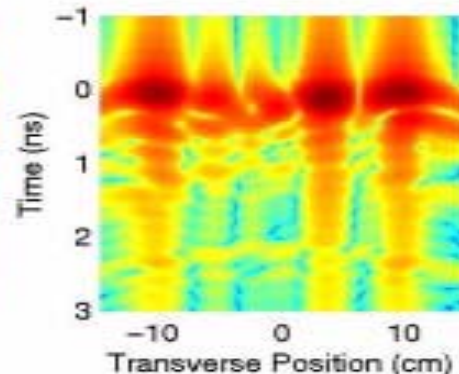
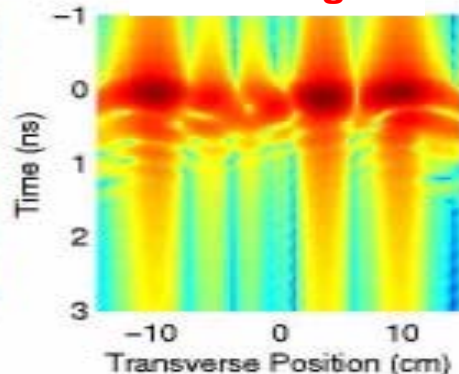
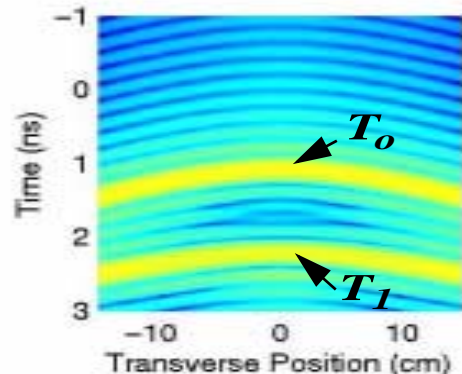
Time Domain Scattering Results for Circular Land Mine Buried 10 cm Deep in Dry Sand

Smooth Surface Interface versus Rough Surface Interface

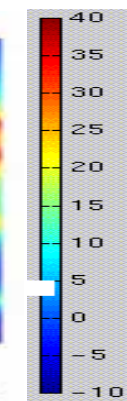
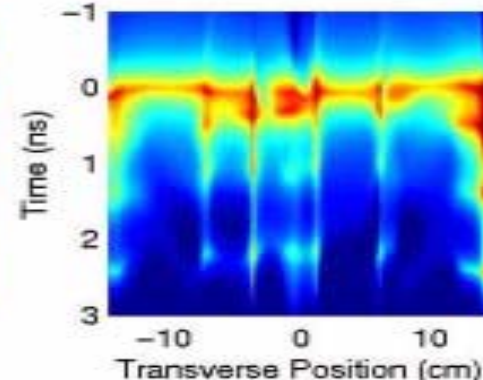
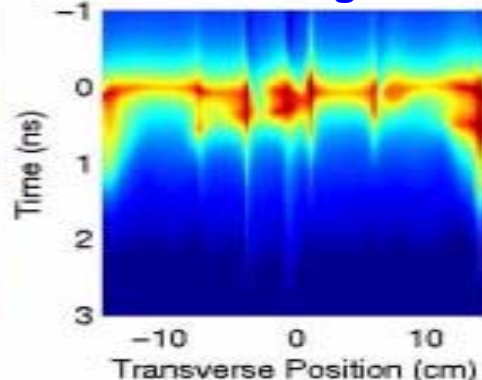
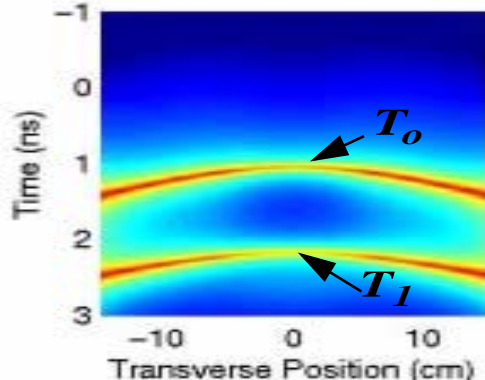
Smooth Surface Interface w/Mine Rough Surface Interface only Rough Surface Interface w/Mine



PSD Results using Welch's Method



PSD Results using MUSIC Method

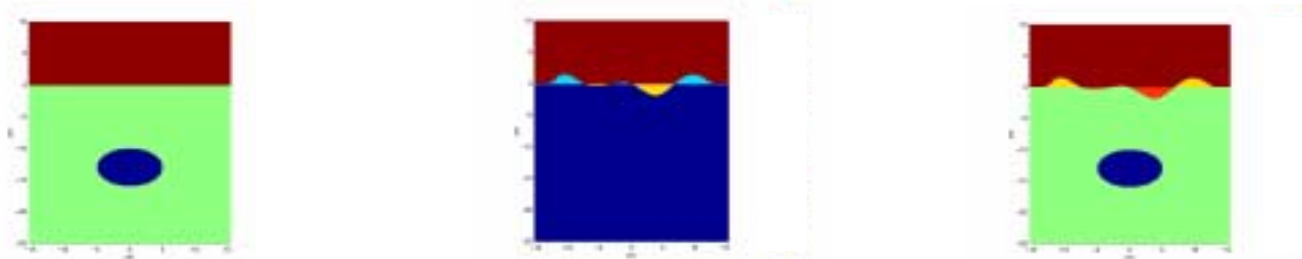




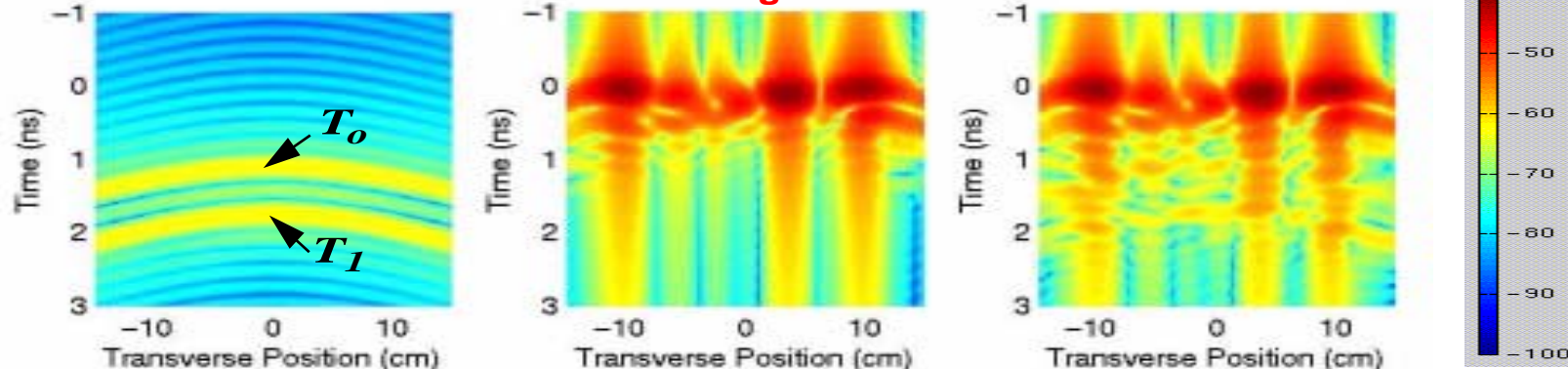
Time Domain Scattering Results for Elliptical Land Mine Buried 10 cm Deep in Dry Sand

Smooth Surface Interface versus Rough Surface Interface

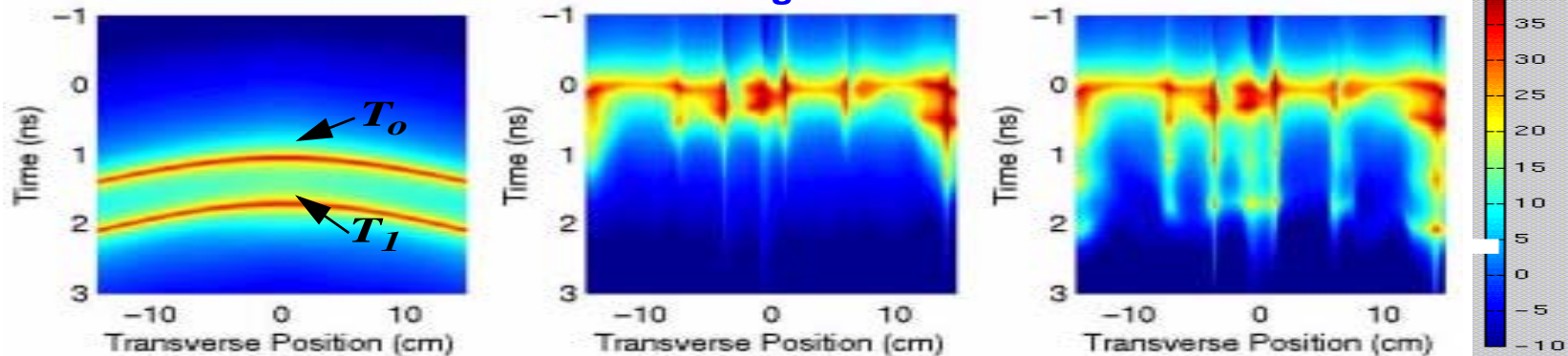
Smooth Surface Interface w/Mine Rough Surface Interface only Rough Surface Interface w/Mine



PSD Results using Welch's Method



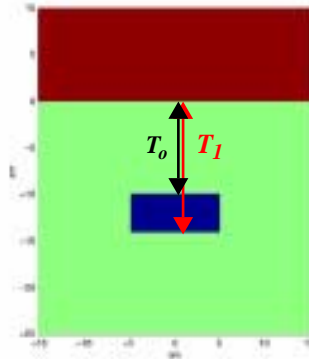
PSD Results using MUSIC Method



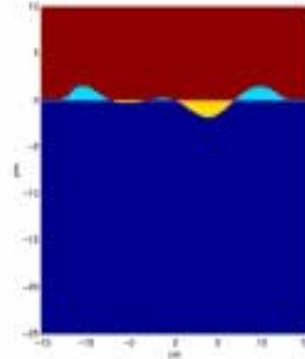


Time Domain Scattering Results for Rectangular Land Mine Buried 10 cm Deep in **Bosnian Soil** Smooth Surface Interface versus Rough Surface Interface

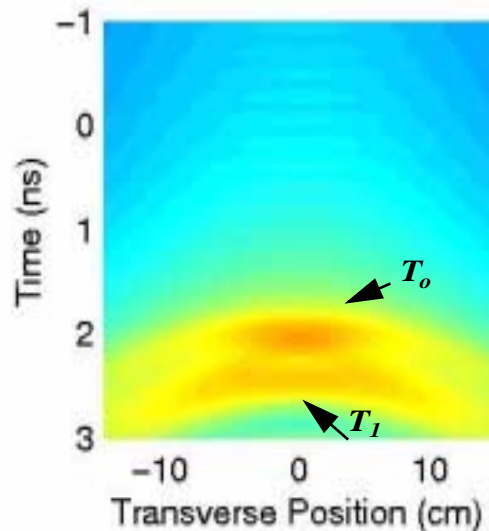
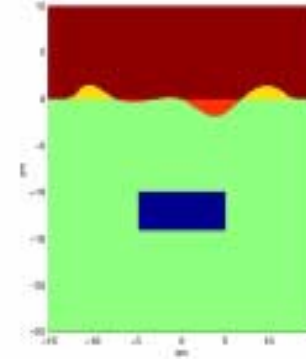
Smooth Surface Interface with Mine



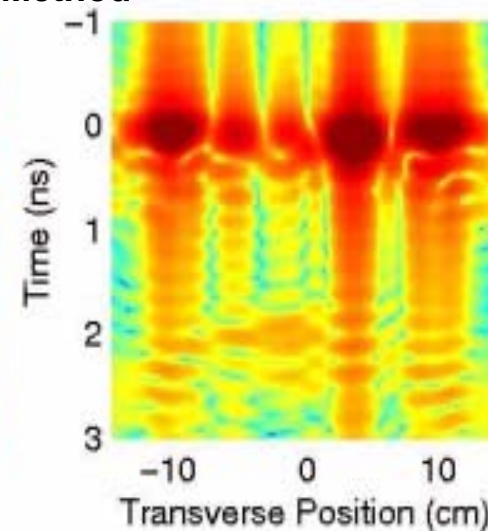
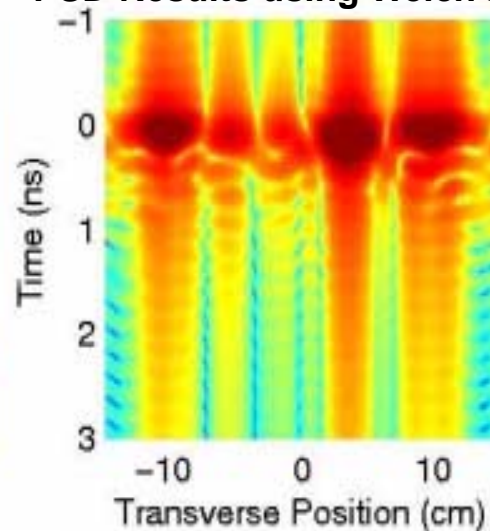
Rough Surface Interface only



Rough Surface Interface with Mine



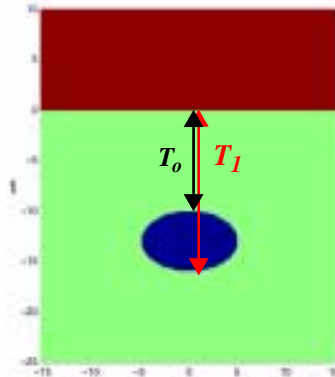
PSD Results using Welch's Method



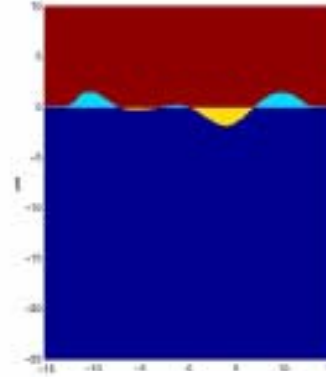


Time Domain Scattering Results for Rectangular Land Mine Buried 10 cm Deep in **Bosnian Soil** Smooth Surface Interface versus Rough Surface Interface

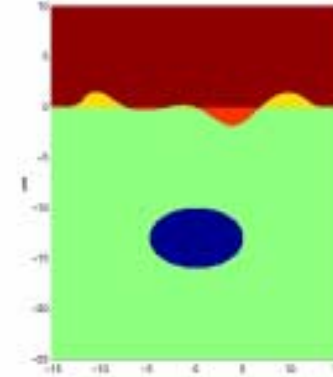
Smooth Surface Interface with Mine



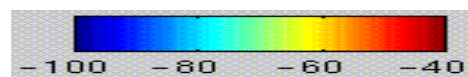
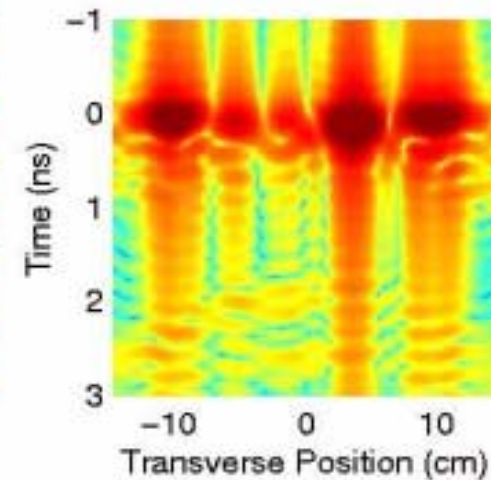
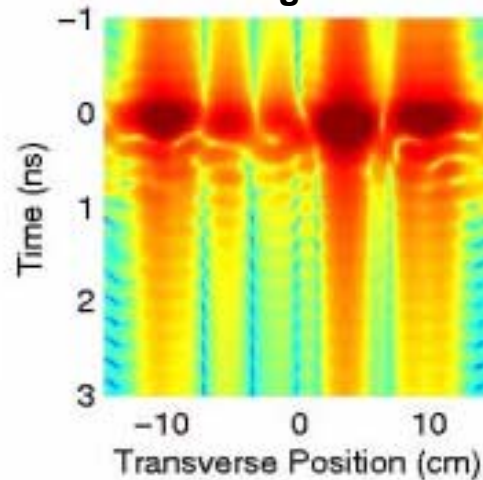
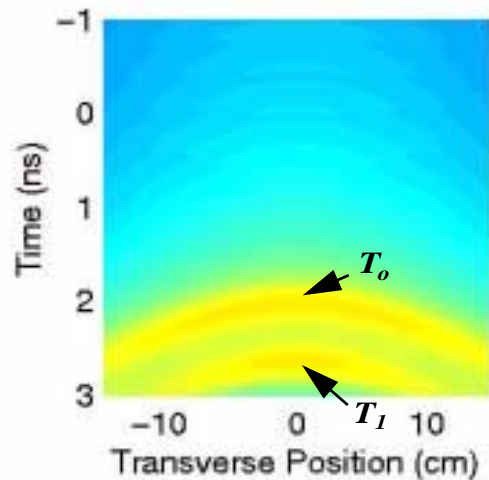
Rough Surface Interface only



Rough Surface Interface with Mine



PSD Results using Welch's Method





Summary

- **Rough surface effects are a dominant scattering mechanism in the GPR detection of buried objects**
- **Frequency domain numerical modeling (FDFD) is a viable means of testing candidate sensing concepts on a wide variety of conditions**
- **Using the appropriate signal processing techniques, land mine identification using spectral timing peaks is a viable means to determine land mine burial depth and size**
 - **wideband processing enables time signatures that highlight invariance**
- **Application: stepped-frequency GPR radar**

Northeastern University Demining MURI 5th Year EXECUTIVE SUMMMARY



the university of
OKLAHOMA



Carey Rappaport
Principal Investigator
Dept. Electrical and Computer Engineering
Northeastern University

Guiding Principles

- **Basic Research**
 - Systems approach: optimize sensor configuration & combination, modeling, processing, and reconstruction
 - Establish quantifiable performance bounds
- **Combined Modalities and Algorithms**
 - Devise sensors with multiple physical effects
 - Fuse multiple sensor information
- **Support Theory with Real Measured Data**
 - Design the sensor configuration first
 - Control and understand data collection

Demining Mission

- **Humanitarian Demining: Find Anti-personnel mines**
 - Small targets: ~10 cm diameter
 - Non-metallic: ~0.5 gram metal firing pin
 - Not deeply buried < 15 cm deep
- **Target in real environment**
 - Inhomogeneous soil
 - Realistic, rough ground interface
- **Practical detection modalities**
 - Efficient
 - Low cost
 - Light weight
 - Robust/Flexible/Rugged
- **Develop modality to suit mine position in ground**

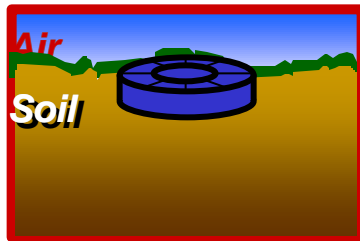
Non-Metallic AP Mine Burial Scenarios; Proposed Solution Modalities

Exposed



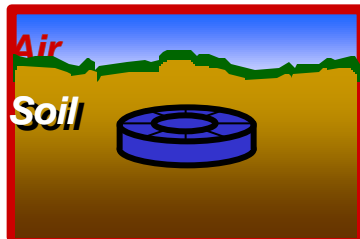
Polarimetric IR, EM Induction

Just Under Surface



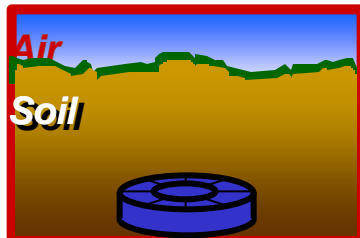
Microwave Enhanced IR Thermography,
Laser/Acoustic, EM Induction

Shallow Buried



Clay: Focused GPR, EM Induction
Sand: Laser/Acoustic, EM Induction

Deep Buried



EM Induction, GPR Spectral Analysis

Technology Transitions



CenSSIS: \$28M NSF ERC
Subsurface Sensing in Soil,
Water and Body

Antenna element analysis for
GeoCenters, Inc.
Humanitarian Demining program

Processed NIITEK (Wichmann)
GPR data

INEEL:
Subsurface
pollutant detection

Processed EG&G GPR data,
Paper accepted at TGARS

SERDP SEED contract:
EMU classification methods
for UXO discrimination and classification

Raytheon:
Counter-Terrorism efforts

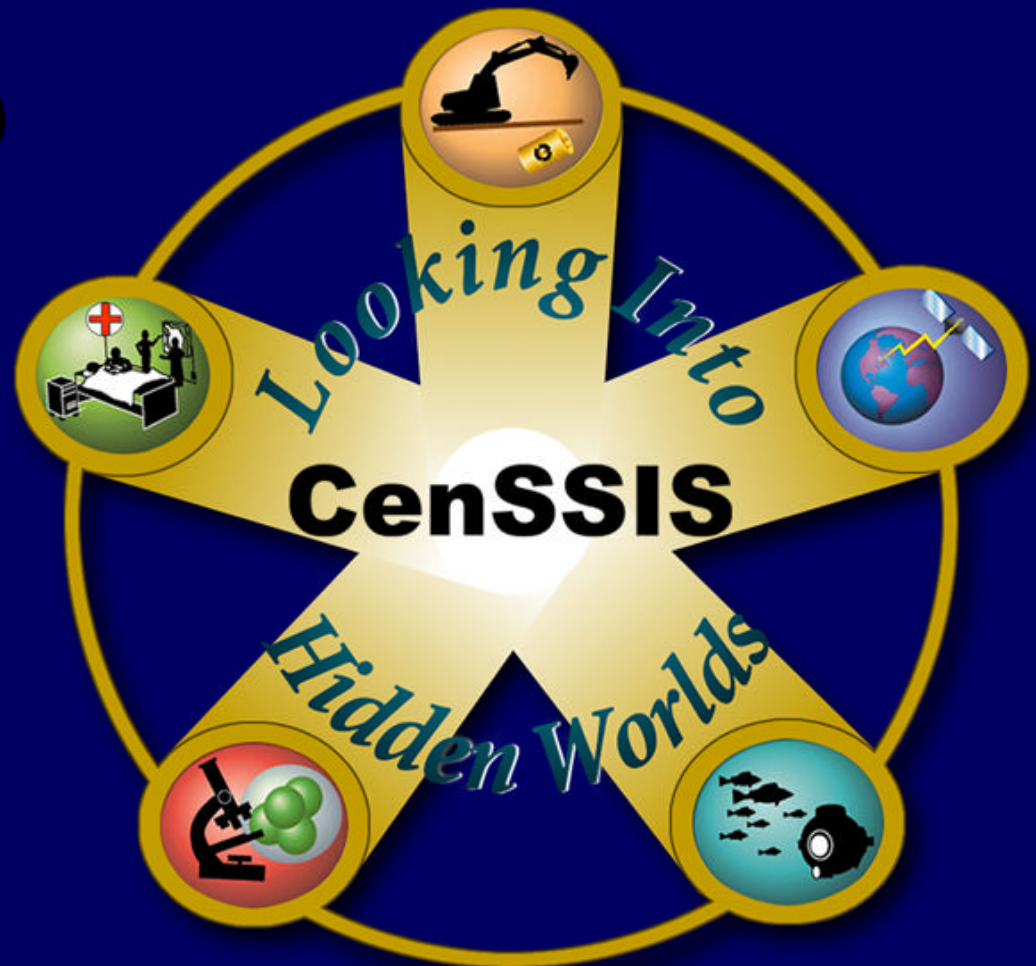
Raytheon:
Vehicles in Foliage

Processed European,
DeTeC consortium GPR data

Center for Subsurface Sensing & Imaging Systems

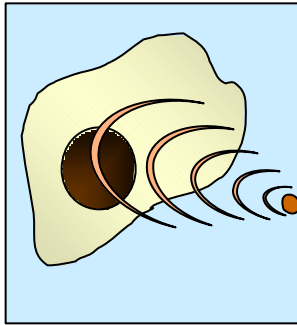


- ***FY 2000 NSF ERC
(Eng. Research Center)***
- ***Very Competitive
2 out of 89 Funded
10 Year Program***
- ***\$28 Million Initial
5 year Investment
NSF/Univ Partners***
- ***Direct MURI Legacy***
- ***www.censsis.neu.edu***



The Scope of CenSSIS Goes Beyond Humanitarian Demining

Subcellular Biology

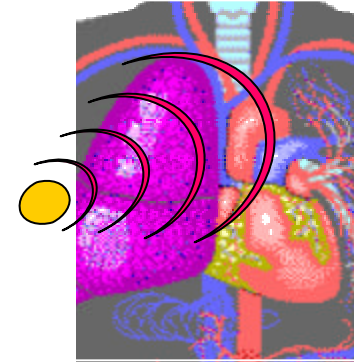


100nm- 0.01 mm

Optics

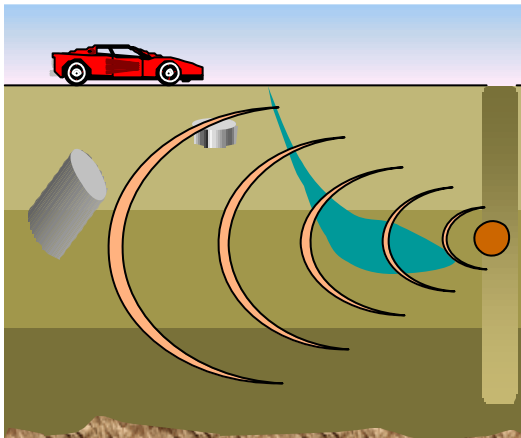
Ultrasound

Tissues & Organs



0.1 mm - 10 cm

**Underground
Diagnosis**



1 cm - 100 m

Radar

Sonar

**Underwater
Exploration**



10 cm - 1 km

The CenSSIS Focus: Multi-Application Subsurface Detection and Sensing Concepts

Long Range Goals

- Attack important real world problems
- Create a unified framework
 - *Diverse Problems – Similar Solutions*
- Combine new multi-sensor instruments & methods into an engineered system
- Implement a multi-disciplinary education program within a distributed university
- Increase the number of women and minorities in engineering
- Create sustained industrial partnerships

Key Milestones of the CenSSIS Master Research Plan

Application to Real World Problems

BED and I-PLUS Integration

SoilBED (1st Gen)

BioBED (1st Gen)

Unified Framework (1st Gen)

MedBED SeaBED (1st Gen)

SoilBED (2nd Gen)

I-PLUS (1st Gen)

BioBED, MedBED, SeaBED (2nd Gen)

Integrated SSI Projects

2D Quadrature Microscope

Indoor Multilayer Hyperspectral Imaging

Quantitative Subretinal and Undersea Mapping

Real Time Molecular Tracking

Diffuse Optical Tomography
High Contrast Ultrasound Imaging

Advanced Electrical Impedance Tomography

3D Quadrature Microscope

R3 Projects

Data and Database Standards

First Gen Collaborative Environment

Physics Based Compression Tools

Second Gen Collaborative Environment

First Gen Solutionware Modules

Data and MetaData Libraries

Fast Scalable Parallel Computation Tools

R2 Projects

Special Purpose Sensor Fusion Tools

Diffuse Tomography Tools

Robust Nonlinear Inversion Tools

Physics Based Image Understanding Tools

General Purpose Sensor Fusion Tools

R1 Projects

Fast 2D Models

Photo-Acoustic Probe

Special Purpose 3D Models

Nonlinear Ultrasound Probe

Acoustic-Photonic Probe

General Purpose 3D Models

Entangled 2-Photon Probe

Yr 1

Yr 2

Yr 3

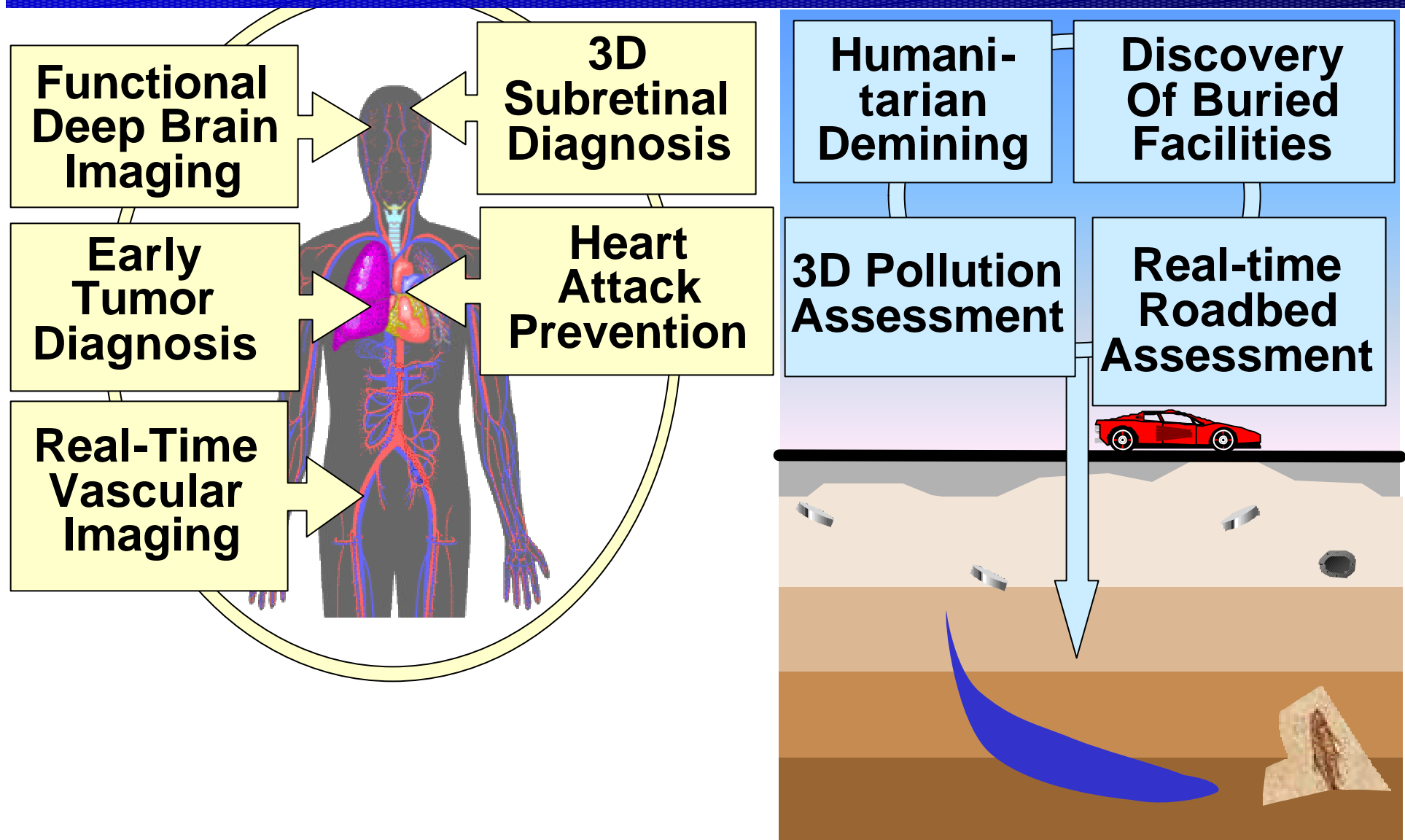
Yr 4

Yr 5

Examples of Year 1 CenSSIS Project Accomplishments

- Diffuse Optical Tomography (DOT) Toolbox
 - Diverse Applications
 - (NU/MGH/Tufts/TMW) Team
- Acoustic Diffraction and Electrical Near Field Probing of Complex Underground Media
 - Buried Waste Assessment
 - (NU/INEEL/UOK) Team
- New Sensor Fusion/ Mosaicing Tools
 - 3D Retinal & Undersea Images
 - (RPI/WHOI) Team
- New Models for Subcellular Dynamics Tested Against Real Data (BWH/NU)
- New Acoustic Method for Cardiac Artery and Underground Assessment
 - (BU/LLNL/NU/Industry) Team

The MURI Legacy: A Vehicle to Attack Important Real World Problems



NEU Demining MURI Publications

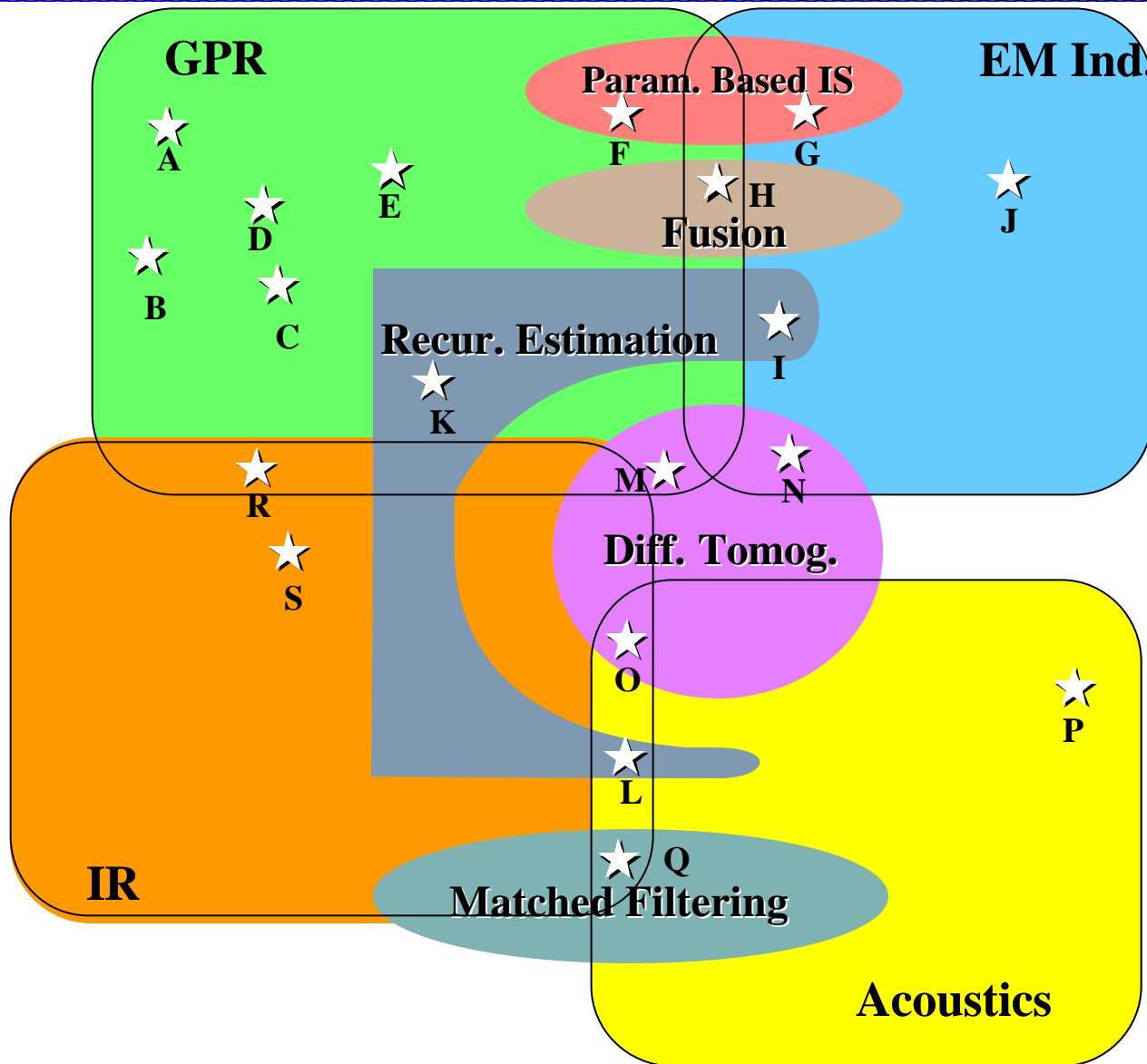
January 1997 – September 2001

- 66 Journal Papers
 - 12 in IEEE Transactions on Antennas and Propagation.
 - 10 in IEEE Transactions on Geoscience and Remote Sensing.
 - 6 in IEEE Transactions on Image Processing.
 - 6 in Journal of Subsurface Sensing Technologies and Applications.
- 149 Conference Papers
 - 25 Invited submissions
 - 3 Keynote presentations

Patents

- DiMarzio, Charles A., Stephen W. McKnight, and Scott C. Lindberg, Optical-Pulse-Induced Acoustic Mine Detection, U. S. Patent 6,069,843. Awarded on 30 May 2000.
- Rappaport, Carey L., and Charles A. DiMarzio, Microwave Enhanced Infrared Thermography, Provisional Application 60/057,253, filed 29 Aug 1997. Full application submitted 27 August 1998.
- Rappaport, Carey M., Handheld Radar Mine Detector with Offset Parabolic Reflector Transmitter and Array Receiver, Provisional Patent filed

NEU MURI Research Integration

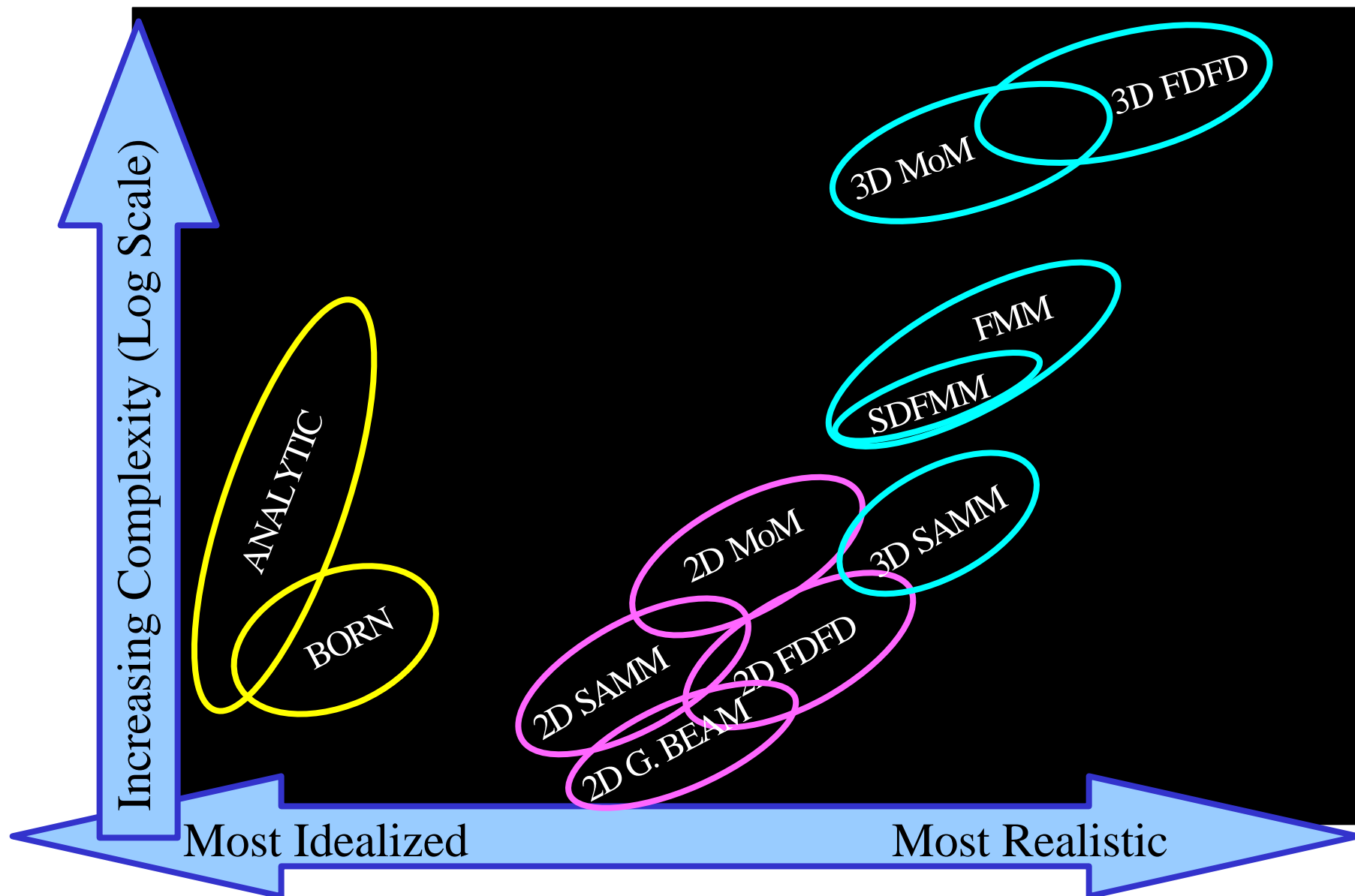


- A. FD Model: rough, wet, shape
- B. Handheld Parabola/ Array
- C. SDFMM
- D. Artif. Diel. Macth Layer
- E. Subsurface Probe
- F. GPR Shape Detection
- G. Diffusive Wave Shape
- H. Anomaly + Fining Pin
- I. Pushbroom EMI
- J. GEM III Spectroscopy
- K. Statist. GPR Array Proc.
- L. Kalman filtering
- M. GPR Psuedo-Inverse.
- N. EMIS
- O. Laser/Acoustic Sensing
- P. Acous Prop. Model/Meas.
- Q. Match Filt. Laser/Acoust.
- R. Microw. Enhanced IR
- S. Polarimetric IR

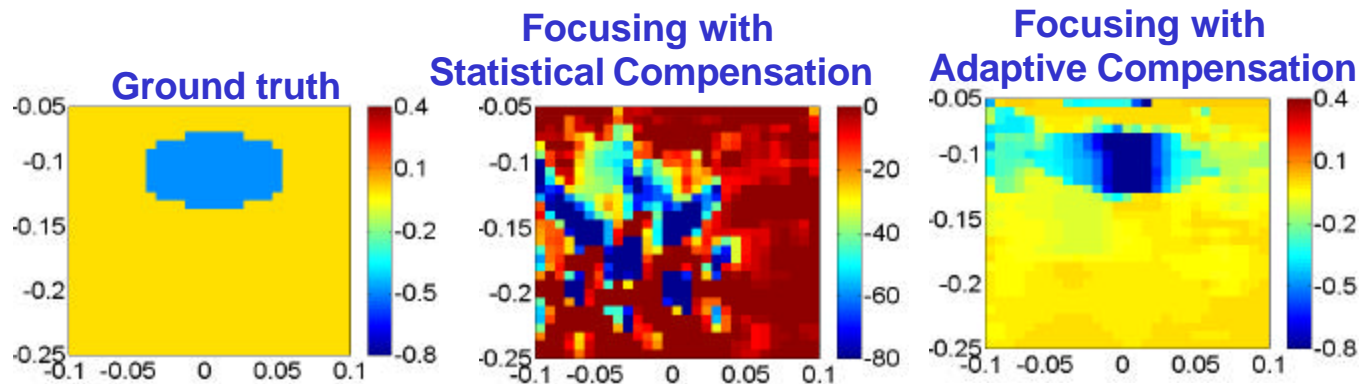
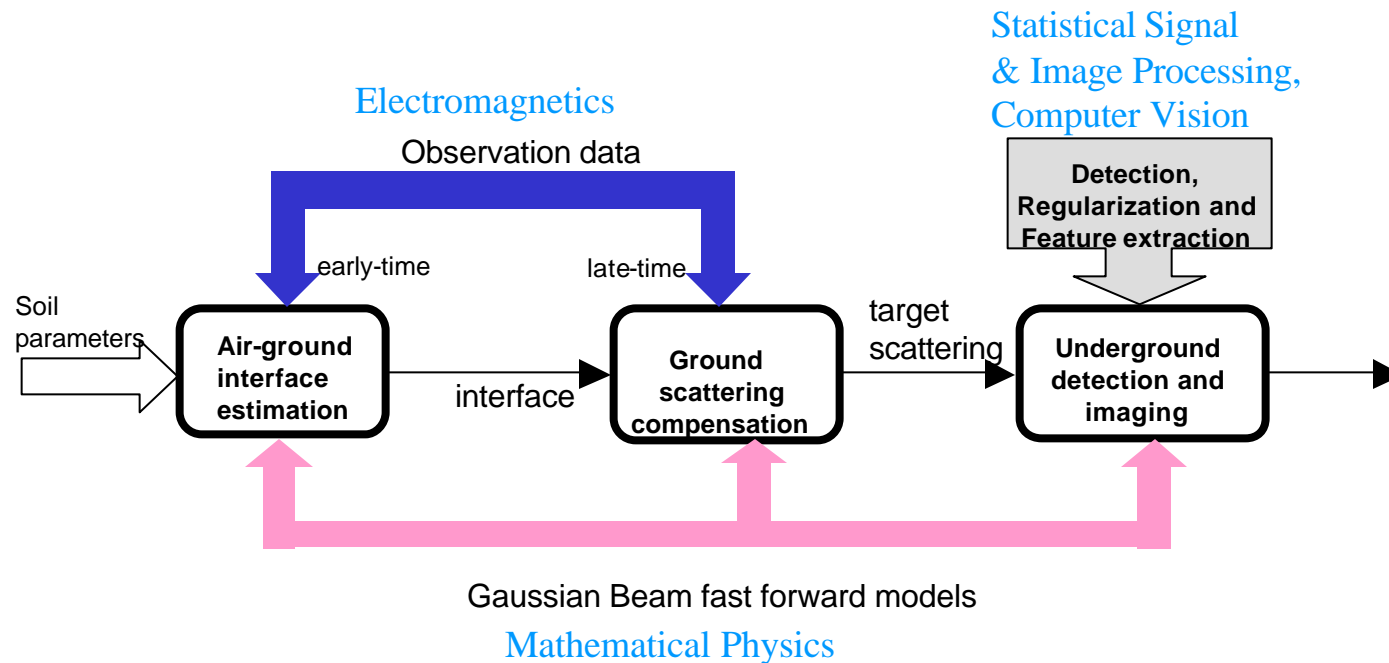
Computational Modeling Applications

- **Gaussian Beam Asymptotics (2-D ~~↗~~ 3-D)**
 - Fast rough surface effect modeling
- **SDFMM (3-D)**
 - Fast frequency domain MoM
 - Two target analysis
- **FDFD (3-D)**
 - Slow but versatile frequency domain
 - Shows target feature invariant
 - Guides clutter removal strategy
- **Dispersive FDTD (3-D)**
 - Monte Carlo rough surface analysis
 - Establishes ground cutter signal ident. and removal strategy

Frequency Domain Computational Electromagnetic Models



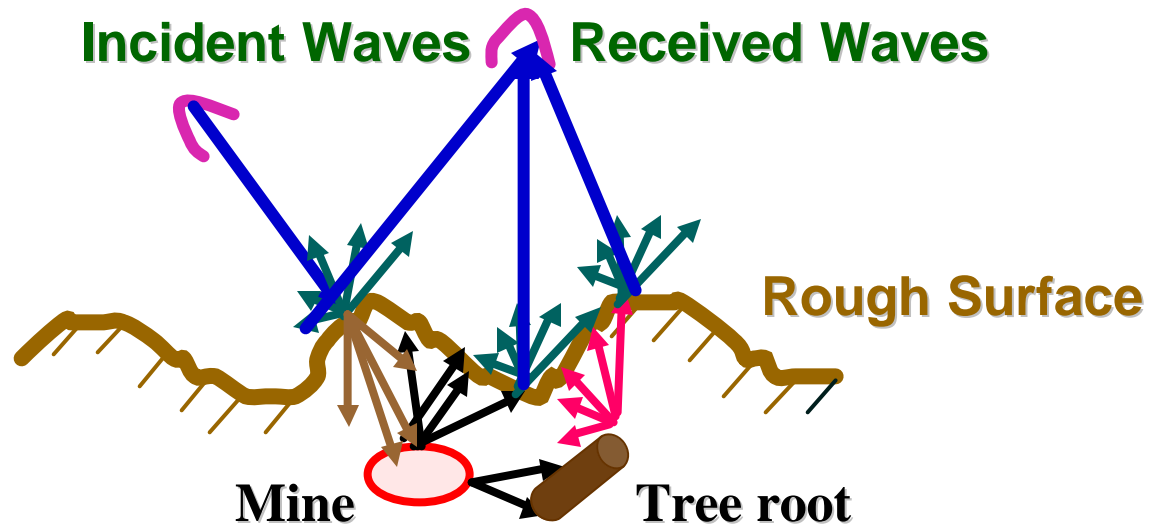
Model-Based Adaptive GPR Detection and Estimation



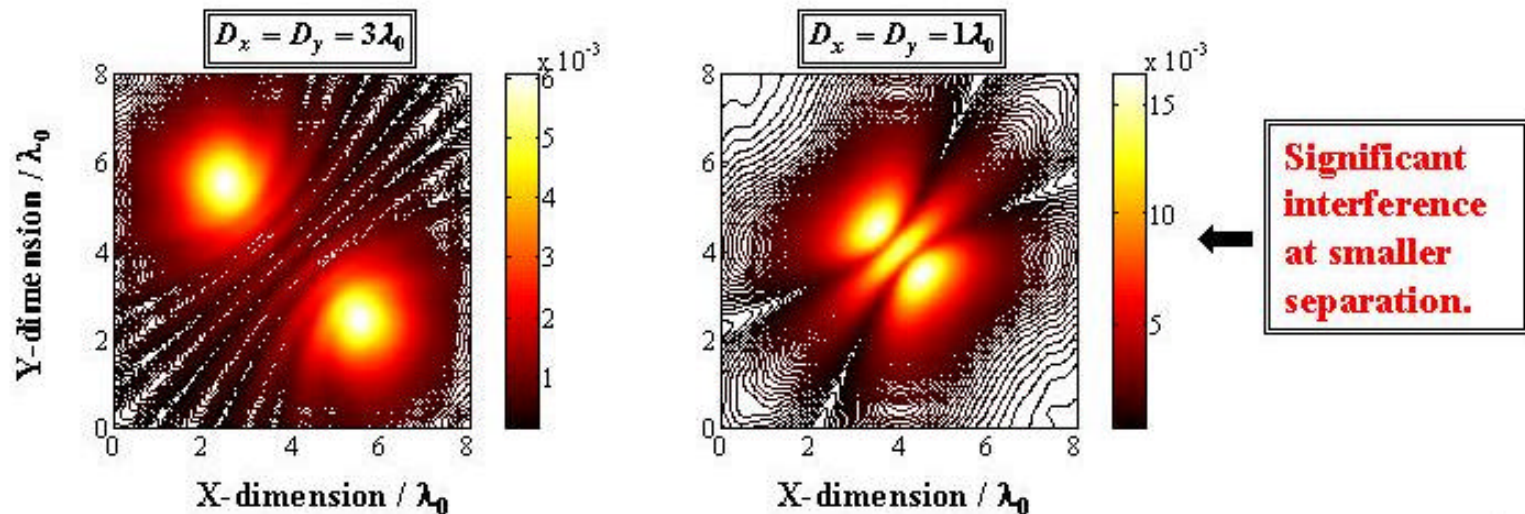
SDFMM for Two Penetrable Objects Buried under Random Rough Ground

“Steepest Descent Fast Multipole Method”

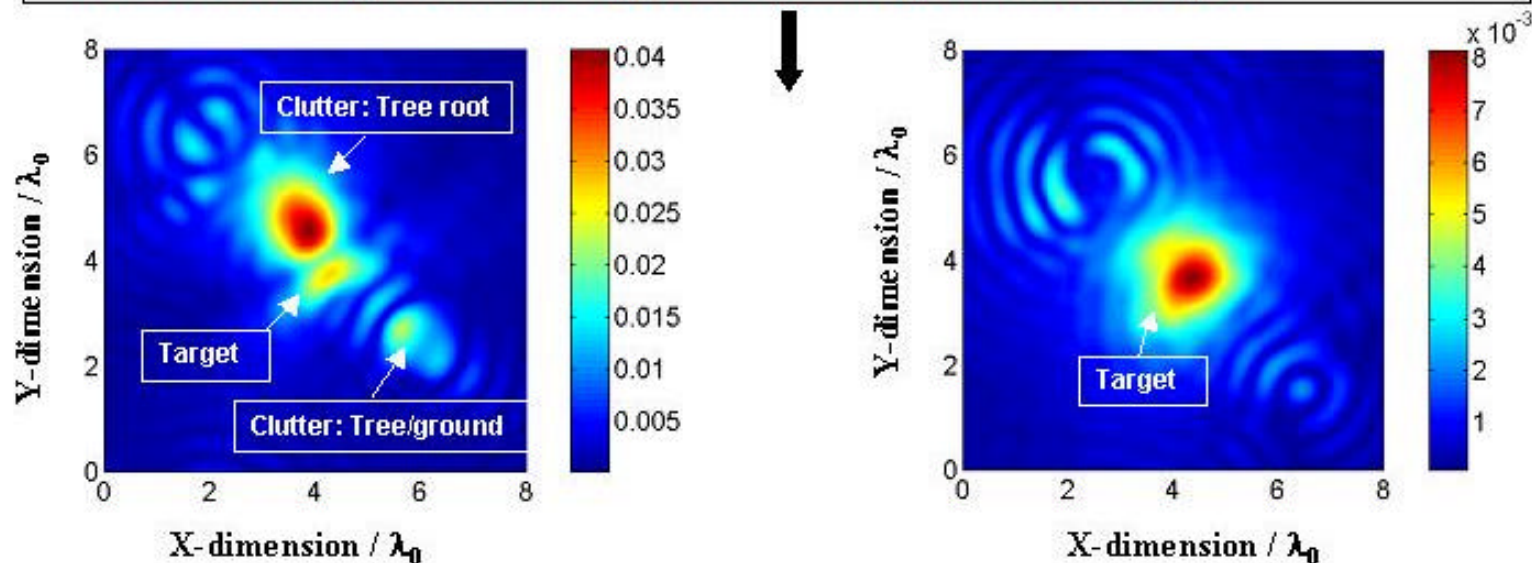
- 3-D Fast MoM Modeling
- Random rough ground.
- Near field computations.
- Target & Clutter objects.
- New integral equation formulations.
- Fast Multipole Method to solve for the unknown surface currents.
- Sources of Clutter: (1) Rough ground, (2) Second object.
- Parallel Implementation of the fast SDFMM computer code.
- Parametric investigations: target/clutter object shape, separation, location, multiple frequencies, Monte Carlo simulations.



Study of Mutual Coupling Effects of Multiple Buried Objects



Statistical average for clutter object is dominating that for the target: False Alarm.

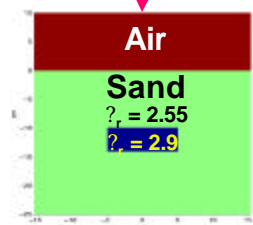


Finite Difference Modeling Predicts Detectability and Aids in Sensor Design

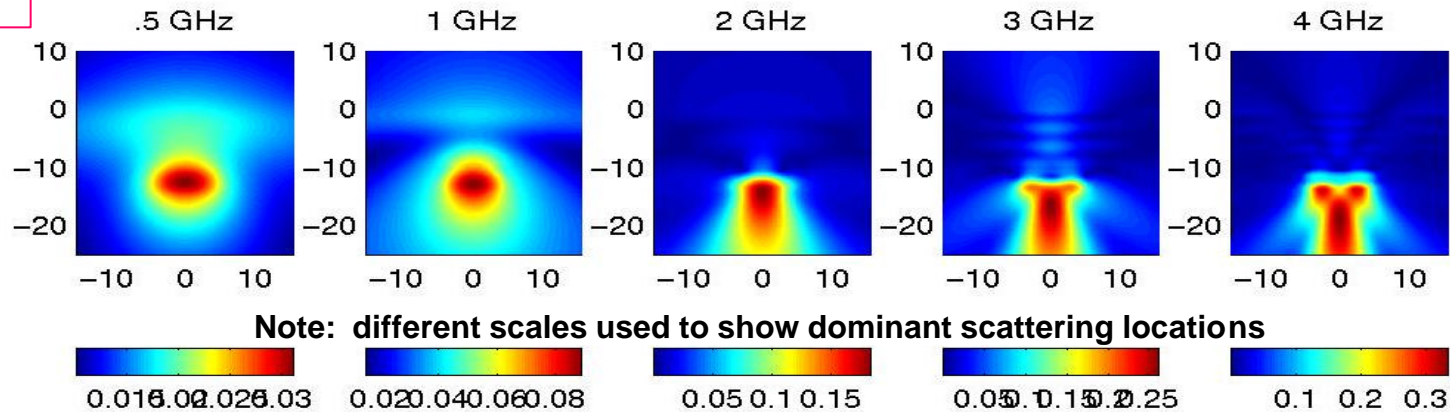
- **Frequency domain (FDFD) suggests features for target discrimination.**
 - Point source and plane wave excitations possible
 - Soil inhomogeneities can be accurately modelled (loose soil)
 - 3-D implemented using sparse techniques with 50X speedup over FORTRAN
- **Time domain (FDTD) assists in characterizing and filtering rough ground clutter effects.**
 - 3-D model validated with measured data for GeoCenters TEMR element on outdoor test track
 - FDTD model for smooth or average surface reflection used to minimize surface clutter and enhance plastic mine target

FDFD Simulation of Rectangular TNT Mine Buried 10cm in Dry Sand with Flat/Rough Ground Surface

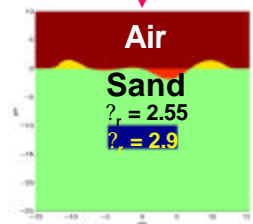
normally incident
TM plane wave



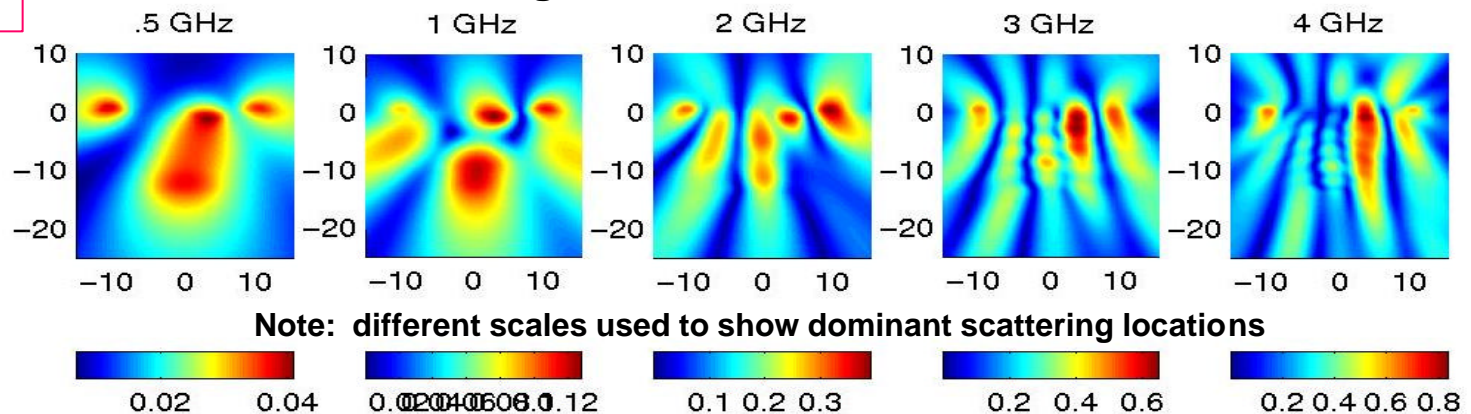
Smooth Surface Interface



normally incident
TM plane wave

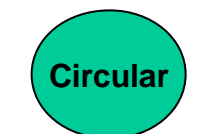


Rough Surface Interface



As resolving capacity increases, rough surface scattering
overwhelms target scatter

Time Domain Scattering Obtained from Stepped Freq. Model of Targets Buried 10cm in Sand



10 cm diameter

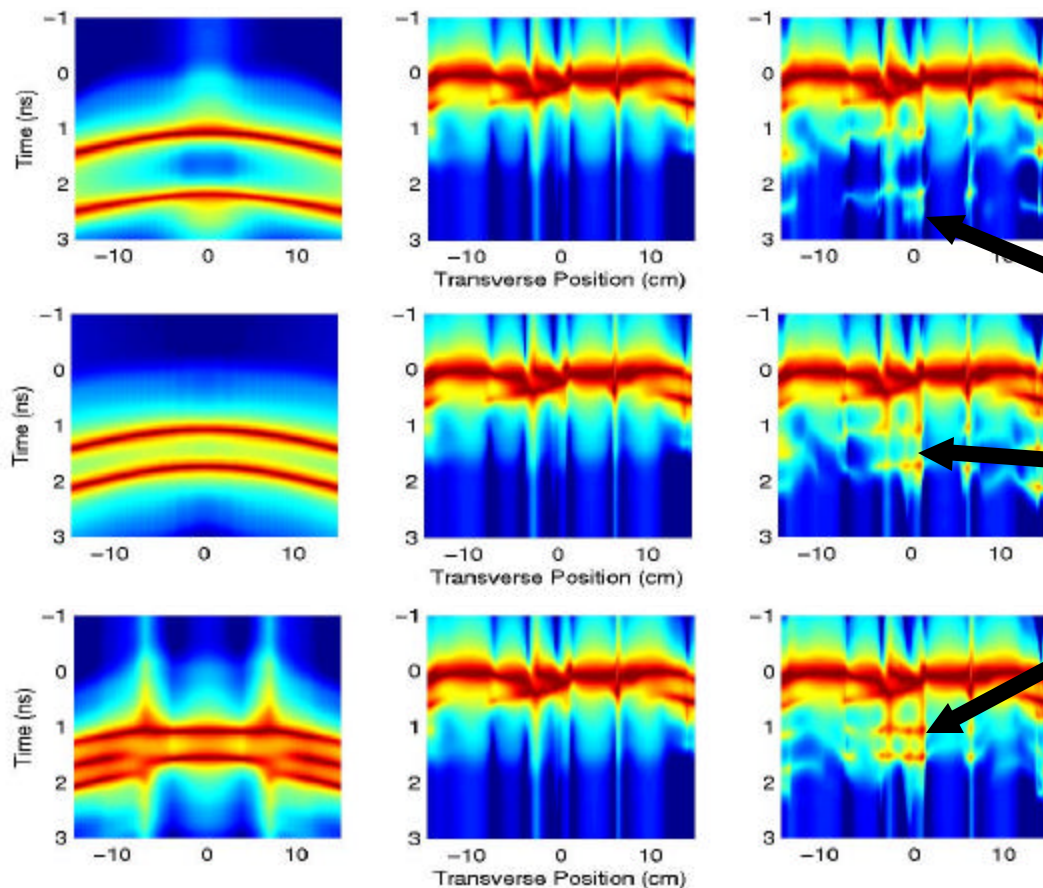


10 cm by 6 cm

Rectangular

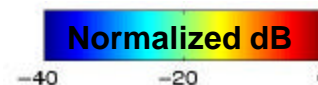
10 cm by 4 cm

Smooth Interface w/Mine Rough Interface only Rough Interface w/Mine



Note: Target features invariant to rough surface clutter

Time domain scattering results were generated using MUSIC algorithm applied to frequency domain scattering data received at the interface

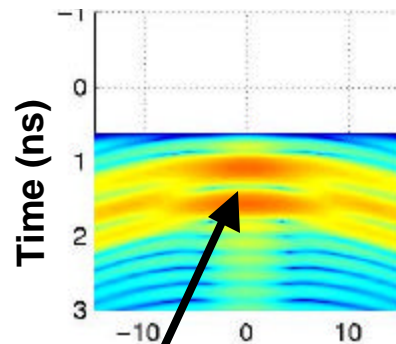


Time-Filtered Frequency Response Rectangle in Dry Sand

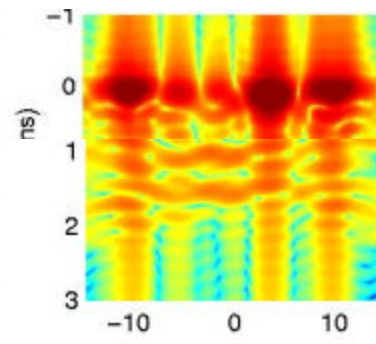
Target features are detectable in Time Domain, but are distorted by ground clutter

Target features

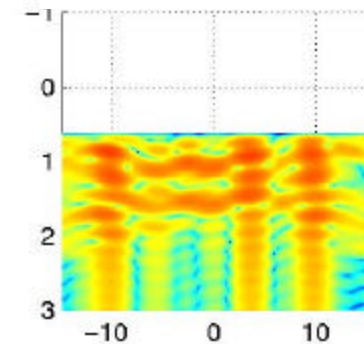
Smooth Surface, filtered



Rough Surface, no filtering

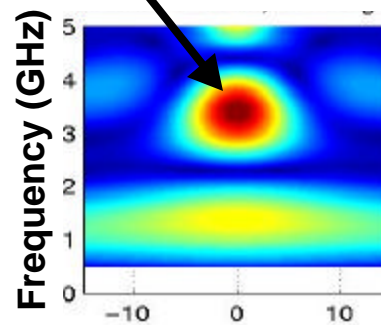


Rough Surface, filtered

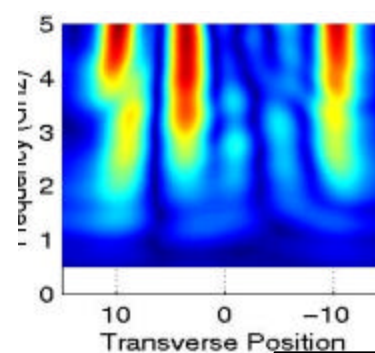


In Frequency Domain, full signal is dominated by ground surface clutter

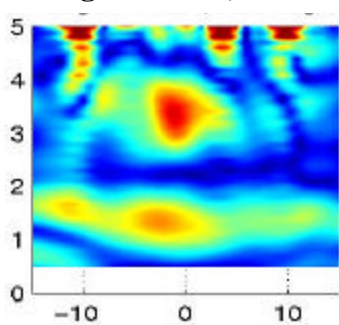
Smooth Surface, filtered



Rough Surface, no filtering



Rough Surface, filtered



Removing early time signals recovers wideband frequency response of target

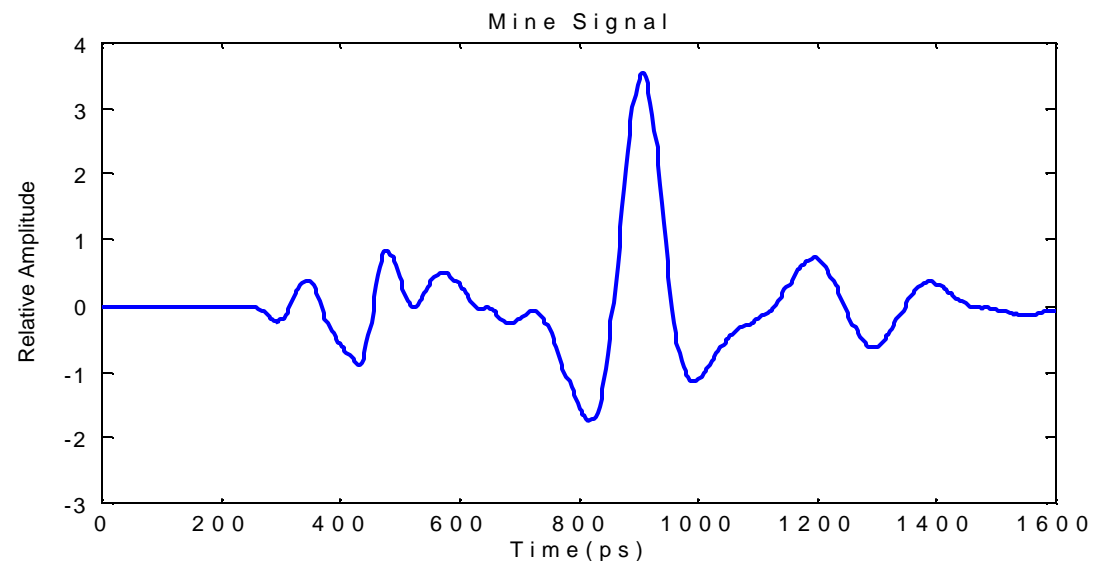
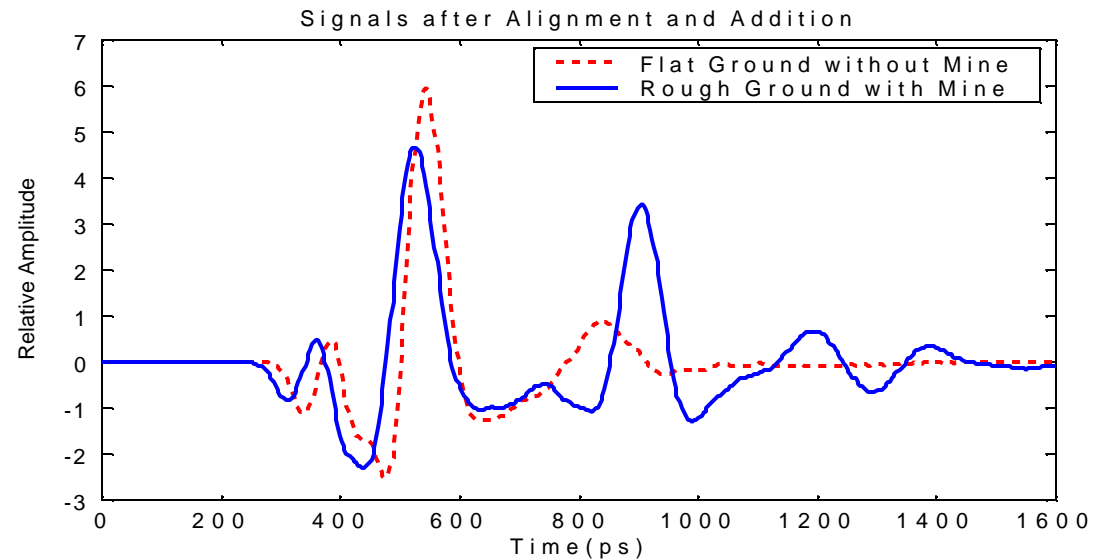
FDTD Ground Surface Clutter Identification and Removal

Computed signals for:

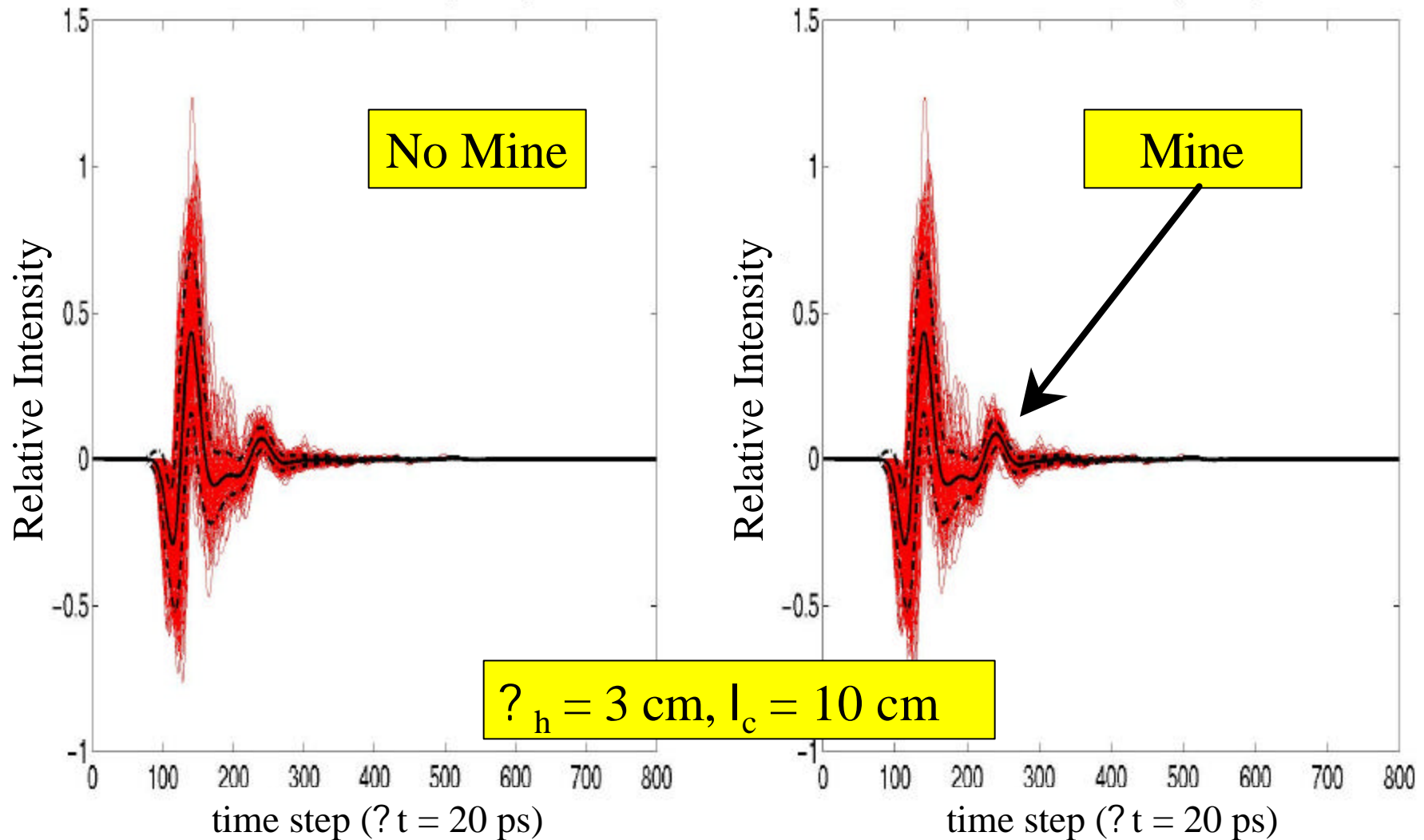
Flat Ground with
No Mine

Rough Ground with
Mine -----

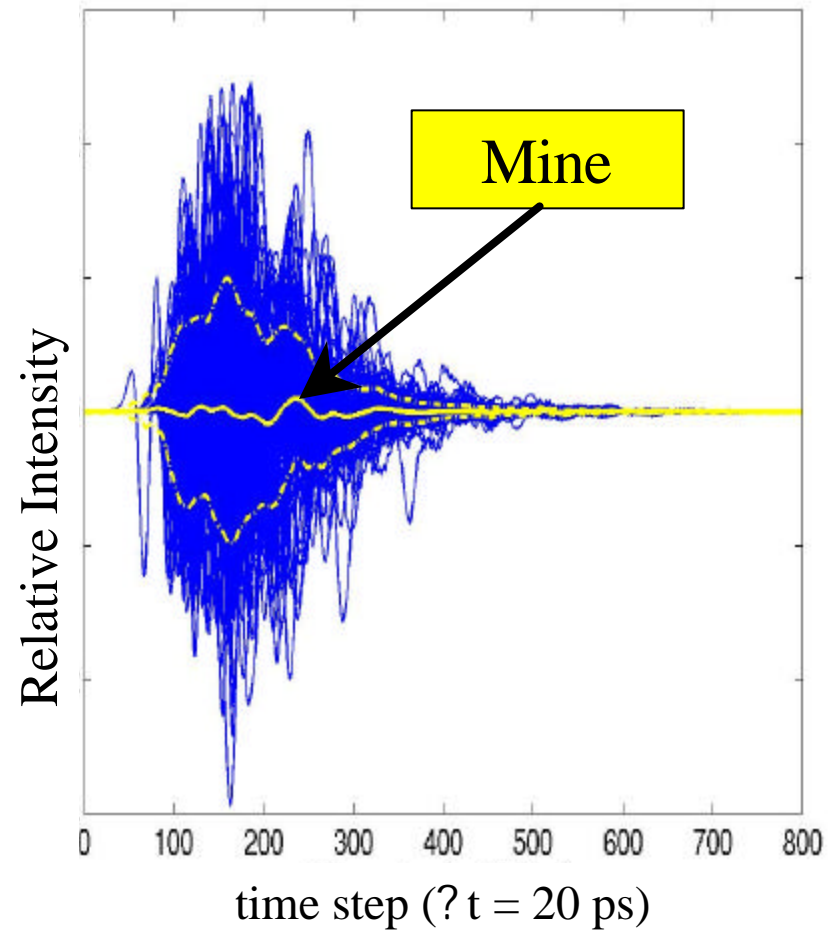
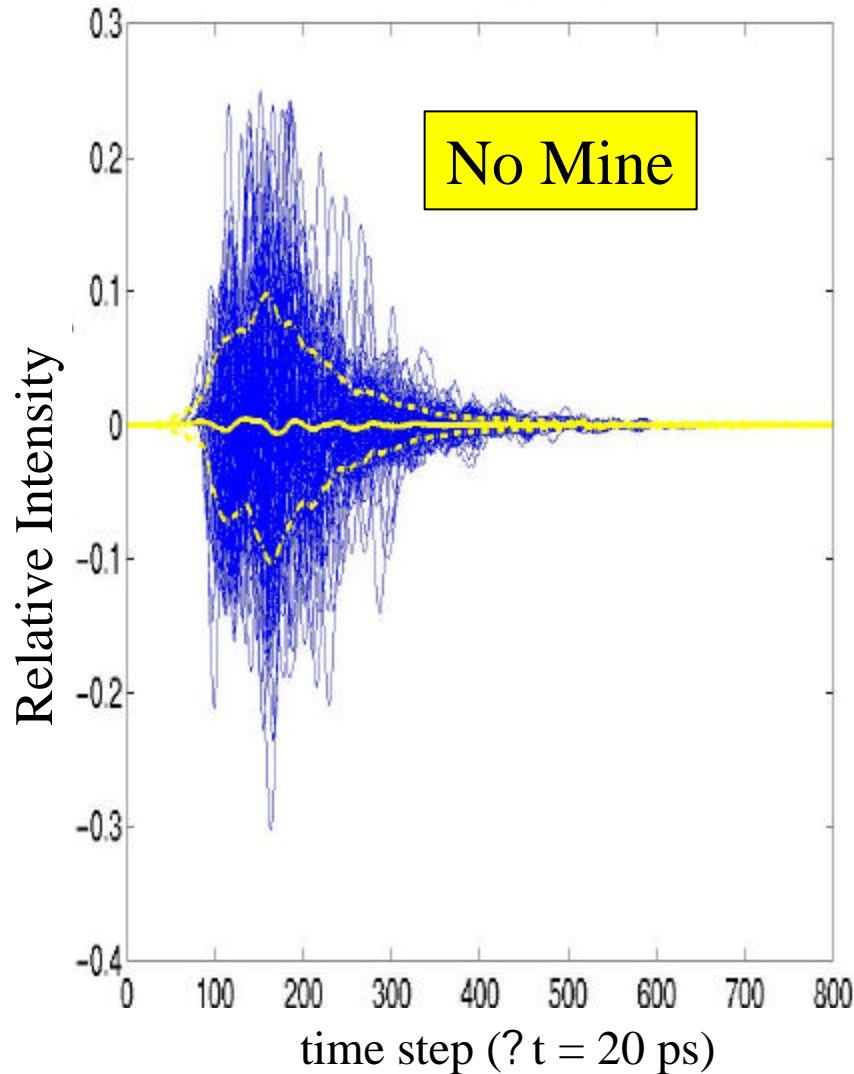
Computed Mine Signal
subtracting scaled and
shifted Flat Ground
with No Mine Signal
from Rough Ground
with Mine Signal



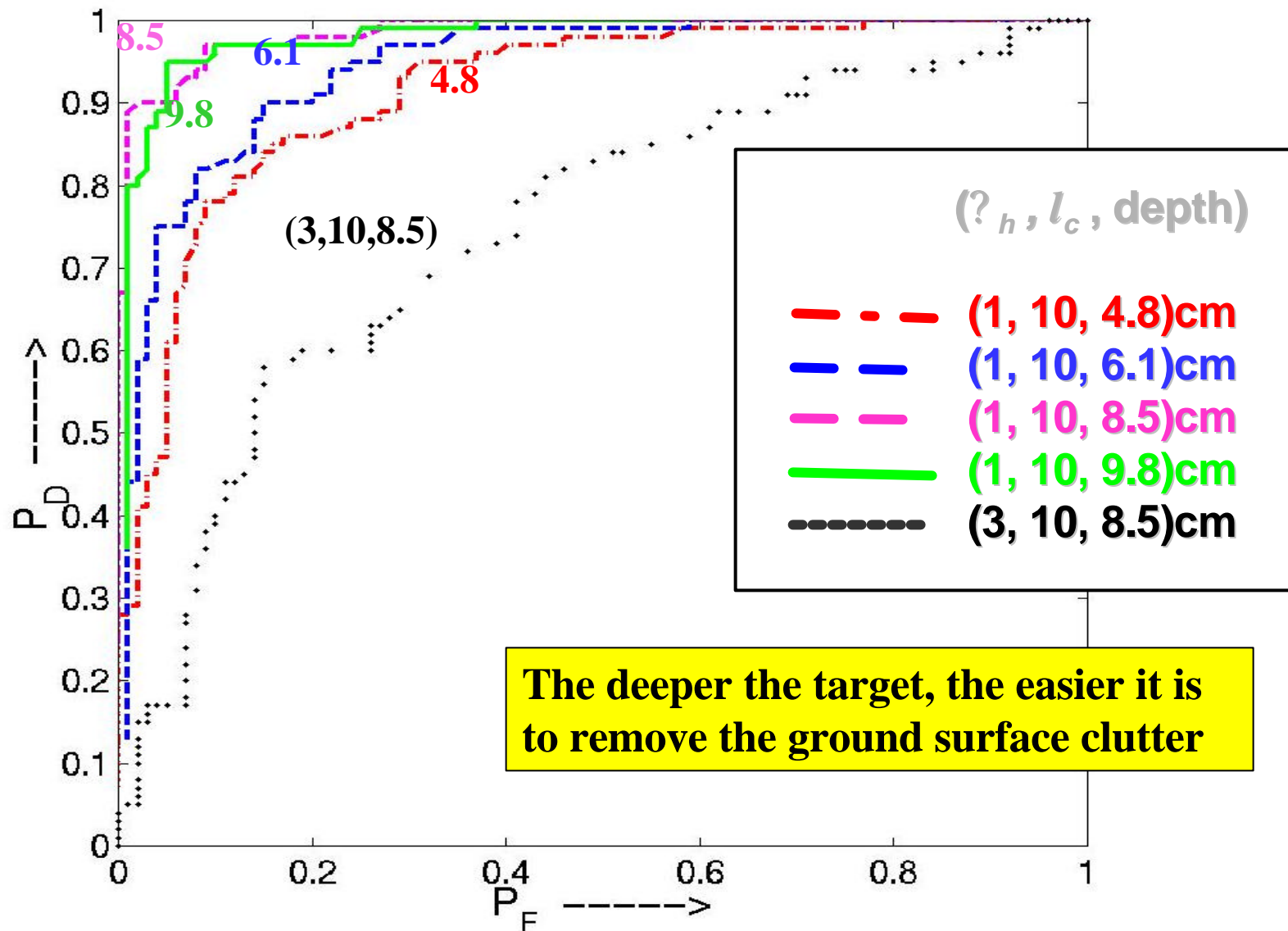
Original Signal Averages Obscure Mine Signal



Identifying and Removing Ground Surface Signal & Realigning Makes Target Signals Stand Out



ROC Curves for Various Target Depths



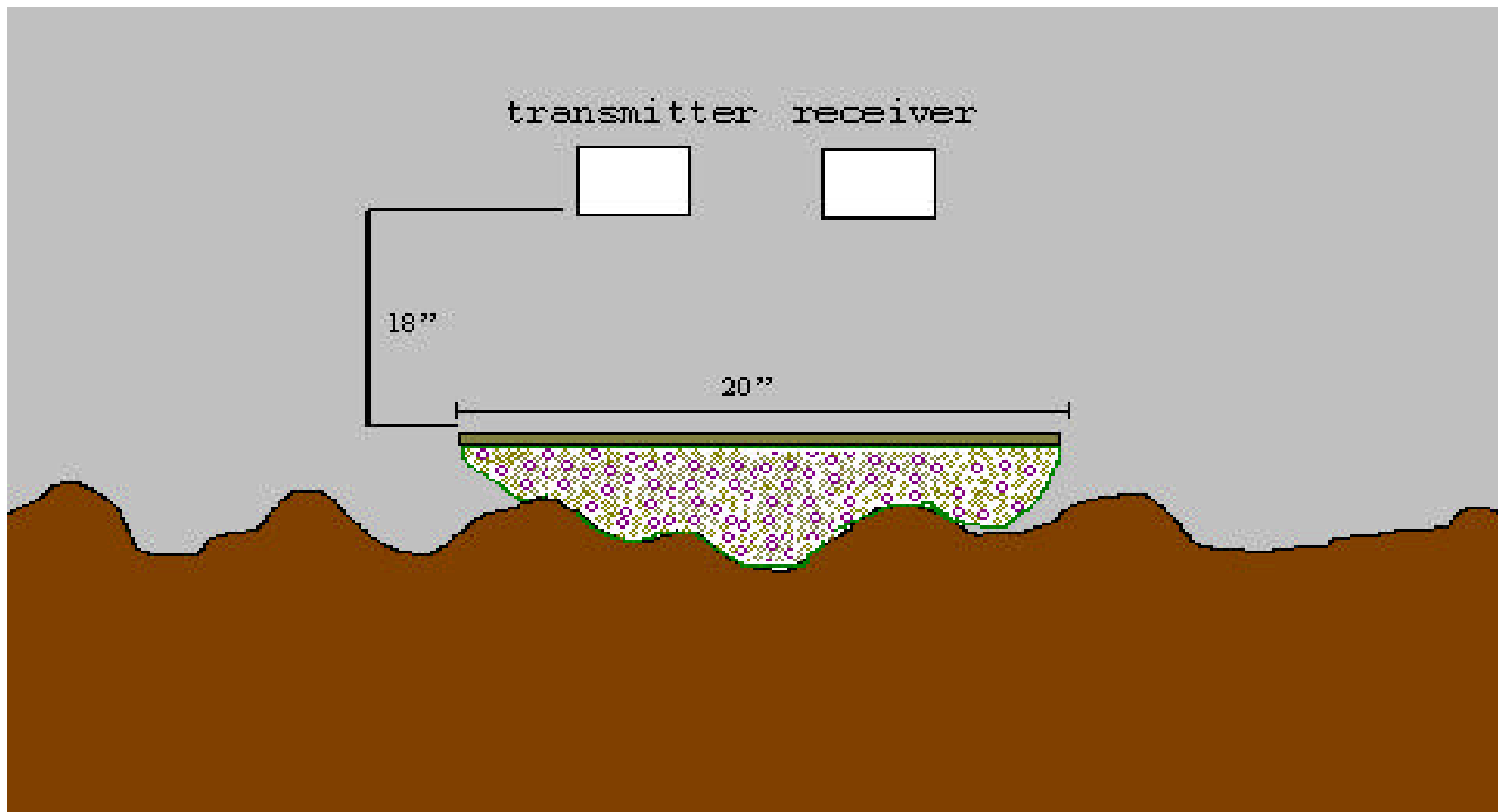
Latest Modeling Accomplishments

- **Validated 2-D and 3-D modeled time domain clutter suppression on measured 3-D data with real targets**
- **Developed fast 3D FDFD preconditioner**
- **Established feature database for FDTD model**
- **Extended SDFMM for multiple buried objects**
- **Parallelized FDFD, SDFMM**
- **Analyzed multistatic focusing advantages**

Novel Experimental Systems

- Artificial Dielectric for ground surface matching
- Antenna confirmatory probe
- Hyperspectral Polarimeter for surface mine detection
- Microwave Enhanced Infrared Thermography with dual frequency discrimination

Conformable Artificial Dielectric Device to “Smooth Out” Rough Ground Clutter



Volume: 6mm dielectric beads ~ 2L
8mm plastic beads ~ 2L

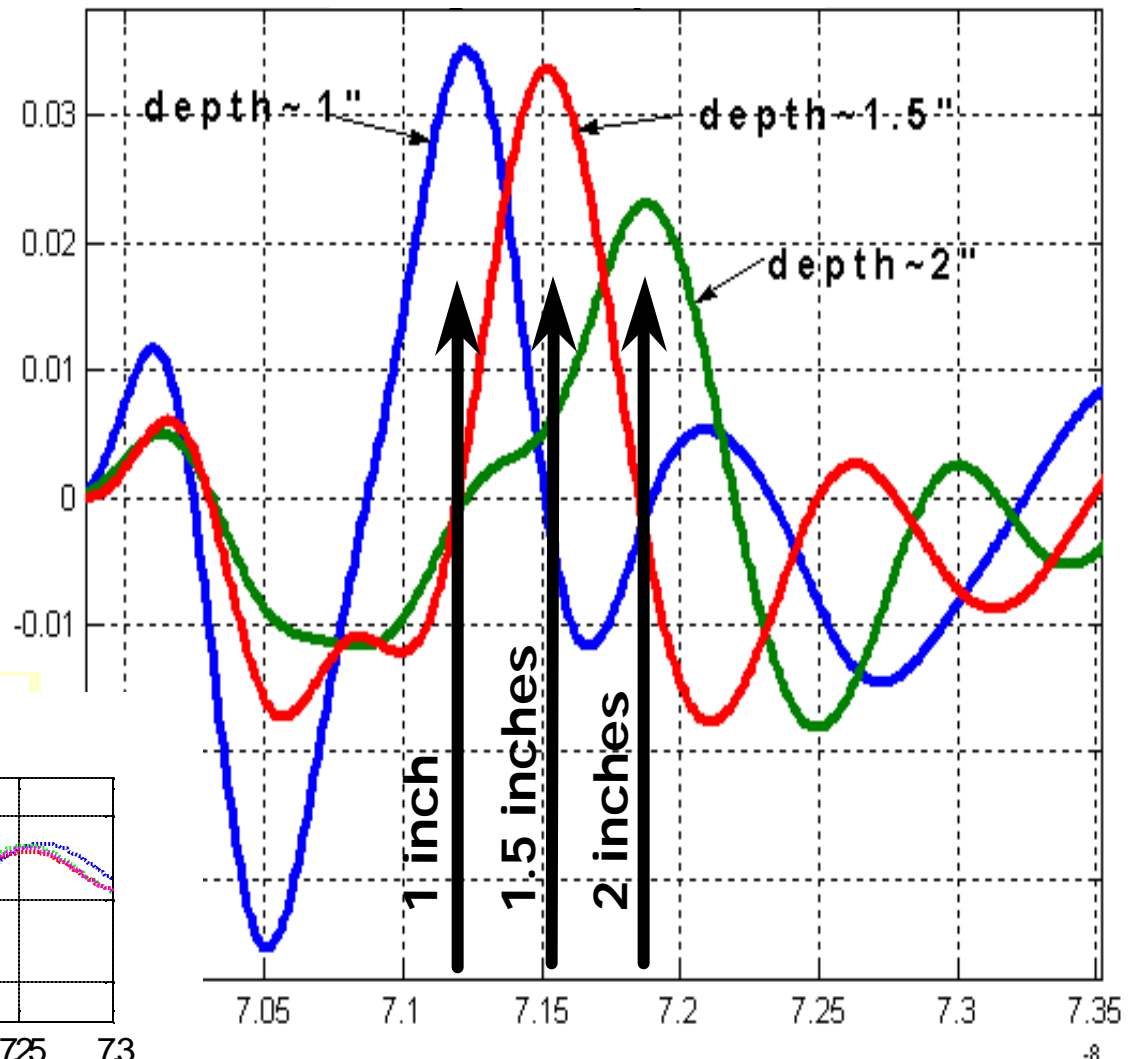
Artificial Dielectric Filling Rough Ground Surface Troughs



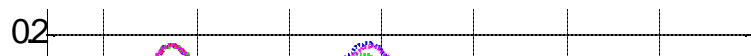
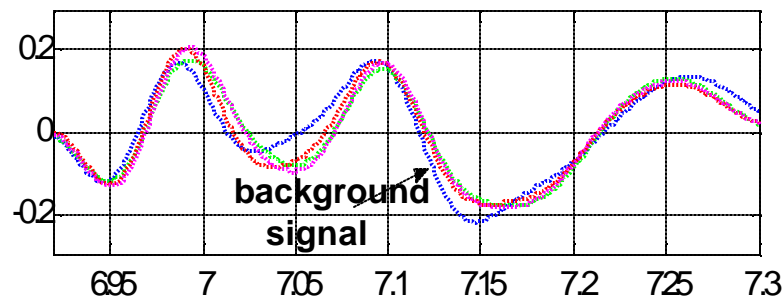
Artificial Dielectric Aids Detection for Targets at Various Depths

AD used to remove rough ground surface clutter shows target responses for nonmetallic AP mines buried at 1", 1.5" and 2"

Target Response



Total Response

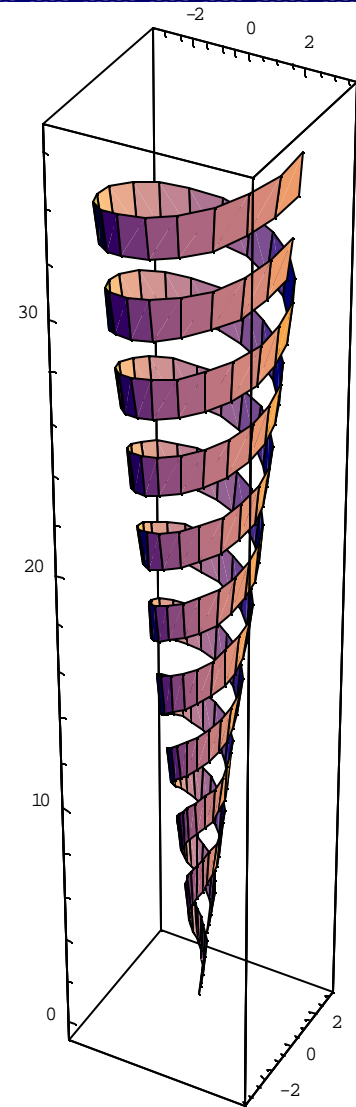


Experimental Results for Conducting Bead Artificial Dielectric Matching Layer

- Lightweight Artificial Dielectrics can be made easily and inexpensively
- Wide range of soils permittivities can be matched with Artificial Dielectrics
- Response from rough ground smoothed with AD:
 - almost as good as undisturbed soil
 - just as good as if holes refilled with original soil
- Target signals greatly enhanced using AD
 - Rough ground clutter effectively suppressed
 - Target depth accurately determined

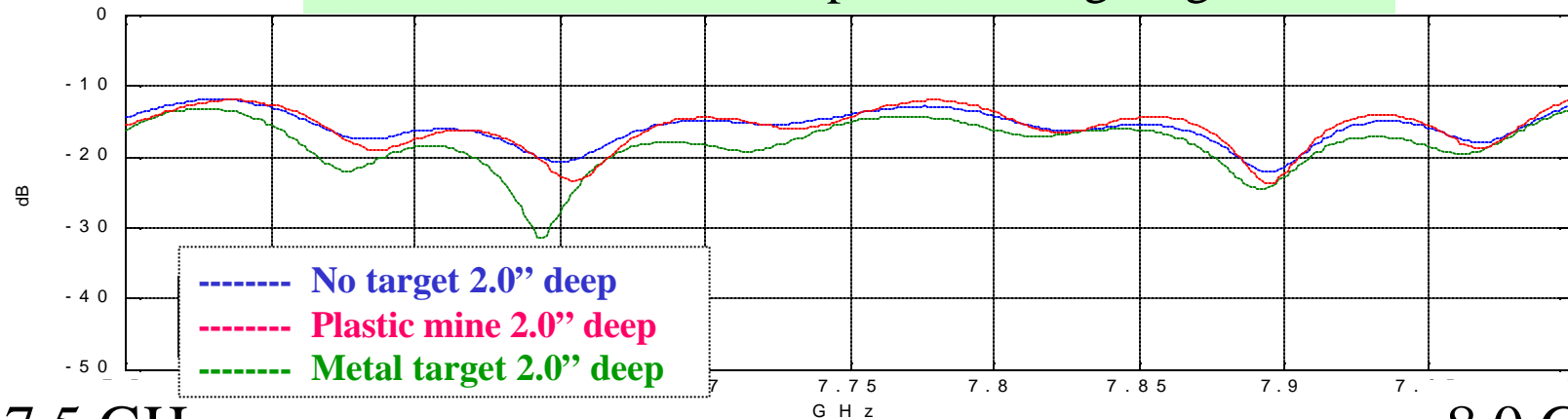
Subsurface Ultra Wideband Radar Probe Confirmatory Sensor

- Use frequency independent **Conical Spiral** antenna, adapted to lossy soil as probe inserted near target
- Illuminate target with 0.5 to 8 GHz circularly polarized wave
- Detect wideband spectral response with same, second, or above surface antenna
- Observe spacing between peaks in FFT signal which depend only on target composition and geometry



2 cm Monopole Antenna in Moist Soil: Measured Sensitivity to Metal/Plastic Target

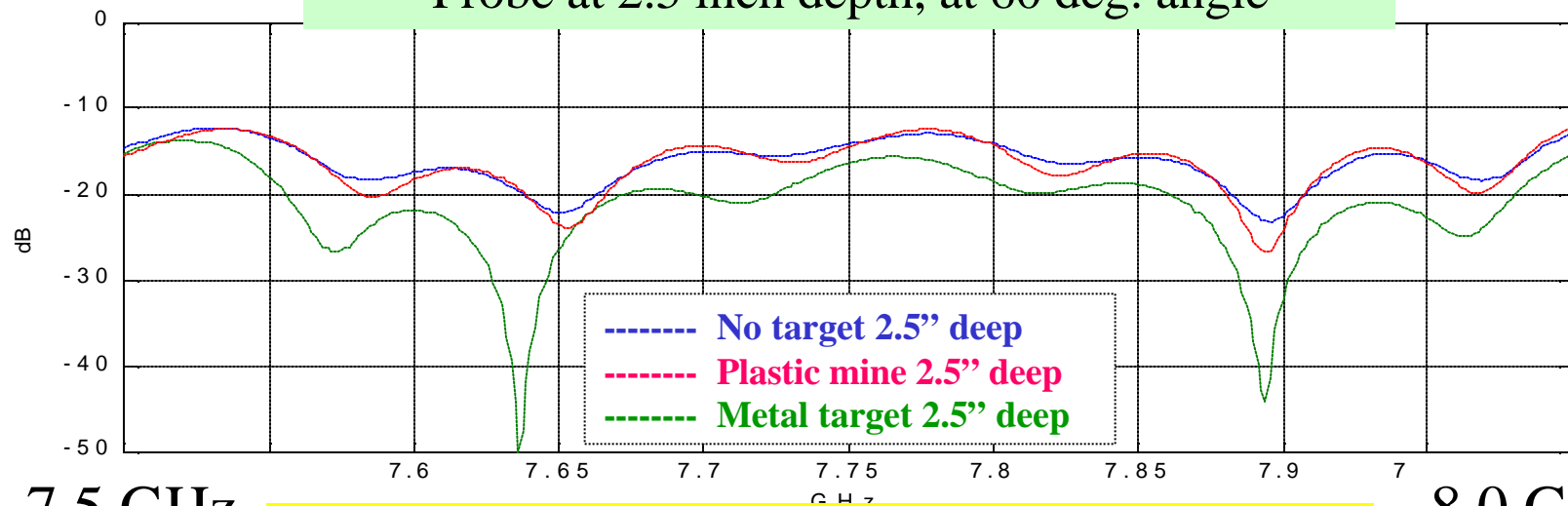
Probe at 2 inch depth, at 60 deg. angle



7.5 GHz

8.0 GHz

Probe at 2.5 inch depth, at 60 deg. angle



7.5 GHz

8.0 GHz

Strong signal dependence for probe near targets

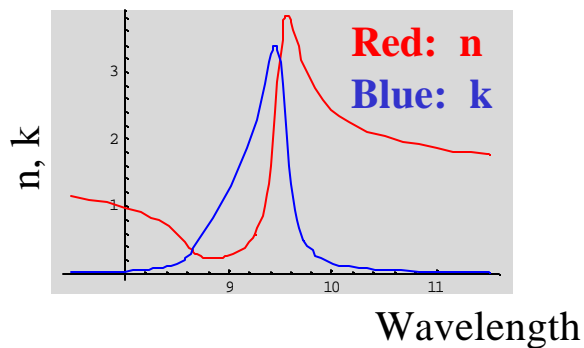
Hyperspectral Polarimetry:

Unique Target Discriminant Measurable Only By Spectrally-Resolved Polarimetry

Spectral Behavior of *Refractive Index*
Across a (Lorentzian) Resonance



Spectral Behavior of *Fresnel Reflectance*
Across a Resonance



Fresnel γ_s (perpendicular)

Fresnel γ_p (parallel)

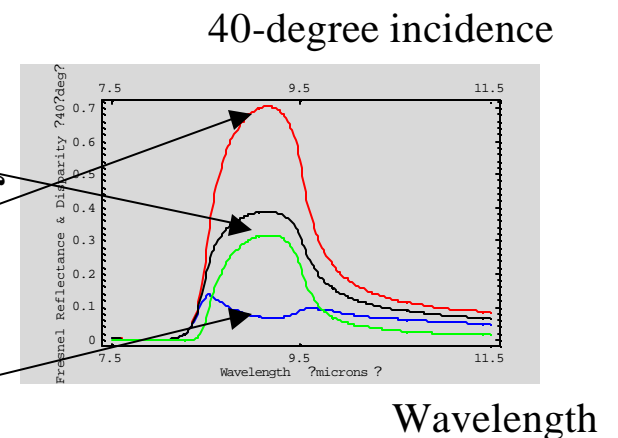
Uni-modal avg.

reflectivity: $(\gamma_s + \gamma_p)/2$

Bi-modal *polarized*

disparity: $(\gamma_s - \gamma_p)/2$

Reflectivity



The discriminant within the Stokes spectrum:

- (a) bi-modality of percent polarization
- (b) coincidence with uni-modal intensity

Discriminant is a Joint Fingerprint of
(a) Material Type and (b) Smoothness

Confirming Data from Visible Band Imaging Hyperspectral Polarimeter*

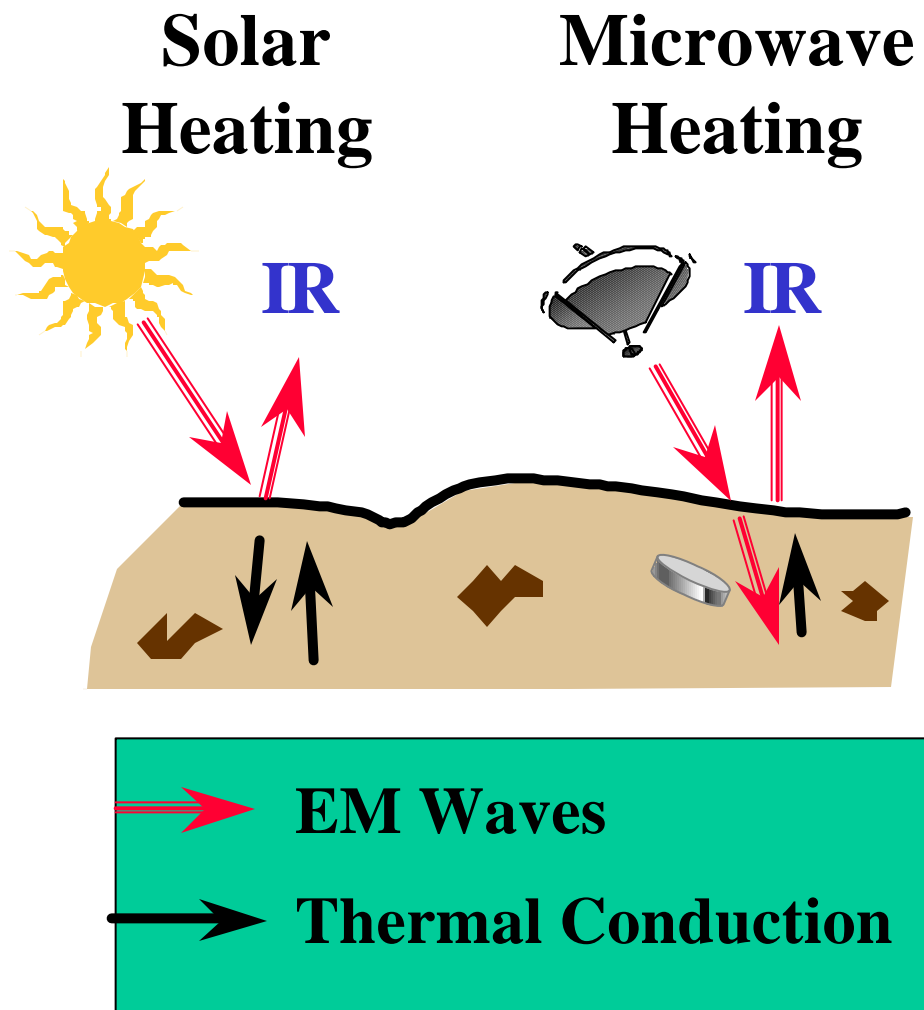
Current Results

- **LWIR implementation proved too costly, so visible band selected for demonstrating effectiveness of approach, including Aerodyne's key concepts**
 - Polarization Spectral Intensity Modulation** (Patent Pending)
 - Attribute Diversity
- **Projected Performance Demonstrated**
 - Eigenvalue analyses of initial data, albeit limited in range of conditions, indicate that polarization enhanced spectral is many times more effective than spectral only for discriminating man made objects in clutter.
- **Benefits of approach realized**
 - Complete Stokes Vector extraction
 - Perfect registration and simultaneous attribute measurements
 - No moving polarization elements required

*Implementation support from Air Force Research Laboratory

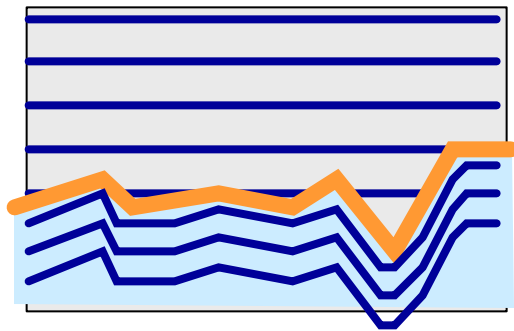
Motivation for Microwave Enhanced IR Thermography (MEIT)

- **Solar Ground Heating**
 - 1000 W/m² Sufficient
 - Dawn/Dusk Required
 - Minimal Penetration
- **Microwave Heating**
 - Comparable Power
 - No Sunny Day, Dawn/Dusk Constraint
 - Variable Penetration



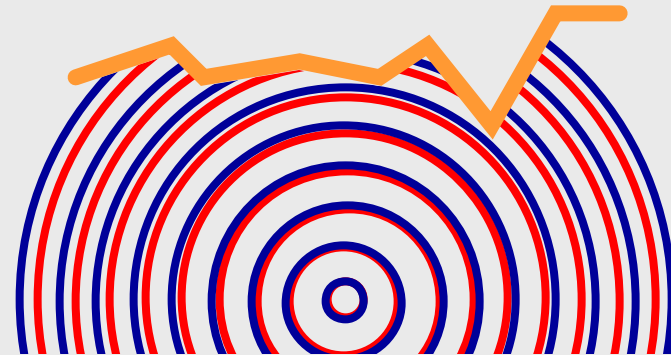
Two-Frequency Microwave Enhanced IR Heating Concept

- Surface Effects Similar at Both Frequencies
- Propagation Effects Different



Incident Fields

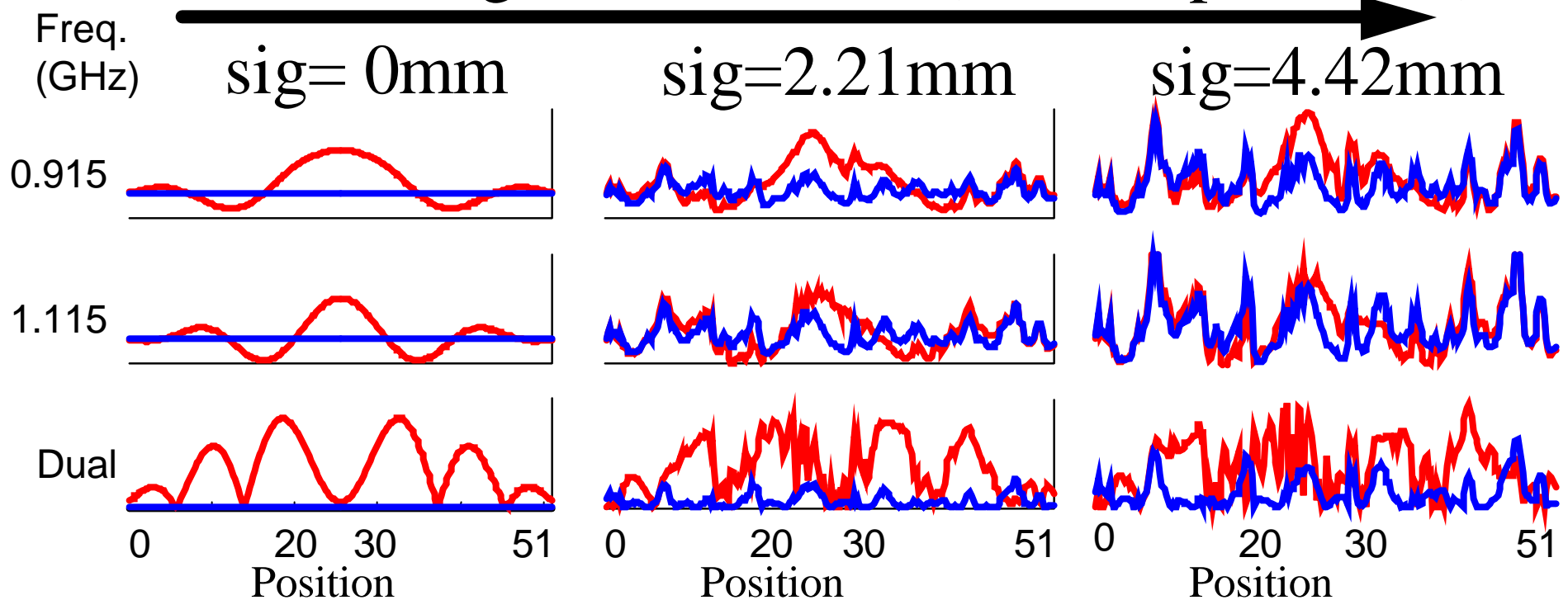
(Similar at Both Frequencies)



Scattered Fields

Two Frequency Microwave Power Distribution along Rough Surface

Surface Roughness increase (fixed depth=6 cm)



Red: mine

Blue: no mine

GPR/EMI Experiments at the NEU Dedham Test Track

- Accessibility
 - 10 miles from NU, BU
 - Ability to take data in different track/weather conditions
 - Quick feedback on hardware/software development
- Convenience
 - Built-in tracking platform
 - Fast set-up
- Customizability
 - Control of targets, clutter objects
 - Create specific target characteristics
- Comparison of multiple modalities on identical track conditions

Track Targets

Location	Target	Type	Depth
36' 9"	EM-3	Simulant	2"
41' 0"	VS-2.2	AT	2"
44' 0"	Screw	Clutter	3"
48' 6"	Valmora 69	AP	2"
51' 6"	VS-1.6	AT	2"
55' 6"	EM-6	Simulant	2"
59' 0"	Screw & Rock A	Clutter	1 ½", 3"
63' ½"	M-14	AP	1"

Clutter/False Alarm Objects

- Simple clutter:
isolated rocks and screws
 - Rocks=radar false alarm
 - Screw=EM Induction false alarm
- Complex clutter:
screws in proximity of rocks



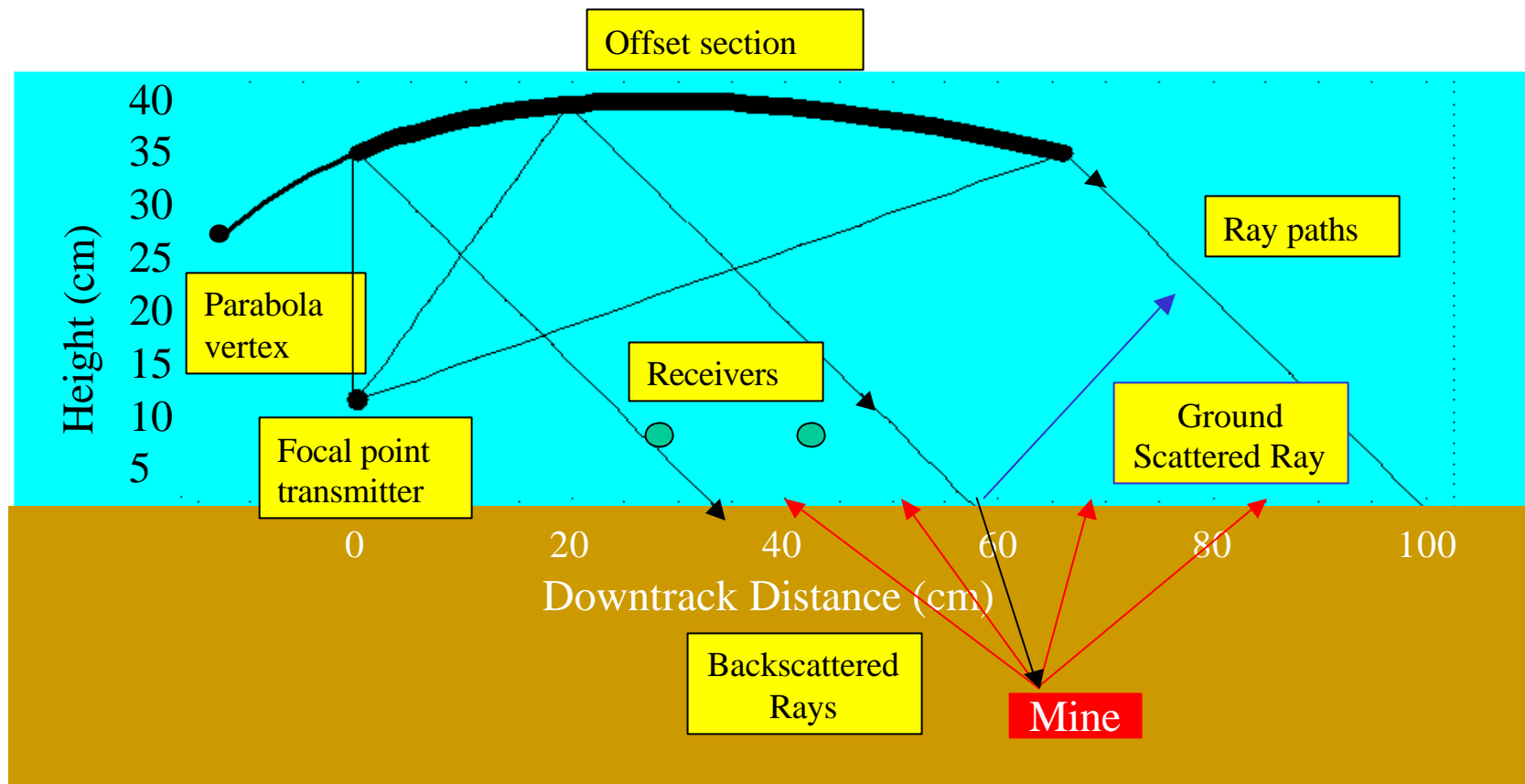
Multistatic Ground-Penetrating Radar

- Parabolic reflector antenna for forward-directed plane wave
- Four detectors for back-scattered radiation
- High-bandwidth (1.7 GHz) pulsed sources
- Scans of track C/L, +/-3", +/-6", -9", -12"

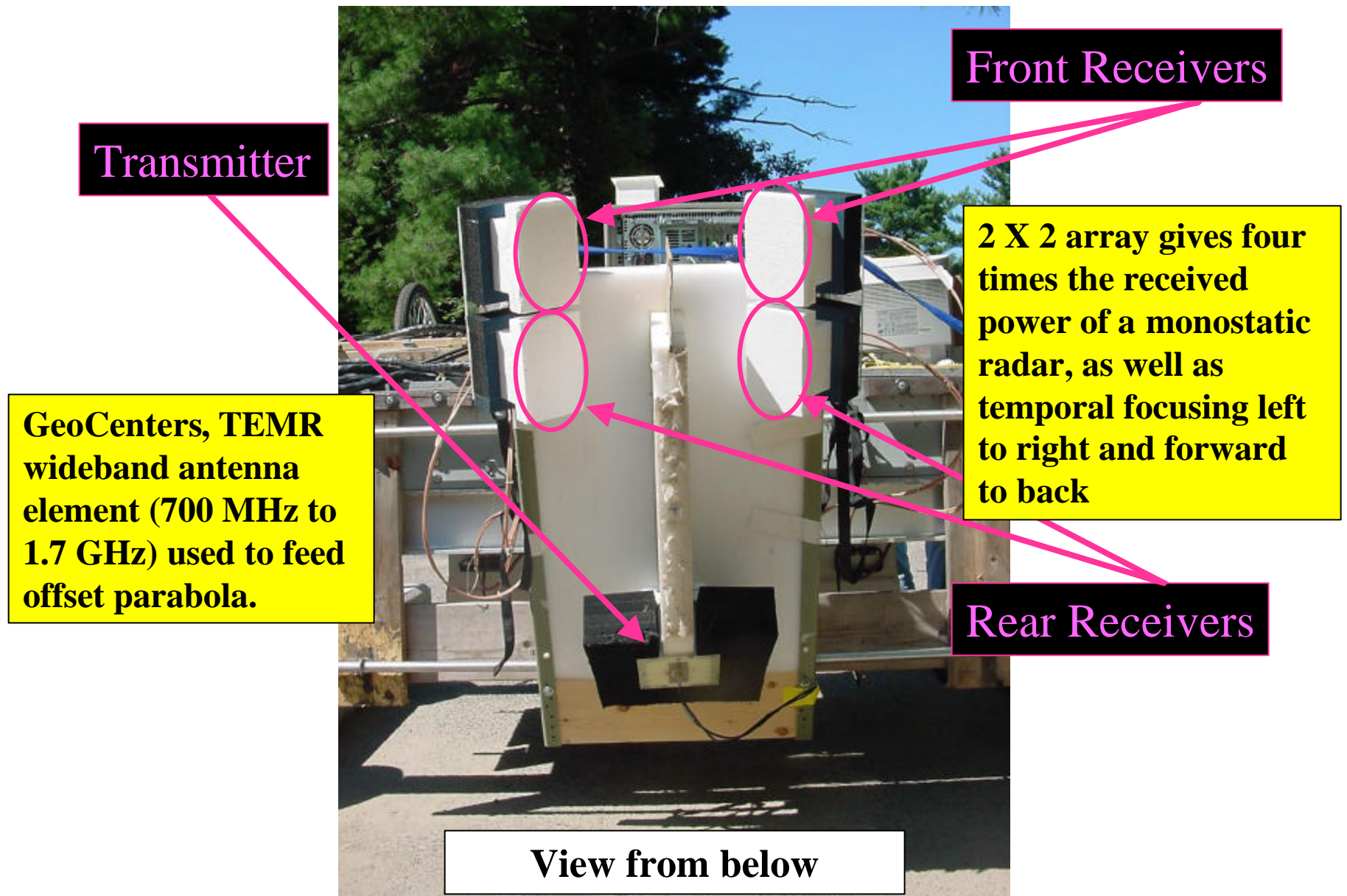


Novel Forward-Looking ‘Handheld’ AP Mine Detector with Planar Incident Beam Excitation

Offset Paraboloid Ray Paths



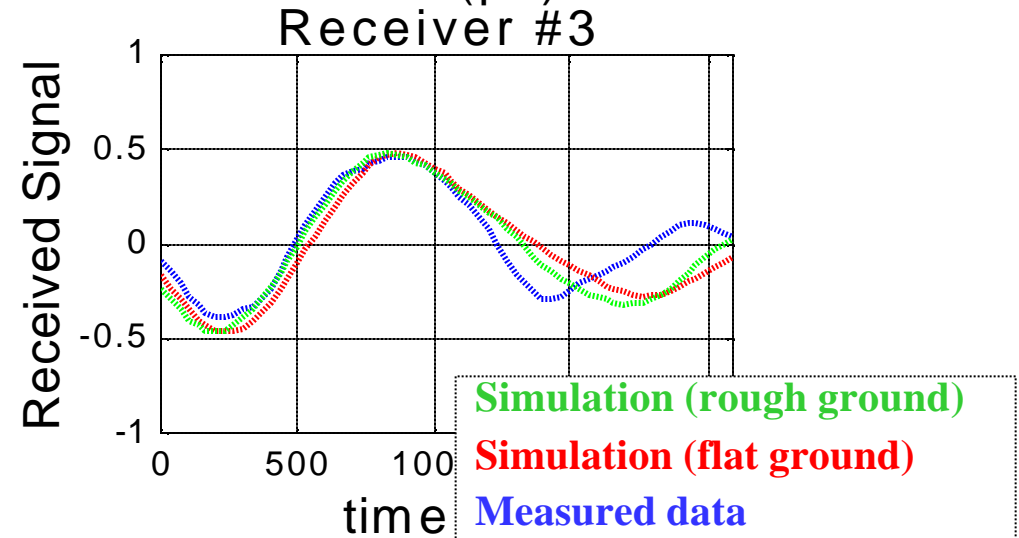
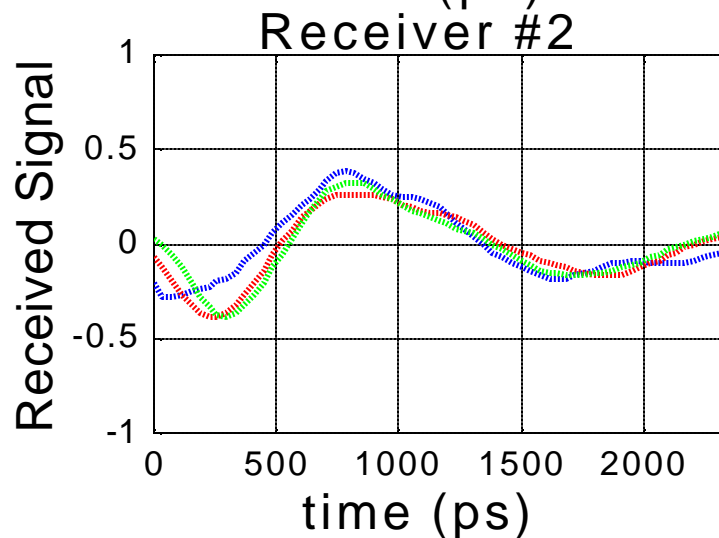
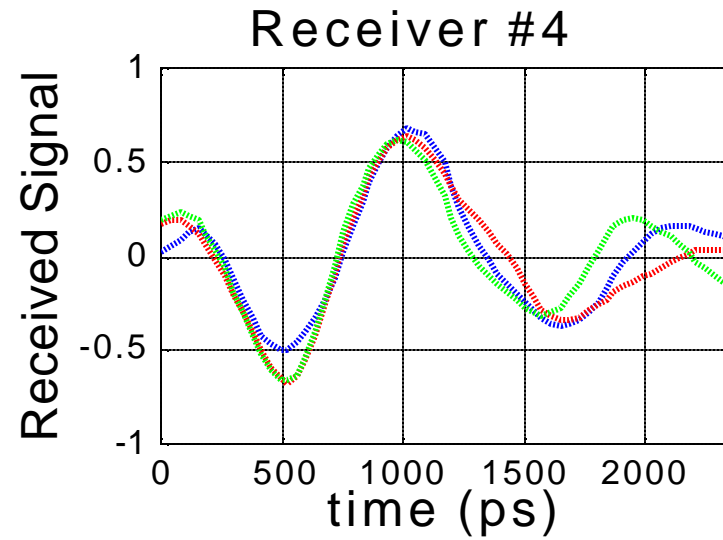
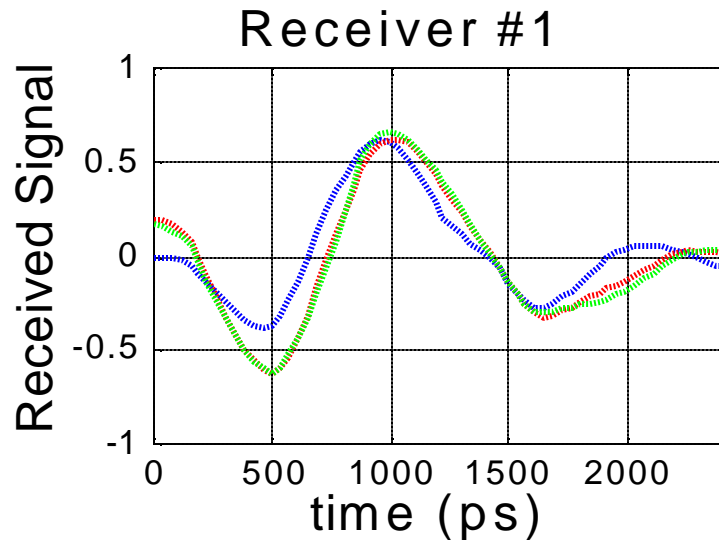
Ground-Penetrating Radar: “Air Shot”



Radar in Operation at NEU Dedham Test Track



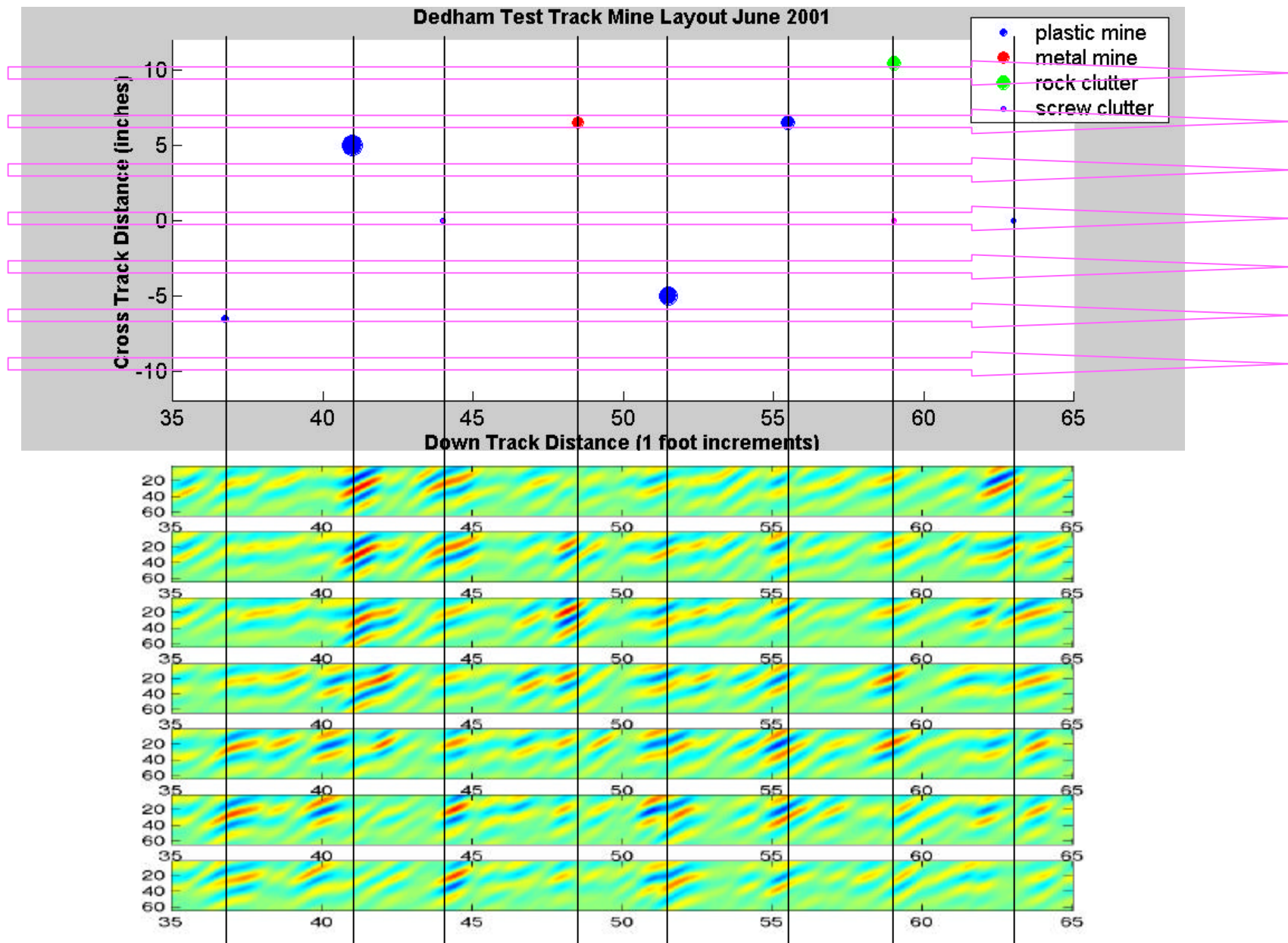
Comparison of Measured GPR Data with Flat and Rough 3-D FDTD Simulations: Ground reflection & target signal



GPR Signal Processing: Programmatic Overview

- **Basic philosophy:**
 - Combination of tractable physical models with sophisticated, statistical signal processing methods to make intellectually interesting advances that address real problems
- **Algorithmic work**
 - Recursive detection methods for array GPR data based on sequential hypothesis testing and modified ANOVA technique
 - Fusion of physics-based backprojection (F-K Migration) with min-entropy image restoration work to obtain robust GPR SAR auto-focus method for object localization
- **Validation: all on real data**
 - Detection method: NU and EG&G data
 - Localization: NU, DeTeC, NITTEK data
- **Continued effort work via Night Vision Forward Looking Program**

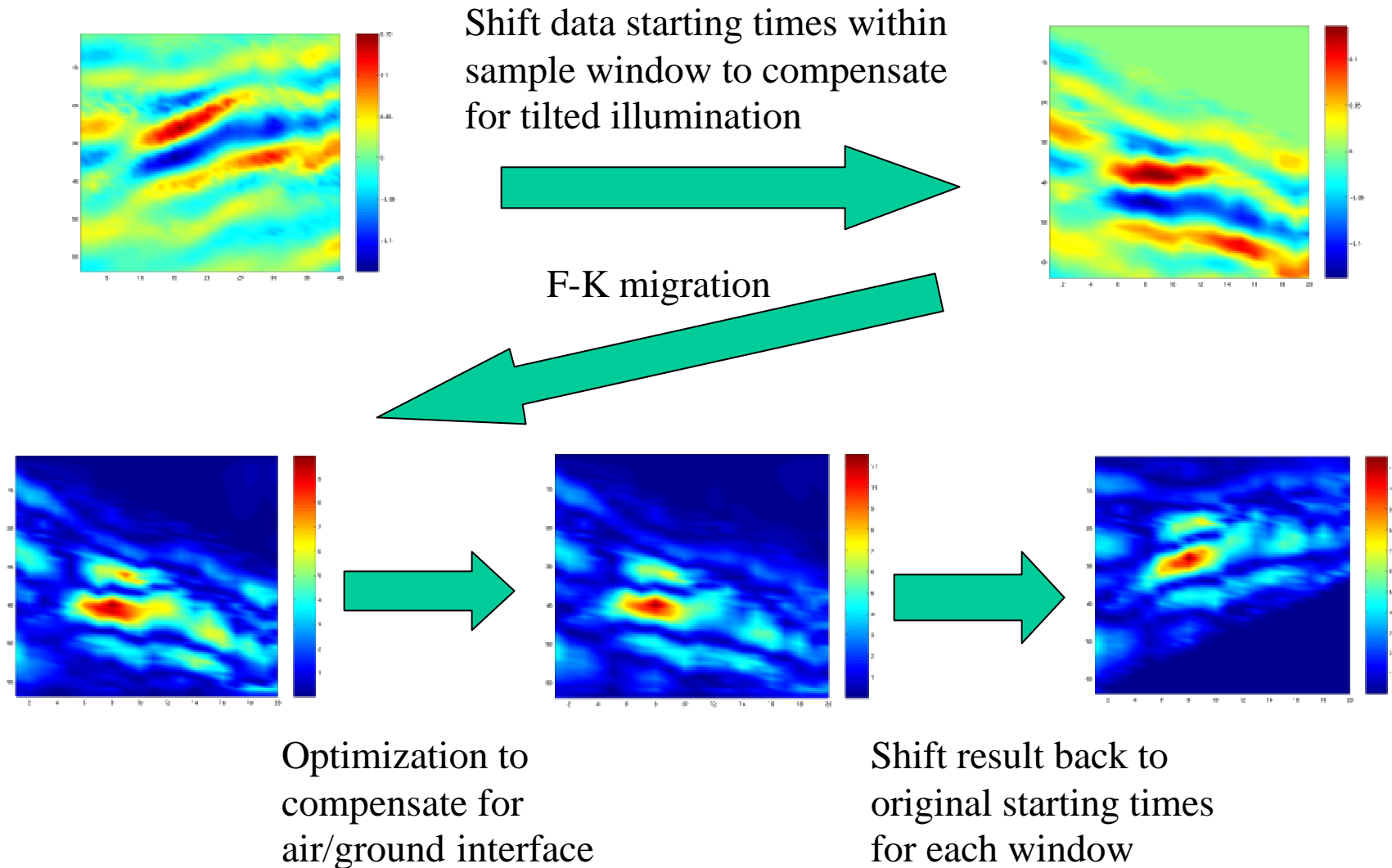
Matched Filter Processed GPR Signals



GPR Signal Processing: Accomplishments

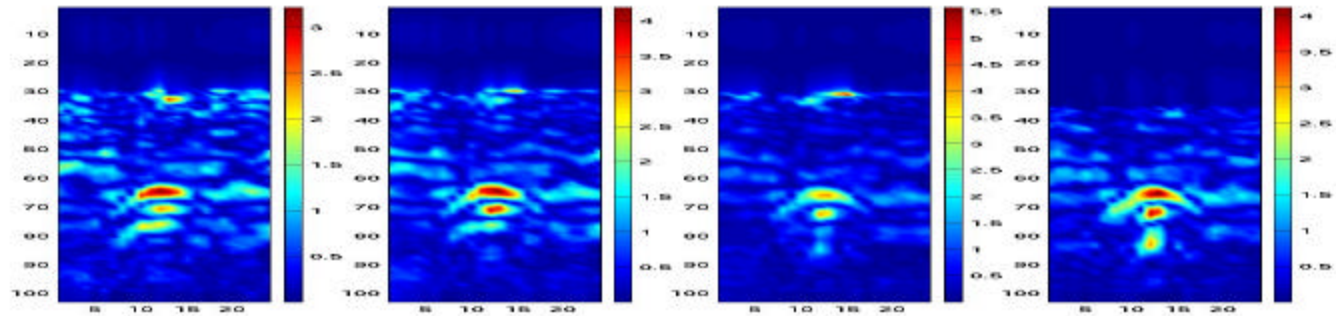
- **Basic problem:**
 - Localization of buried targets from multi-monostatic and multistatic GPR data
- **Approach:**
 - Novel fusion of simple physics and solid image restoration in the form of entropy-regularized, F-K migration technique
- **Results:**
 - Strong localization ability demonstrated on real data sets from multiple sources (NU, NIITEK, DeTeC).

Measured NEU Data Processing with Optimized F-K Migration: PMN AP Plastic Mine

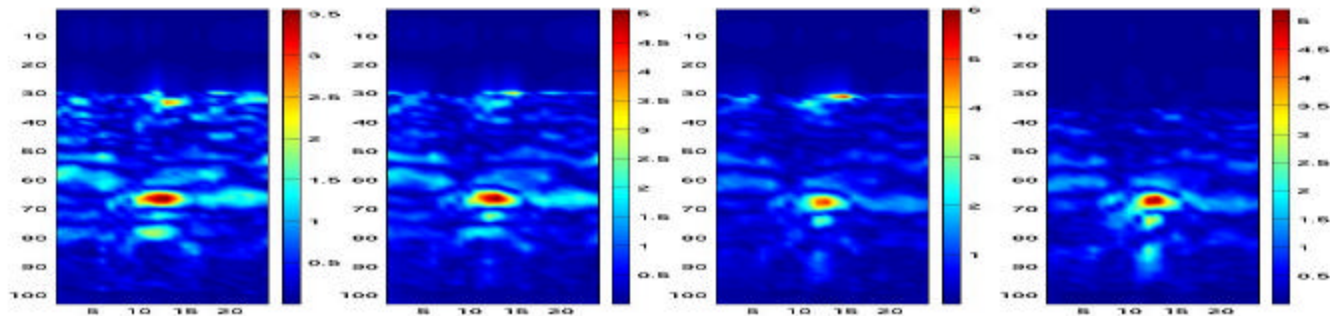


Measured Data: Wichmann Radar for M19 target

FK
result



Opt'ed
FK result

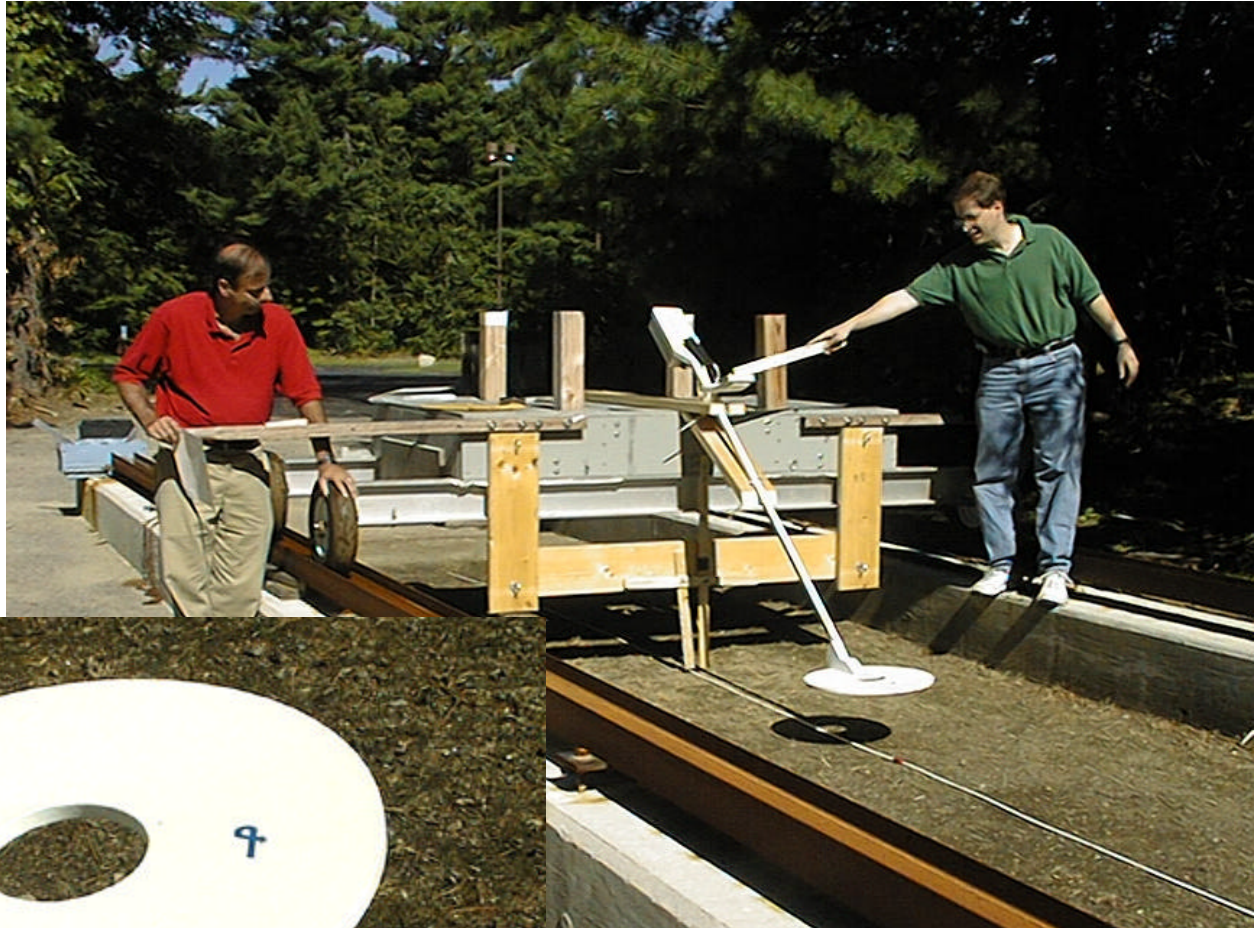


EMI: Programmatic Development

- **Philosophy:**
 - Simple, parametric physical models combined with “robustifying” processing schemes to arrive at novel methods for solving real problems
- **Algorithmic advances**
 - Optimal processing of data collected as a function of time, space, and frequency
 - Recursive detection methods designed to function progressively as the data are acquired
 - Batch data approach to object characterization robust to unknown orientation and location of object
- **Proof-of-concept level validation on real sensor data**
- **Transition to UXO problem via SERDP SEED grant**

EMI Track Scan

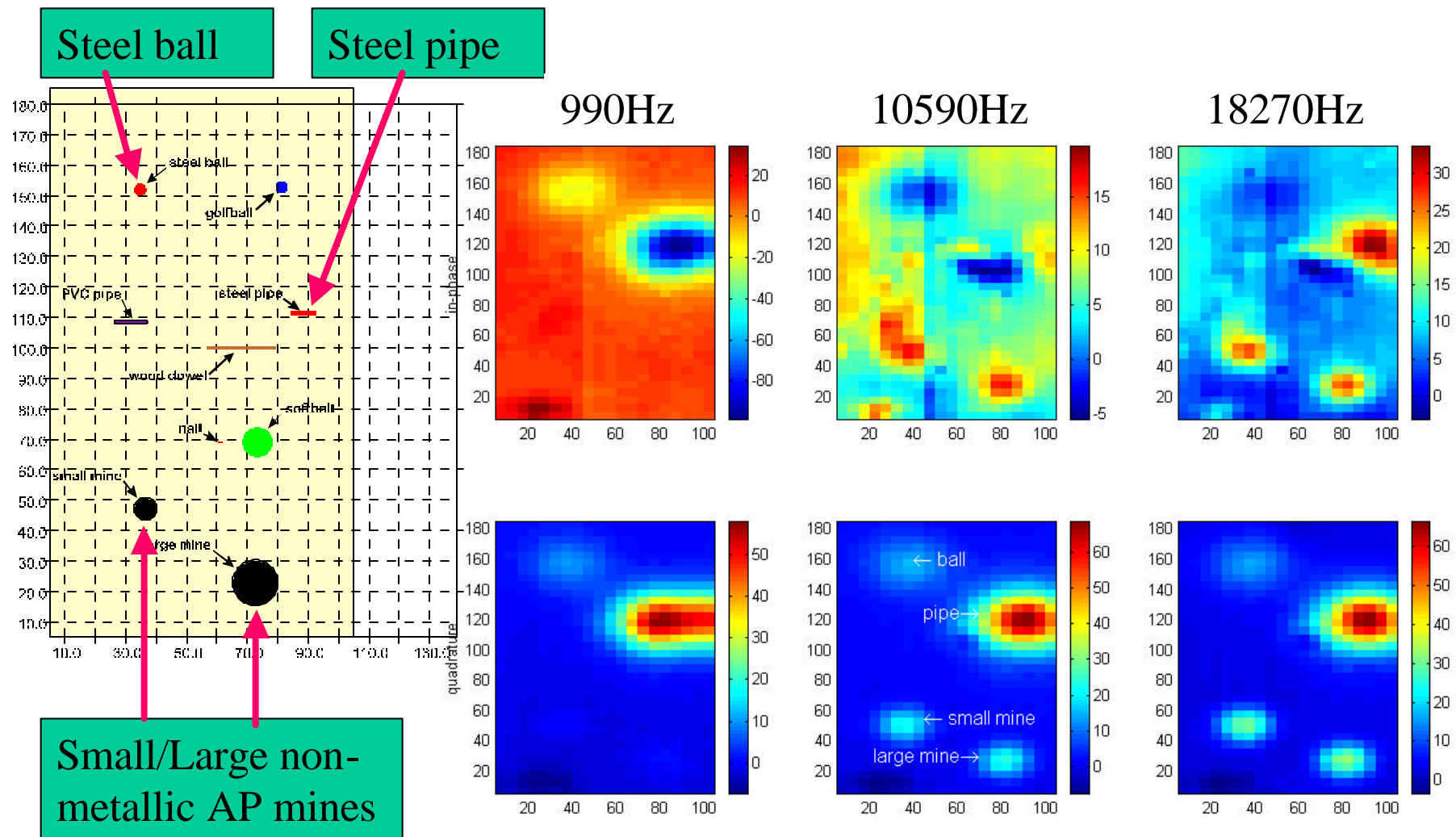
- Geophex GEM-3 EMI sensor
- Ten frequency bands sampled
- Full track scans at 6" resolution
- Detail target scans at 3" resolution



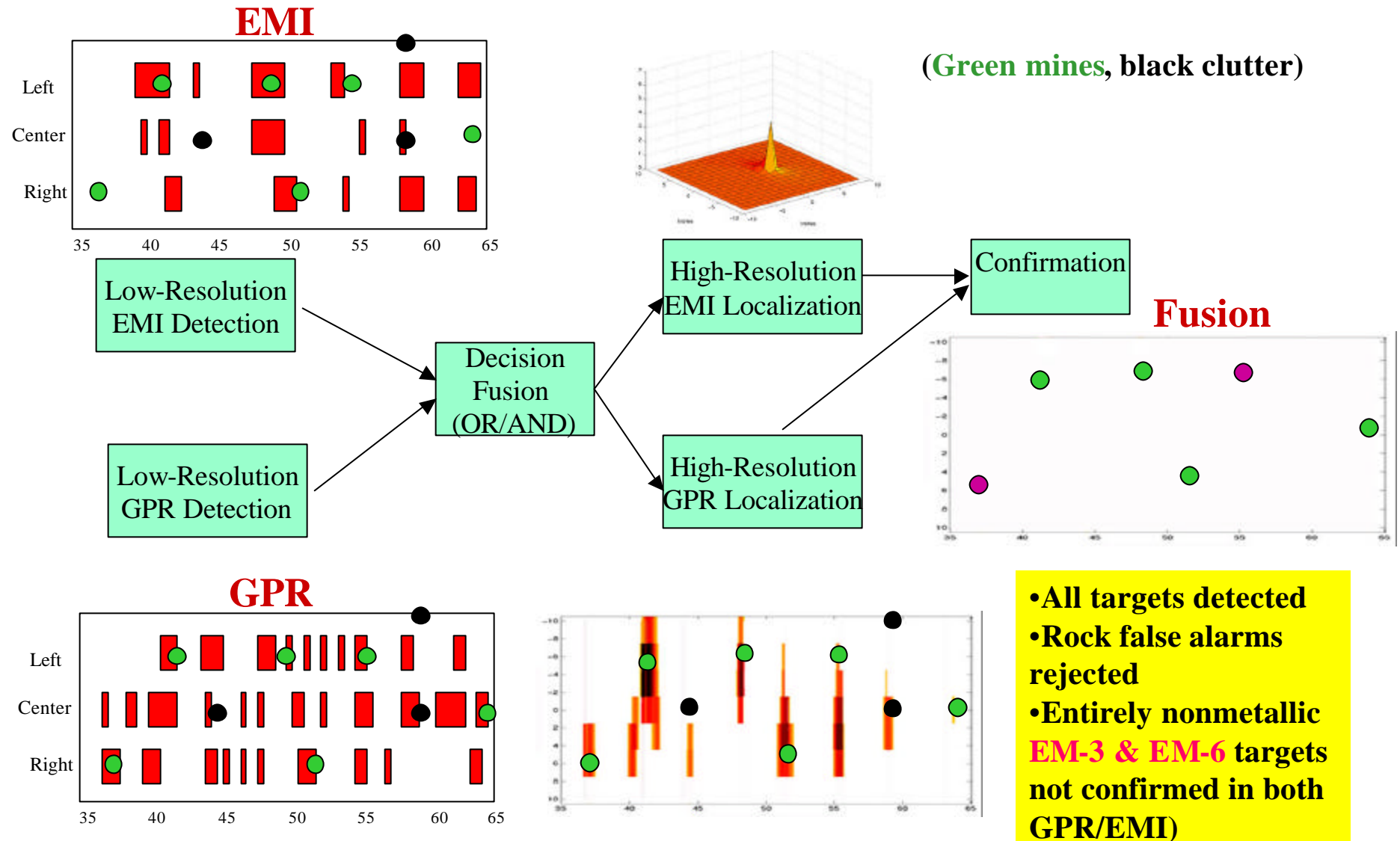
EMI: Accomplishments This Year

- Implemented and evaluated statistically based method for discriminate already detected-objects
- Algorithm characteristics
 - Based on analytical dipole scattering model
 - Able to process using time or frequency domain data
 - Capable of handling (and works best for) spatially sampled EM data
 - Inherently robust to modeling errors as well as uncertainty in position and orientation of the object
- Evaluated using
 - Synthetic data (time and frequency domain sensors)
 - Real GEM3 (frequency domain sensor)
 - Results encouraging
- Next up:
 - A min-max optimization approach for dealing with uncertainties in positioning of sensors

Multi-Frequency EMI Measurements: Mines and Clutter 5 cm deep - Background Removed Wet Sand (with Columbia University)



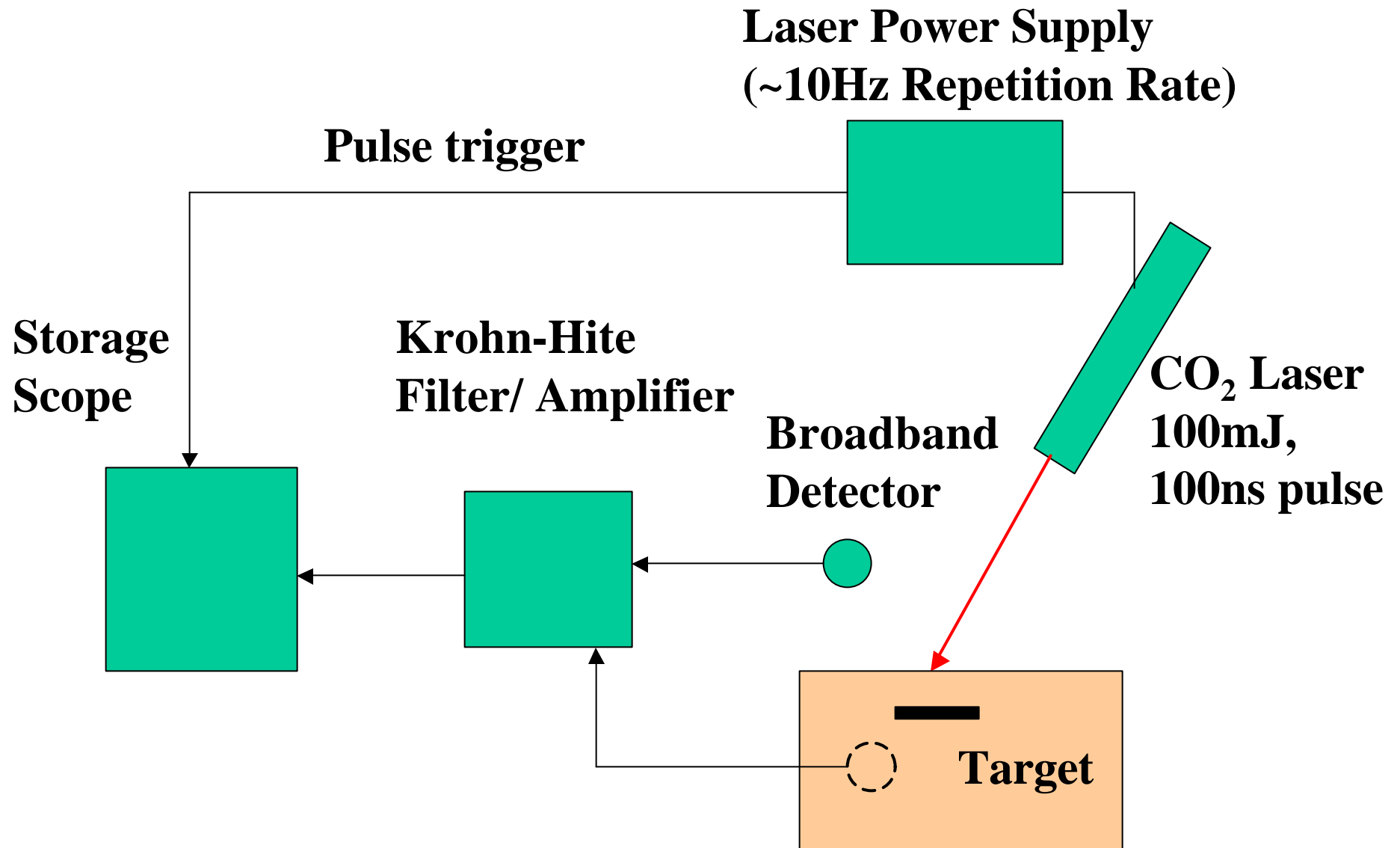
Multi-resolution Architecture for Clutter Rejection from GPR/EMI Fusion



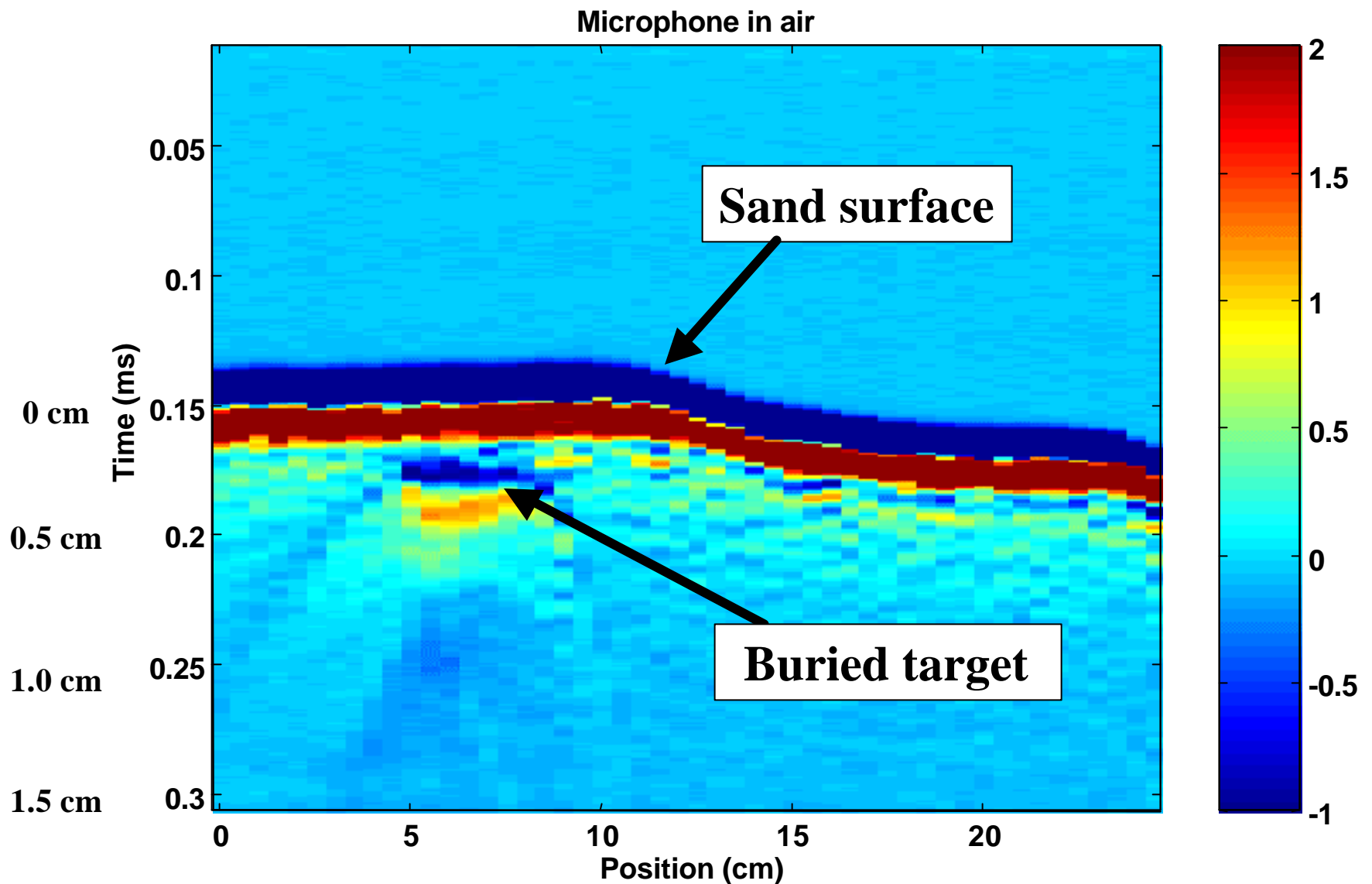
Conclusion

NEU's Dedham Test Track provides an essential resource to enable the evaluation and verification of processing and fusion algorithms with real data from a well-characterized "ground truth" facility

Laser/Acoustic Experimentation



Broadband Detection of Laser-Induced Subsurface Scattering

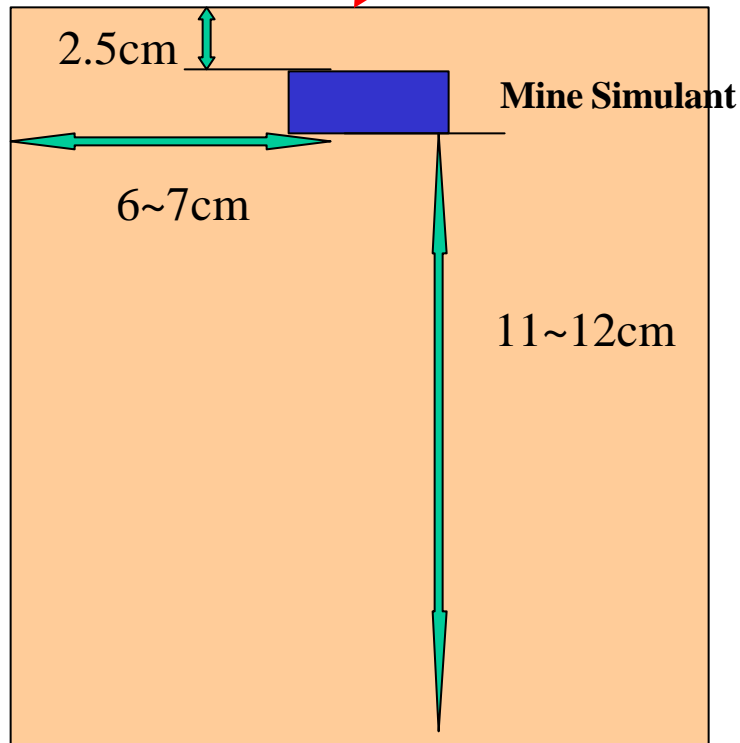


Acoustic Array Buried Object Detection

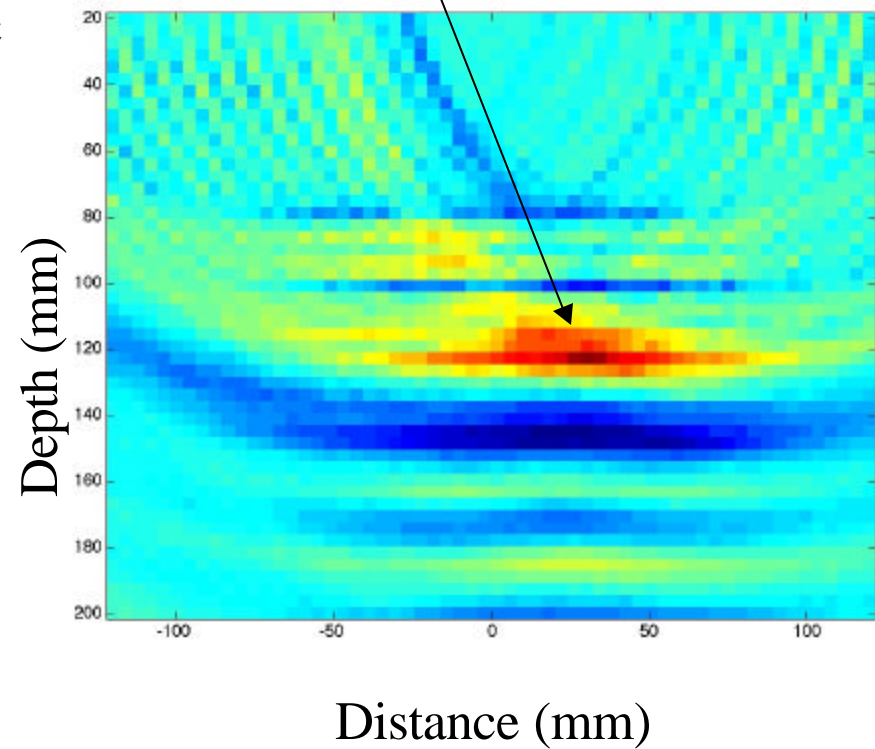
Hydrophone
Array



Focused
CO₂ laser
beam



Hard rubber target
5cm x 2.5 cm thick



Overall Summary

- **Integrated Approach:** develop models, sensors, and algorithms in conjunction with each other for particular problem
- **New Sensors:** Multistatic Array GPR, Laser/Acoustic, MEIT, Microwave Probe, Artificial Dielectric Matching Layer
- **New Models:** Dispersive 3-D FDTD, Multi-object SDFMM, Sparse Preconditioned 3-D FDFD, Gaussian Beam Asymptotics
- **New Rough Surface Clutter Reduction Methods:** Time-Gated Frequency Response, Ground Clutter Pulse Capture and Removal
- **New Reconstruction:** Sequential Array Estimation, EMI Imaging, Multistatic Template Field Matching, Optimized F-K Migration

NEU Demining MURI Related Publications

January 1997 - September 2001

Journal Papers

1. E. Miller and A. Willsky, "Multiscale, statistical anomaly detection analysis and algorithms for linearized inverse scattering problems," *IEEE Trans. On Geoscience and Remote Sensing*, Vol. 8, No. 1, pp. 151-184, January 1997.
2. W. Weedon and C. Rappaport, "A general method for FDTD modeling of wave propagation in arbitrary frequency- dispersive media," *IEEE Transactions on Antennas and Propagation*, pp. 401--410, March 1997.
3. A. Sahin and E. Miller, "Recursive T-Matrix algorithm for multiple metallic cylinders," *Microwave and Optical Technology Letters*, vol. 15, No. 6, pp.360-363, August 1997.
4. J. Kaufhold, M. K. Schneider, W. C. Karl, and A. S. Willsky, "A statistical method for efficient segmentation of MR imagery," *International Journal of Pattern Recognition and Artificial Intelligence*, Special issue on Processing of MRI Imagery, Vol. 11, No. 8, December 1997, pp. 1213-1231.
5. D. Keiswetter and I.J. Won, "Multifrequency electromagnetic signature of the Cloud Chamber, Nevada Test Site," *Jour. Environmental and Engineering Geophysics*, Vol. 2, Issue 2, pp. 99-104, 1997.
6. R. Deming and A. Devaney, "Diffraction tomography for multi-monostatic ground penetrating radar imaging," *Inverse Problems*, vol. 13, pp. 29-45, 1997.
7. A. Witten, I.J. Won, and S. Norton, "Subsurface imaging with broadband electromagnetic induction," *Inverse problems*, Vol. 13, pp. 1621-1639, 1997.
8. A. Witten, I.J. Won, and S. Norton, "Imaging underground structures using broadband electromagnetic induction," *Jour. of Environmental and Engineering Geophysics*, Vol. 2, Issue 2, pp. 105-114, 1997.
9. I.J. Won and D. Keiswetter, "Comparison of magnetic and electromagnetic data for underground structures," *Jour. Environmental and Engineering Geophysics*, Vol. 2, Issue 2, pp. 115-126, 1997.
10. I.J. Won, D. Keiswetter, D. Hanson, E. Novikova, and T. Hall, "GEM-3: a monostatic broadband electromagnetic induction sensor," *Jour. Environmental and Engineering Geophysics*, Vol. 2, Issue 1, pp. 53-64, 1997.
11. E. Marengo, A. Devaney, and E. Heyman, "Analysis and characterization of ultrawide-band scalar volume sources and the fields they radiate: Part II - square pulse excitation," *IEEE Transactions of Antennas and Propagation*, vol. 46, No. 2, February 1998.
12. A. Sahin and E. Miller, "Recursive T-Matrix Methods for Scattering From Multiple Dielectric and Metallic Objects," *IEEE Trans. Antennas and Propagation*, Vol. 46, No. 5, May 1998, pp. 672-678.
13. E. Marengo and A. Devaney, "Time-dependent plane wave and multipole expansions of the electromagnetic field," *Journal of Mathematical Physics*, vol. 39, no. 7, July 1998.
14. I.J. Won, D. Keiswetter, and E. Novikova, "Electromagnetic induction spectroscopy," *Jour. Environmental and Engineering Geophysics*, Vol. 3, Issue 1, pp. 27-40, 1998.
15. E. Miller, "Statistically Based Methods for Anomaly Characterization in Images from Observations of Scattered Radiation," *IEEE Trans. On Image Processing*, Vol. 8, No. 1, January, 1999, pp. 92-101.
16. C. Rappaport, S. Wu, and S. Winton, "FDTD Wave Propagation Modeling in Dispersive Soil Using a Single Pole Conductivity Model," *IEEE Transactions on Magnetics*, vol.35, May 1999, pp. 1542-1545.
17. E. Marengo, C. Rappaport, and E. Miller, "Optimum PML ABC Conductivity Profile in FDTD," *IEEE Transactions on Magnetics*, Vol.35, No. 3, May 1999, pp. 1506-1509.
18. G. Tsihrantzis, A. Devaney, and E. Heyman, "Estimation of Object Location from Wideband Scattering Data", *IEEE Transactions on Image Processing*, Vol. 8, No.7, July 1999, pp. 996-1001.
19. A. Sahin, and E. Miller, "Electromagnetic Scattering-Based Array Processing Methods for Near-Field Object Characterization", *Journal of Electromagnetic Waves and Applications*, Vol. 13, 1999, pp. 1209-1236.
20. M. E. Kilmer and G. W. Stewart, "Iterative Regularization and MINRES", *SIAM Journal for Matrix Analysis and Applications* Vol. 21, No. 2, pp. 613-628, 1999.
21. G.A. Tsihrantzis, A.J. Devaney and. Johansen, P.M, "Buried Object Detection and Location Estimation from Electromagnetic Field Measurements", *IEEE Transactions on Antennas and Propagation*, Vol. AP-47, pp. 1742-1744, 1999.

22. M. Silevitch, S. McKnight, and C. Rappaport, "A Unified Discipline of Subsurface Sensing and Imaging Systems," (Invited paper) *Journal of Subsurface Sensing and Applications*, January 2000, pp. 3-22.
23. M. K. Schneider, P. W. Fieguth, W. C. Karl, and A. S. Willsky, "Multiscale Statistical Methods for the Segmentation of Images," *IEEE Transactions on Image Processing*, Vol. 9, No. 3, pp. 456-468, March, 2000.
24. M. Belge and E. L. Miller, "A Sliding Window RLS-like Adaptive Filtering Algorithm for Filtering α -stable Noise," *IEEE Signal Processing Letters*, Vol. 7, No. 4, April 2000, pp. 86-89.
25. M. Belge, M. Kilmer and E. L. Miller, "Wavelet Domain Image Restoration with Adaptive Edge-Preserving Regularization," *IEEE Trans. On Image Processing*, Vol. 9, No.4, April 2000, pp. 598-608.
26. M. Belge, and E. L. Miller, "A Sliding Window RLS-like Adaptive Filtering Algorithm for Filtering α -stable Noise," *IEEE Signal Processing Letters*, Vol. 7, No. 4, April 2000, pp. 86-69.
27. C. Rappaport and S. Winton, "Using the PML ABC for Air/Soil Wave Interaction Modeling in the Time and Frequency Domains", *Journal of Subsurface Sensing and Applications*, Vol. 1, No. 3, July 2000, pp. 289-304.
28. E. Miller, M. Kilmer, and C. Rappaport, "A New Shape-Based Method for Object Localization and Characterization from Scattered Field Data," Special Issue of *IEEE Transaction on Geoscience and Remote Sensing* entitled "Computational Wave Issues in Remote Sensing, Imaging and Target Identification, Propagation, and Inverse Scattering," Vol. 38, No.4, July 2000, pp. 1682-1696.
29. E. L. Miller, I. Yavuv, L. Nicolaides, and A. Mandelis, "An adaptive, multiscale inverse scattering approach to photothermal depth profilometry," (Invited paper) *Circuits, Systems, and Signal Processing* special issue on Advanced Signal/Image Restoration, August 2000, Vol. 19, No. 4, pp. 339-363.
30. S. Winton and C. Rappaport, "Specifying the PML Conductivities by Considering Numerical Reflection Dependencies", *IEEE Trans. on Antennas and Propagation*, Vol.48, No. 7, September 2000, pp. 1005-1063.
31. C. Rappaport, M. Kilmer, and E. Miller, "Accuracy Considerations in Using the PML ABC with FDFD Helmholtz Equation Computation", (Invited paper) *International Journal of Numerical Modeling* Vol. 13, No. 471, September 2000, pp. 471-482.
32. V. Galdi, D.A. Castañón, and L.B. Felsen, "Multifrequency reconstruction of moderately rough interfaces via quasi-ray Gaussian beams," submitted to *IEEE Trans. Geosci. and Remote Sensing*, Sept. 2000.
33. O. Dorn, E. L. Miller, and C. M. Rappaport, "A shape reconstruction method for electromagnetic tomography using adjoint fields and level sets," (Invited paper) *Inverse Problems* special issue on Electromagnetic Imaging and Inversion of the Earth Subsurface, Vol. 16, No. 5, October 2000, pp. 1119-1156.
34. W. Li, R. Roy, R. Cleveland, L. Berg, and C. DiMarzio, "Laser-Induced Acoustic Waves for Measurement and Imaging," submitted to ASME/ICEME 2000.
35. G.A. Tsihrintzis and A.J. Devaney, "Higher-Order (Nonlinear) Diffraction Tomography: Reconstruction Algorithms and Computer Simulation", *IEEE Transactions on Image Processing*, Vol. IP-9, pp. 1560-1572, 2000.
36. G.A. Tsihrintzis and A.J. Devaney, "Higher-Order (Nonlinear) Diffraction Tomography: Inversion of the Rytov Series," *IEEE Transactions on Information Theory: Special Issue on Information Theoretic Imaging*, Vol. IT-46, pp. 1748-1761, 2000.
37. G.A. Tsihrintzis and A.J. Devaney, "A Volterra Series Approach to Nonlinear Traveltime Tomography," *IEEE Transactions on Geoscience and Remote Sensing*, Special Issue on Computational Wave Issues in Remote Sensing, Imaging and Target Identification, Propagation, and Inverse Scattering, vol. GRS-38, pp. 1733-1742, 2000.
38. A. Sahin, and E. L. Miller, "Model-Based Multiple Object Detection Using High-Resolution Near Field Array Processing," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, No. 1, January 2001, pp. 136-141.
39. V. Galdi, L.B. Felsen, and D.A. Castañón, "Time-domain radiation from large two-dimensional apertures via narrow-waisted Gaussian beams," submitted to *IEEE Trans. Antennas and Propagat.*, Jan. 2001.
40. R. Dufour, E. L. Miller, and N. Galatsanos, "Template Matching Based Object Recognition with Unknown Geometric Parameters," submitted to *IEEE Trans. Image Processing*, March 2001.
41. S. W. McKnight, C. A. DiMarzio, W. Li, and J. Stott, "Laser-Induced Acoustic Imaging of Buried Objects," *Journal of Subsurface Sensing Technologies and Applications*, Vol. 2, pp. 119-126, April 2001.
42. V. Galdi, L.B. Felsen, and D.A. Castañón, "Quasi-ray Gaussian beam algorithm for short-pulse two-dimensional scattering by moderately rough dielectric interfaces," submitted to *IEEE Trans. Antennas and Propagat.*, Apr. 2001.
43. X. Xu, E.L. Miller, C. Rappaport, and G. Sower, "Statistical Method to Detect Subsurface Objects Using Array Ground Penetrating Radar-Data," submitted to *IEEE. Trans. Geoscience and Remote Sensing* May 2001.

44. M. El-Shenawee, C. Rappaport, E. Miller, M. Silevitch, "3-D Subsurface Analysis of Electromagnetic Scattering from Penetrable/PEC Objects Buried under Rough Surfaces: Use of the Steepest Descent Fast Multilevel Multipole (SDFMM) Technique," *IEEE Transactions on Geoscience and Remote Sensing*, special issue on subsurface sensing, Vol. 39, No. 6, June 2001, pp. 1174-1182.
45. M. El-Shenawee, "The Steepest Descent Fast Multipole Method for Scattering from Two Penetrable Shallow Objects Buried Under Two-Dimensional Randomly Rough Surface," submitted to the *IEEE Trans. Antennas and Propagations*, June 2001.
46. B. Yang and C. Rappaport, "Response of Realistic Soil for GPR Applications with Two Dimensional FDTD," (Invited Paper) *IEEE Trans. on Geoscience and Remote Sensing* June 2001, pp. 1198-1205.
47. C. Rappaport, M. Silevitch, S. McKnight, C. DiMarzio, E. Miller and H. Raemer, "Multimode Subsurface Sensing and Imaging for Landmine Detection", *SSTA Journal*, Vol.2, 3, July, 2001.
48. D. Jiang, W. Meleis, M. El-Shenawee, E. Mizan, M. Ashouei and C. Rappaport, "Parallel Implementation of the Steepest Descent Fast Multipole Method (SDFMM) On a Beowulf Cluster for Subsurface Sensing Applications," submitted to the *IEEE MWCL*, July 2001.
49. M. El-Shenawee, C. Rappaport, D. Jiang and W. Meleis, "Electromagnetics Computations Using the MPI Parallel Implementation of the Steepest Descent Fast Multipole Method (SDFMM)," submitted to the *Journal of Applied Computational Electromagnetics (ACES)*, August 2001.
50. C. Rappaport, M. Silevitch, S. McKnight, C. DiMarzio, E. Miller and H. Raemer, "Multi-mode Subsurface Sensing and Imaging for Land Mine Detection," (Invited paper) accepted for publication in *Journal of Subsurface Sensing Technologies and Applications*, August 2001.
51. M. El-Shenawee and C. Rappaport, "Modeling Clutter from Bosnian and Puerto Rican Rough Ground Surfaces for GPR Subsurface Sensing Applications Using the SDFMM Technique," accepted for publication in *Journal of Subsurface Sensing Technologies and Applications*, August 2001.
52. Jiang, D., Meleis, W., El-Shenawee, M., Mizan, E., Ashouei, M., and Rappaport, C., "Parallel Implementation of the Steepest Descent Fast Multipole Method (SDFMM) On a Beowulf Cluster for Subsurface Sensing Applications," accepted for publication in *IEEE Microwave and Guided Wave Letters*, August 2001.
53. M. El-Shenawee, "Multiple Scattering Model for Non-Shallow Object Buried Under Two-Dimensional Random Rough Surface," submitted to the *IEEE Trans. on Geoscience and Remote Sensing*, September 2001.
54. V. Galdi, L.B. Felsen, and D.A. Castañon, "Quasi-ray Gaussian beam algorithm for time-harmonic two-dimensional scattering by moderately rough interfaces," *IEEE Trans. Antennas and Propagat.*, Vol. 49, No. 9, pp. 1305-1314, Sept. 2001.
55. V. Galdi, L.B. Felsen, and D.A. Castañon, "Narrow-waisted Gaussian beam discretization for short-pulse radiation from one-dimensional large apertures," *IEEE Trans. Antennas and Propagat.*, vol. 49, No. 9, pp. 1322-1332, Sept. 2001.
56. V. Galdi, J. Pavlovich, D.A. Castañon, W.C. Karl, and L.B. Felsen, "Moderately rough dielectric interface reconstruction via short-pulse quasi-ray Gaussian beams," submitted to *IEEE Trans. Antennas and Propagat.*, Sept. 2001.
57. V. Galdi, H. Feng, D.A. Castañon, W.C. Karl, and L.B. Felsen, "Moderately rough surface underground imaging via short-pulse quasi-ray Gaussian beams," submitted to *IEEE Trans. Antennas and Propagat.*, Sept. 2001.
58. E. Bahar and M. El-Shenawee, "Double Scatter Cross Sections for Two Dimensional Random Rough Surfaces that Exhibit Backscatter Enhancement," *J. of Optical Society of America*, to appear in December 2001.
59. M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo Simulations of Electromagnetic Wave Scattering from Random Rough Surface with 3-D Penetrable Buried Object: Mine Detection Application Using the SDFMM," *J. Optical Society of America A*, to appear in December 2001.
60. C. A. DiMarzio, T. Shi, C. M. Rappaport, G. O. Sauermann, "Two-Frequency Microwave Infrared Thermography to Improve Depth of Imaging of Objects Under a Rough Surface," in preparation for *Journal of Subsurface Sensing Technologies and Applications*
61. X. Xu, E. Miller, C. Rappaport, and G. Sower, "Statistical Method to Detect Subsurface Objects Using Array Ground Penetrating Radar Data," accepted for publication in *IEEE Trans. on Geoscience and Remote Sensing*.

- 62. M. Kilmer, E. Miller, and C. Rappaport, "Preconditioners for Structured Matrices Arising in Subsurface Object Detection," in review for publication in *Journal of Computational Physics*.
- 63. E. Kilmer and D. P. O'Leary, "Choosing Regularization Parameters in Iterative Methods for Ill-Posed Problems," *SIAM Journal on Scientific Computation*, in review.
- 64. C. Rappaport, M. El-Shenawee, and H. Zhan, "Suppressing GPR Signal Degradation from Randomly Rough Ground Surfaces," in review for publication in *IEEE Trans. on Geoscience and Remote Sensing*.
- 65. M. Kilmer, E. L. Miller, and C. Rappaport, "QMR-Based Projection Techniques for the Solution of Non-Hermitian Systems with Multiple Right Hand Sides," to appear, *SIAM Journal on Scientific and Statistical Computing*
- 66. L.B. Felsen and V. Galdi, "Complex-source-point narrow-waisted ray-like Gaussian beams for frequency and time domain radiation and scattering," in *Ultra-Wideband, Short Pulse Electromagnetics 5*, S.R. Cloude and P.D. Smith, Eds., Kluwer/Plenum Publishers, New York, NY, USA, 2001 (in print).

A Sliding Window RLS-like Adaptive Algorithm for Filtering alpha-stable Noise

Murat Belge and Eric L. Miller

Abstract— We introduce a sliding window adaptive RLS-like algorithm for filtering alpha-stable noise. Unlike previously introduced stochastic gradient type algorithms, the new adaptation algorithm minimizes the L_p norm of the error exactly in a sliding window of fixed size. Therefore, it behaves much like the RLS algorithm in terms of convergence speed and computational complexity compared to previously introduced stochastic gradient based algorithms which behave like the LMS algorithm. It is shown that the new algorithm achieves superior convergence rate at the expense of increased computational complexity.

I. INTRODUCTION

In the vast majority of signal processing applications it has been assumed that the signal or noise under investigation can be modeled by a Gaussian distribution law. This assumption has been justified by the central limit theorem and strong analytical properties of Gaussian pdf which leads to linear algorithms. However, in many real-world problems the noise encountered is more impulsive in nature than that predicted by a Gaussian distribution. Examples are underwater acoustic noise, low frequency atmospheric noise and many types of man-made noise [1]. Systems optimized under the Gaussian assumption often yield unacceptable performance when subjected to impulsive, non-Gaussian noise [12]. There exists a class of distributions, called alpha-stable distributions, than can be used to model these types of noise [1].

With the introduction of alpha-stable distributions to the signal processing community, a number of different adaptive filtering approaches have been proposed for filtering these processes [3], [5], [4], [7]. All of the algorithms that have been introduced so far can be classified as an LMS variant [8] which basically updates the filter coefficients by using an instantaneous approximation to the gradient of the cost function. The only difference of these algorithms from the conventional LMS algorithm is the minimization of the L_p norm of the error at the output of the adaptive filter instead of the usual Euclidean norm. The Least Mean p-Norm (LMP) algorithm [6] was derived exactly as described above. Later, motivated by the normalized versions of the LMS algorithm, Arikan *et. al.* developed the Normalized LMP (NLMP) algorithm [3]. Both LMP and NLMP suffer from the same problem that has plagued the LMS algorithm. Namely, when the input to the adaptive filter is highly correlated the convergence is very

slow. The RLO algorithm [7] is an alternative to LMP and NLMP, however its implementation requires some a-priori information on the error statistics and the filter inputs [7].

In this work, we develop a sliding window adaptation algorithm which is similar to the RLS algorithm [2] both in terms of derivation and convergence characteristics. The new algorithm provides much increased convergence rate at the expense of increased computational complexity. A block implementation of the new algorithm decreases the computational cost substantially.

II. SLIDING WINDOW LEAST MEAN P-NORM ADAPTATION ALGORITHM

The objective of an adaptation algorithm is to minimize the averaged error at the output of the filter by adjusting the coefficients of the filter. Adaptive estimation of a time-varying finite impulse response system is usually obtained by limiting the filtering memory. Here, we adopt a true finite memory or a sliding window approach for the adaptation of filter coefficients. That is we minimize the averaged L_p norm of the error in a window of size L :

$$J_{\underline{w}}(n) = \sum_{k=n-L+1}^n |d(k) - \underline{w}^t(n)\underline{x}(k)|^p = \sum_{k=n-L+1}^n |e(k)|^p \quad (1)$$

where $d(k)$ is the desired signal at time k , $\underline{w}(n)$ is the vector of optimal filter coefficients at time n , $\underline{x}(k) = [x(k) \ x(k-1) \dots x(k-N+1)]^t$ contains the N most recent samples of the input signal and $1 \leq p < 2$. Taking the gradient of $J_{\underline{w}}(n)$ with respect to $\underline{w}(n)$ and equating the result to zero we obtain:

$$\sum_{k=n-L+1}^n |e(k)|^{p-1} \text{sign}[e(k)] \underline{x}(k) = \underline{0} \quad (2)$$

which can be written as

$$\sum_{k=n-L+1}^n u(k) \underline{x}(k) \underline{x}^t(k) \underline{w}(n) = \sum_{k=n-L+1}^n u(k) d(k) \underline{x}(k) \quad (3)$$

where $u(k) = |e(k)|^{p-2}$, and (3) is obtained by substituting $\frac{e(k)}{|e(k)|}$ for $\text{sign}[e(k)]$ in (2) and rearranging. Defining $\underline{r}(n) = \sum_{k=n-L+1}^n u(k) d(k) \underline{x}(k)$ and $R(n) = \sum_{k=n-L+1}^n u(k) \underline{x}(k) \underline{x}^t(k)$ we obtain the following expression for $\underline{w}^*(n)$ which minimizes the p -norm of the error in a window of size L :

$$\underline{w}^*(n) = R^{-1}(n) \underline{r}(n) \quad (4)$$

Note, however, that $\underline{w}^*(n)$ cannot be readily obtained from (4) since $R(n)$ and $\underline{r}(n)$ are functions of $\underline{w}^*(n)$. However,

Dept. of Electrical and Computer Engineering, Northeastern University, Boston, MA 02148. This work was supported by an ODDR&E MURI under Air Force Office of Scientific Research contract F49620-96-1-0028, a CAREER Award from the National Science Foundation MIP-9623721, and the Army Research Office Demining MURI under Grant DAAG55-97-1-0013.

we can devise an iterative scheme to solve for $\underline{w}^*(n)$ at each point in time. Such an approach leads to the following algorithm:

Algorithm 1.

1. $\underline{w}^0(n) = \underline{w}^*(n-1)$
2. Compute $u^j(k) = |d(k) - \underline{x}^t(k)\underline{w}^j(n)|^{p-2}$, $k = n-L+1, \dots, n$
3. Compute $R^j(n) = \sum_{k=n-L+1}^n u^j(k)\underline{x}(k)\underline{x}^t(k)$ and $\underline{r}^j(n) = \sum_{k=n-L+1}^n u^j(k)d(k)\underline{x}(k)$
4. $\underline{w}^{j+1}(n) = [R^j(n)]^{-1} \underline{r}^j(n)$
5. If $\frac{\|\underline{w}^{j+1}(n) - \underline{w}^j(n)\|}{\|\underline{w}^j(n)\|} < \epsilon$ then, $\underline{w}^*(n) = \underline{w}^{j+1}(n)$, stop; else $j = j+1$, goto step 2

For each new input sample, we apply algorithm 1 to obtain the optimal filter coefficients, $\underline{w}^*(n)$, for the current time. Steps 2-5 of Algorithm 1 constitute the so-called *iterative re-weighted least squares* (IRLS) method which has been suggested and applied in several contexts [11], [9], [10]. The convergence of the IRLS algorithm can be guaranteed by making the following modification [9]:

$$u^j(k) = \begin{cases} u^j(k) & \text{if } u^j(k) \leq \frac{1}{\mu} \\ \frac{1}{\mu} & \text{if } u^j(k) > \frac{1}{\mu} \end{cases} \quad (5)$$

where μ is a small positive constant.

Note that, in step 4 of Algorithm 1, we need the inverse of $R^j(n)$. Rather than first computing $R^j(n)$ and then inverting this matrix to obtain $\underline{w}^{j+1}(n)$ we may consider computing $[R^j(n)]^{-1}$ directly in step 3 of Algorithm 1. To this end, consider the following expression for $i = L, \dots, 1$:

$$R^j(n-i+1) = R^j(n-i) + u^j(n-i+1)\underline{x}(n-i+1)\underline{x}^t(n-i+1) \quad (6)$$

Defining, $P^j(n) = [R^j(n)]^{-1}$, and applying the matrix inversion lemma [2] to (6) we obtain:

$$P^j(n-i+1) = P^j(n-i) - \frac{1}{\alpha(n-i)} \underline{g}(n-i) \underline{g}^t(n-i) \quad (7)$$

where $\underline{g}(n-i) = P^j(n-i)\underline{x}(n-i+1)$ and $\alpha(n-i) = 1/u^j(n-i+1) + \underline{x}^t(n-i+1)\underline{g}(n-i)$. Equation (7) implies that the matrix $P^j(n)$ can be obtained by a series of recursive updates starting from the matrix $P^j(n-L)$ at the beginning of the window. We assume that $P^j(n-L) = \frac{1}{\sigma}I$, which corresponds to a soft initialization [2]. Then, in step 4 of Algorithm 1, we compute $P^j(n)$ by using (7) instead of $R^j(n)$ and then obtain $\underline{w}^{j+1}(n)$ in step 5 by $\underline{w}^{j+1}(n) = P^j(n)\underline{r}^j(n)$.

The complexity of the algorithm given above is $O(MLN^2)$ where M is the number of IRLS iterations (steps 2-5 of Algorithm 1) needed. Because of the similarity of the algorithm to RLS, we call the new approach the recursive least mean p-norm algorithm (RLMP). The direct implementation of the RLMP algorithm is infeasible for most applications because of its high computational complexity which is dominated by construction of $P^j(n)$. However, a *subsampling* version of the RLMP algorithm where the filter coefficients are updated once at every k iterations, $k > 1$ being the subsampling rate, can be considered. In

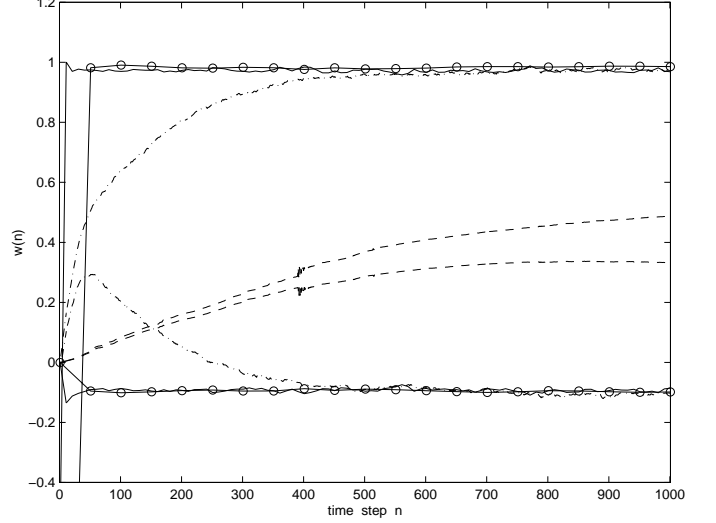


Fig. 1. Transient behavior of tap weight adaptations for RLMP with $L = 50$, $k = 10$ (solid line), RLMP with $L = 100$, $k = 50$ (circles), NLMP (dashed line) and LMP (dash-dotted line) algorithms.

particular, if $k = L$, the complexity of the RLMP algorithm is $O(MN^2)$ per iteration. In this case, the filter coefficients are updated once for every data block of length L . As we will see in Section 3, the average number of IRLS iterations can be quite low making the subsampled version of the RLMP algorithm a viable alternative to its stochastic gradient type counterparts.

III. SIMULATION STUDY

In this section, we compare the performance of the RLMP algorithm to that of NLMP and LMP algorithms. In the NLMP and LMP algorithms, the coefficients of the adaptive filter is updated as follows [3]:

$$\underline{w}(n+1) = \underline{w}(n) + \alpha \frac{|e(n)|^{p-1} \text{sign}[e(n)]}{h(n)} \underline{x}(n) \quad (8)$$

where $h(n) = 1$ for LMP and $h(n) = \|\underline{x}(n)\|_p^p + \gamma$, with $\gamma > 0$ being a small constant, for the NLMP algorithm. Following [3], we consider the following AR process:

$$x(n) = 0.99x(n-1) - 0.1x(n-2) + u(n) \quad (9)$$

where $u(n)$ is an alpha-stable sequence of i.i.d. random variables with $\alpha = 1.2$, $\beta = 0$ and $\gamma = 1$. A simulation is performed to identify the coefficients of the AR process with the $p = 1.1$ norm. Figure 1 shows the transient behavior of the tap weights of the adaptive filter and Fig. 2 shows the norm of the error between the true and the estimated parameters, defined as $E(n) = 20 \log_{10} \|\underline{w}_{true} - \underline{w}(n)\|_2$. Both figures were obtained by averaging the results of 100 independent trials. The parameters of the RLMP algorithm are: $\mu = 10^{-6}$, $\sigma = 10^{-4}$, $\epsilon = 10^{-2}$. The RLMP algorithm was implemented for two different window sizes corresponding to $L = 50$ and $L = 100$ samples. For $L = 50$, the filter coefficients were updated once for every 10 iterations and for $L = 100$, the filter coefficients were updated once per 50 data samples. For a window of $L = 50$ samples, the RLMP algorithm produces a steady-state tap

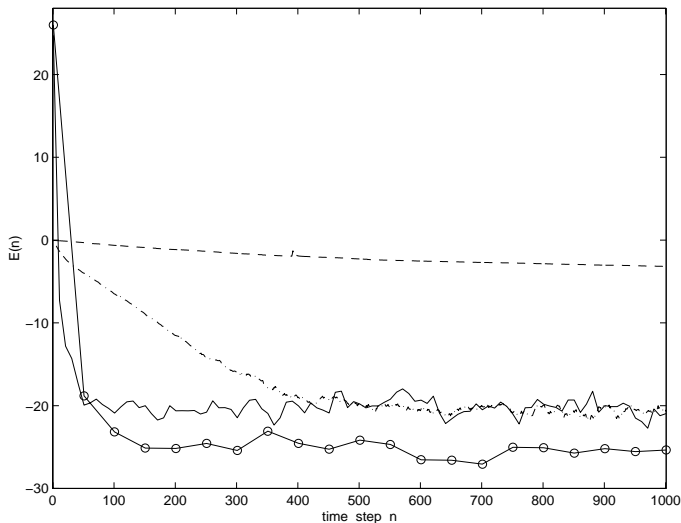


Fig. 2. Transient behavior of tap weight error powers for RLMP with $L = 50$, $k = 10$ (solid line), RLMP with $L = 100$, $k = 50$ (circles), NLMP (dashed line) and LMP (dash-dotted line) algorithms.

weight error of approximately -20dB. This figure can be made smaller/larger by adjusting the window length. The step sizes of the NLMP and the LMP algorithms were set to produce the same steady state tap weight error as RLMP. The step size for NLMP was found to be 3.8×10^{-2} and 5×10^{-5} for LMP. Figure 1 and 2 show that RLMP provides a large improvement in convergence rate over the NLMP and LMP. To give an idea about the number of iterations needed for the outer IRLS iterations to converge, we computed the number of IRLS iterations at each discrete time, n , by averaging 100 trials. The results are displayed in Fig. 3. It is seen that at most 3 iterations are sufficient to obtain a relative error $\frac{\|w^{j+1}(n) - w^j(n)\|}{\|w^j(n)\|}$ of about 10^{-2} .

Examining Algorithm 1 in detail, we see that the most efficient implementation of the RLMP algorithm requires $ML(2N^2 + 3N + 2)$ multiplications, $2ML$ divisions and ML nonlinear operations per update of the filter coefficients. In general, for $k = L$ (i.e. filter coefficients are updated once for each data block of L samples), the computational effort required by the RLMP algorithm is approximately M times that of a single RLS update plus M nonlinear operations. From Fig. 3, we see that, for this example, the ensemble average of M is actually quite low and approximately 2.5 at the steady state.

IV. CONCLUSION

In this letter, we described a novel adaptation algorithm for filtering alpha-stable noise. The new algorithm is derived by minimizing the averaged L_p error at the output of the filter in a window of fixed size. Simulations show that the new algorithm provides much improved convergence rate compared to other stochastic gradient based adaptation algorithms for alpha-stable noise environments. The major disadvantage of the algorithm is its computational complexity. We proposed a subsampled implementation of the RLMP algorithm which reduces the computational complexity to $O(MN^2)$ per data sample. Current

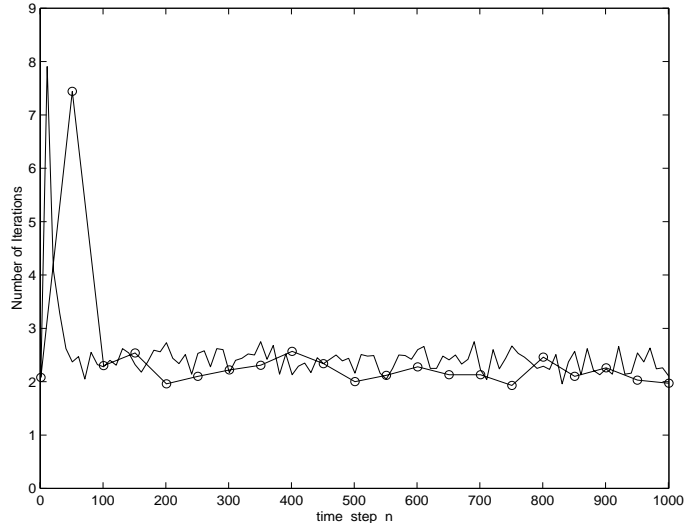


Fig. 3. Fig. 3 Ensemble average of the number of IRLS iterations needed at each time instant for RLMP with $L = 50$, $k = 10$ (solid line) and RLMP with $L = 100$, $k = 50$ (circles).

research is focused on frequency domain implementations of the RLMP algorithm which will further reduce the computational complexity.

REFERENCES

- [1] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, New York: Wiley, 1995.
- [2] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.
- [3] O. Arikan, M. Belge, E. Cetin, E. Erzin, "Adaptive filtering approaches for non-Gaussian stable processes," *Proc. ICASSP'95*, v. 2, pp. 1400-1403, 1995.
- [4] G. Aydin, O. Tanrikulu, E. Cetin, "Robust least mean mixed norm adaptive filtering for α -stable random processes," *Proc. IEEE ISCAS'97*, v. 4, pp. 2296-2299, 1997.
- [5] S. Kalluri, G. R. Arce, "Adaptive weighted myriad filter algorithms for robust signal processing in α -stable noise environments," *IEEE Trans. Signal Processing*, v. 46, n. 2, pp. 322-334, 1998.
- [6] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proc. IEEE*, v. 81, pp. 986-1009, 1993.
- [7] J. S. Bodenschatz and C. L. Nikias, "Recursive local orthogonality filtering,"
- [8] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, *IEEE Trans. Signal Processing*, v. 45, n. 9, pp. 2293-2300, 1997. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [9] J. Schroeder, R. Yarlagadda, J. Hershey, " L_p normed minimization with applications to linear predictive modeling for sinusoidal frequency estimation," *Signal Processing*, v. 24, n. 2, pp. 193-216, 1991.
- [10] E. E. Kuruoglu, P. J. W. Rayner, W. J. Fitzgerald, "Least l_p -norm impulsive noise cancellation with polynomial filters," *Signal Processing*, v. 69, pp. 1-14, 1998.
- [11] M. Belge, M. E. Kilmer, E. L. Miller, "Wavelet domain image restoration with adaptive edge preserving regularization," <http://www.cdsp.neu.edu/info/students/belge/index.html>, to appear in *IEEE Trans. Image Processing*.
- [12] G. A. Tsihrintzis and C. L. Nikias, "Performance of optimum and suboptimum receivers in the presence of impulsive noise modeled as an alpha-stable process," *IEEE Trans. Commun.*, v. 43, pp. 904-914, March 1995.

Under review, "IEEE Trans. Geoscience and Remote Sensing"

A New Shape-Based Method for Object Localization and Characterization from Scattered Field Data *

Eric L. Miller, Misha Kilmer, and Carey Rappaport
Dept. of Electrical and Computer Engineering
Northeastern University
360 Huntington Ave.
Boston, MA 02215
Tel: (617) 373-8386
email: elmiller@cdsp.neu.edu

July 22, 1999

Abstract

The problem of characterizing the structure of an object buried in an inhomogeneous half-space of unknown composition is considered. We develop a non-linear inverse scattering algorithm based on a low dimensional parameterization of the unknown object and the background. In particular, we use low order polynomials to represent the contrast in the real and imaginary parts of the object and background complex permittivities. The boundary separating the target from the unknown background is described using a periodic, quadratic B-spline curve whose control points can be individually manipulated. We determine the unknown control point locations and contrast expansion coefficients using a greedy-type approach to minimize a regularized least-squares cost function. The regularizer used here is designed to constrain the geometric structure of the boundary of the object and is closely related to snake methods employed in the image processing community. We demonstrate the performance of our approach via extensive numerical simulation involving 2D, TM_z scattering geometries.

*This work was supported by an OSD MURI on Demining under Grant DAAG55-97-1-0013 and a CAREER Award from the National Science Foundation, MIP-9623721

1 Introduction

We consider the problem of localizing and characterizing the structure of an object buried in a halfspace given noisy observations of scattered electromagnetic fields collected near the interface. Such problems arise in application areas including environmental remediation, humanitarian demining, medical imaging and non-destructive testing [6, 11, 14, 26]. Here we are particularly concerned with addressing two of the many challenges associated with such inverse scattering problems. First, these problems are known to be highly ill-posed in that the quantity of information contained in the data, which are collected only near the interface, is quite limited relative to the amount of information one would like to extract. This leads to considerable instabilities in terms of reliably characterizing the behavior of the subsurface [21, 25]. A second difficulty is the need to deal with *clutter*. Here we take clutter to be physical characteristics of the medium, in particular volume inhomogeneities, whose presence impedes our ability to characterize the object of interest.

Typical methods for overcoming these two problems are based on forming an image of the subsurface and then post processing the results to extract target information [5, 27]. The issue of ill-posedness is addressed through the use of a regularization procedure [3, 23, 25] to stabilize the imaging portion of the algorithm while clutter suppression is achieved through the use of image processing methods to separate targets from background. In forming an image, however, one must solve a large scale, non-linear optimization problem whose size is equal to the number of pixels (voxels) in the region of interest; a highly computationally intensive process. Moreover, under this approach one uses the limited data to generate values for all of the pixels only a few of which contain useful information about the underlying target. By diffusing the information in the data in this manner, target detection can be problematic [24].

There has been considerable work in the past decade on methods that extract directly from the data geometric information regarding the shape and location of the object [8–13, 19, 21, 22, 29, 32].

The assumption underlying most of these methods is that the object is embedded in a medium of otherwise known structure (homogeneous or halfspace). Rather than parameterizing the problem in terms of a large number of pixel values, a relatively small number of unknowns are used to describe the shape of the target. For these methods the issue of ill-posedness is either addressed through a regularization procedure related to the shape of the object [21] or through the use of more traditional Tikhonov or minimum norm least squares methods [11, 15, 32].

For our problems, these shape-based methods are not applicable. For example, the approach taken in [13] requires that the object be surrounded on all sides by transmitters and receivers. While this assumption is satisfied in, for example, medical imaging applications, we are restricted here to problems where we have only reflection-type data at our disposal. Additionally, most of the previously developed techniques assume that the electrical properties of the background are fully known and in many cases, the object's contrast function is also specified *a priori* [22]. Because we wish to address the issue of volume inhomogeneity, we must look to a different approach.

Hence we consider a parameterization based on a concise description of the object's geometric structure that provides for the recovery of a limited amount of information regarding the spatial variations of the complex permittivity over the target and the background. As described in § 2, the variations in the background and the object contrast are modeled via a superposition of a small number of expansion functions; one set of functions for the background and separate set for the object. Thus, the unknowns here are the corresponding expansion coefficients.

As in [1, 11, 21], we seek a representation for the shape of the object in terms of a small number of unknowns. In previous work, a Fourier-type expansion is used in conjunction with an underlying assumption that the object of interest is star-like. Here we choose to describe the boundary of the objects differently, using a linear combination of quadratic B-splines [4, Chap. 3]. The motivation for this choice comes from the fact that the parameters governing this representation (known as

control points) impact the shape of the object only over a small portion of the perimeter. Thus, the control points provide direct and local control over the object's shape. This control is exploited in the development of a simple and efficient inversion scheme described in § 3. Finally, the control-point parameterization leads to a natural, shape-based regularization technique related to *snake* methods [4] used in image processing for contour representation and image segmentation.

We view this approach as a compromise between an ill-posed imaging method that allows for arbitrary variation throughout the region and the more constrained shape-based methods in which homogeneity is assumed. By restricting the contrast variations to lie within the linear span of a set of basis functions, we clearly limit the classes of variations which can be recovered from this inversion process. The motivation for this decision comes from the fact that a detailed reconstruction of the medium is often not what is desired nor is it supported by the information in the data. Rather, the primary information of interest concerns the geometric structure of the object. Thus we are willing to settle for a coarse reconstruction of the object and background contrast functions.

The remainder of this paper is arranged as follows. In § 2, the scattering problem and the models for the object and background are presented; § 3 gives a presentation of the inversion procedure with examples of its performance shown in § 4. Conclusions and future work are described in § 5.

2 Problem formulation

We consider a two-dimensional scattering problem illustrated in Fig. 1. Time harmonic $e^{-j\omega t}$ y -polarized plane waves at various incident angles and frequencies are used to probe the medium. The complex permittivity of the lower halfspace is represented as the sum of a nominal, constant value, $\epsilon_1 + (j/\omega)\sigma_1$, and a space varying perturbation,

$$g(\mathbf{r}) = \epsilon_0 \epsilon_p(\mathbf{r}) + \frac{j}{\omega} \sigma_p(\mathbf{r}), \quad (j = \sqrt{-1}), \quad (1)$$

that encompasses both the object of interest as well as the clutter. Here, $\mathbf{r} = [x \ z]^T$ is a two vector indicating the position of a point in the plane, ϵ_0 is a constant that denotes the permittivity of free space, and ω denotes angular frequency. The scattered fields generated by g are observed along an array of point receivers also located in the upper halfspace. Under this 2D model, there is only a single component of the electric field thereby resulting in a scalar scattering problem.

The model linking the structure of g to the observed scattered field at the k th point along the array, \mathbf{r}_k , is [7, Chap. 9]

$$y(\mathbf{r}_k) = \omega^2 \mu_0 \int_W G(\mathbf{r}_k, \mathbf{r}') E(\mathbf{r}') g(\mathbf{r}') d\mathbf{r}' + n(\mathbf{r}_k), \quad (2)$$

where $y(\mathbf{r}_k)$ is the datum at \mathbf{r}_k for a given incident field, W is the region over which g is nonzero, n is zero mean, additive white Gaussian measurement noise with variance ζ^2 , and G and E denote the Green's function and total electric field, respectively. The constant μ_0 denotes the permeability of free space. In (2), we are only concerned with evaluating the Green's function, $G(\mathbf{r}, \mathbf{r}')$, when \mathbf{r} is above the interface and \mathbf{r}' is below in which case [10]

$$G(\mathbf{r}, \mathbf{r}') = \frac{j}{2\pi} \int_{-\infty}^{\infty} \frac{1}{\gamma_0 + \gamma_1} \exp [j\alpha(x - x') + j(\gamma_1 z - \gamma_0 z')] d\alpha \quad (3)$$

with $\gamma_i = (k_i^2 - \alpha)^{1/2}$ where $k_i = \omega \sqrt{\epsilon_i \mu_0 + j\sigma_i \mu_0 / \omega}$ denotes the wavenumber in the upper ($i = 0$) and lower ($i = 1$) halfspace. The total electric field $E(\mathbf{r})$ is governed by the Helmholtz equation

$$[\nabla^2 + k^2(\mathbf{r})] E = 0$$

with a Sommerfeld radiation condition where $k^2(\mathbf{r})$ is equal to k_0^2 above the interface and $k_1^2 + \omega^2 \mu_0 g(\mathbf{r})$ below. Finally, the total electric field at \mathbf{r} is the sum of the unperturbed electric field generated by a plane wave impinging on a halfspace, $E_{unp}(\mathbf{r})$, [20, § 3.2] and the scattered electric field, $E_s(\mathbf{r})$. With $E = E_s + E_{unp}$, it is easy to show that:

$$[\nabla^2 + k^2(\mathbf{r})] E_s = -\omega^2 \mu_0 g(\mathbf{r}) E_{unp}. \quad (4)$$

As $y(\mathbf{r})$, the observed field, depends on the total field in the region W , it therefore implicitly depends on the particular unperturbed field, E_{unp} , on $g(\mathbf{r})$, and on the frequency ω .

2.1 A Model for $g(\mathbf{r})$

As described in the introduction, we model the unknown portion of the permittivity in the lower halfspace as a superposition of a slowly varying, compactly supported anomaly (the target) on a background of partially known structure (clutter). Mathematically $\epsilon_p(\mathbf{r})$ is decomposed as

$$\epsilon_p(\mathbf{r}) = S(\mathbf{r}) \sum_{i=1}^{N_a} a_{1,i} b_{1,i}(\mathbf{r}) + [1 - S(\mathbf{r})] \sum_{i=1}^{N_b} a_{2,i} b_{2,i}(\mathbf{r}) \equiv S(\mathbf{r}) B_1(\mathbf{r}) \mathbf{a}_1 + [1 - S(\mathbf{r})] B_2(\mathbf{r}) \mathbf{a}_2 \quad (5)$$

where $S(\mathbf{r})$ is one over the (unknown) support of the object and zero elsewhere. The functions $b_{1,i}$ represent basis functions describing the contrast of the real part of the anomaly while the $b_{2,i}$ do the same but for the background. The $a_{i,j}$ are the expansion coefficients determining the weight of each function. The row vectors B_i , $i = 1, 2$ hold the expansion functions and \mathbf{a}_i are vectors of expansion coefficients. An analogous model holds for $\sigma_p(\mathbf{r})$, which is proportional to the imaginary portion of $g(\mathbf{r})$:

$$\sigma_p(\mathbf{r}) \equiv S(\mathbf{r}) B_1(\mathbf{r}) \mathbf{c}_1 + [1 - S(\mathbf{r})] B_2(\mathbf{r}) \mathbf{c}_2. \quad (6)$$

While in principle the expansion functions could differ between ϵ_p and σ_p , the support function S is the same since it represents a physical boundary between the object and the rest of the medium.

The particular choice of the $b_{i,j}$ depends on the application at hand. If one thought that there was a homogeneous dielectric anomaly of contrast $a_{1,1}$ against a real-valued homogeneous background of value $a_{2,1}$ then one would take $b_{1,1}(\mathbf{r}) = b_{2,1}(\mathbf{r}) = 1$. Use of higher order polynomials, trigonometric functions etc. provide greater flexibility in capturing true, underlying inhomogeneities. In any event, assuming the B_i are known, the objective of the problem in this paper is to determine the structure of S along with the \mathbf{a}_i and \mathbf{c}_i given the observed scattered field data.

In this work, we employ a B-spline model to describe the contour of the anomaly, that is, the boundary of the set $S(\mathbf{r})$. To define this curve, we start with an interval $[0, L]$ on which a *knot* sequence k_0, \dots, k_{K-1} with $0 \leq k_i \leq L$ is defined. For this sequence, there exists a periodic basis

of quadratic B-splines $C_{k_i}(s)$ such that

$$b(s) \equiv [x(s), z(s)] = \sum_{i=0}^{K-1} C_{k_i}(s)[x_i, z_i], \quad s \in [0, L] \quad (7)$$

for a given set of x_i, z_i expansion coefficients, or *control points*, uniquely defines a closed, C^2 parametric curve in the plane. The support of each $C_{k_i}(s)$ is $[k_i, k_{i+3}]$. Since the basis is taken to be periodic, the control points are assumed to be wrapped; that is, $[x_0, z_0] = [x_{K-2}, z_{K-2}]$ and $[x_1, z_1] = [x_{K-1}, z_{K-1}]$, so there are a total of K control points, $K - 2$ of which are unique.

To implement our model, we assume that the boundary of the anomaly is in the form:

$$b^*(s) \equiv \sum_{i=0}^{K-1} C_{k_i}(s)[x_i^*, z_i^*]. \quad (8)$$

If $\mathbf{r} = [x \ z]^T$ is a point inside $b^*(s)$, then for example, $\epsilon_p(\mathbf{r})$ is $B_1(\mathbf{r})\mathbf{a}_1$, while for \mathbf{r} outside, $\epsilon_p(\mathbf{r}) = B_2(\mathbf{r})\mathbf{a}_2$.

3 Algorithm

In this section we describe in detail the algorithm we use to solve the inverse problem. We begin by discussing the discretized form of the model. The reader is referred to Table 1 for a summary of the indexing parameters that have been or will be identified.

3.1 Discretization

Using the method of moments [17] with a pulse basis and point matching to discretize (2) and a lexicographical ordering of the unknowns yields the matrix equation:

$$\mathbf{y} = \mathbf{G}\mathcal{D}(\mathbf{E})\mathbf{g} + \mathbf{n}, \quad \mathbf{y} \in C^{M \times 1} \quad (9)$$

where M is the number of source/receiver pairs, \mathbf{y} and \mathbf{n} are vectors with components $y(\mathbf{r}_k)$ and $n(\mathbf{r}_k)$ respectively, \mathbf{G} the discrete Green's function premultiplied by $\omega^2\mu_0$, $\mathcal{D}(\cdot)$ is the diagonal matrix formed from the vector argument, and \mathbf{E} is the vector containing the total electric field (which depends on \mathbf{g}) at each pixel. Finally, $\mathbf{g} \in C^{N \times 1}$ is a vector holding the intensity values of

$g(\mathbf{r})$ at the N grid points in our discretization of the subsurface and from (5) and (6) is given as:

$$\mathbf{g} = \epsilon_0[\mathbf{S}\mathbf{B}_1 - (\mathbf{I} - \mathbf{S})\mathbf{B}_2]\mathbf{a} + \frac{j}{\omega}[\mathbf{S}\mathbf{B}_1 - (\mathbf{I} - \mathbf{S})\mathbf{B}_2]\mathbf{c} \equiv \left[\epsilon_0\mathbf{V} \quad \frac{j}{\omega}\mathbf{V} \right] \mathbf{w} \quad (10)$$

where $\mathbf{a} = [\mathbf{a}_1^T \quad \mathbf{a}_2^T]^T$, $\mathbf{c} = [\mathbf{c}_1^T \quad \mathbf{c}_2^T]^T$, $\mathbf{w} = [\mathbf{a}^T \quad \mathbf{c}^T]^T$, \mathbf{S} is a diagonal matrix corresponding to S with $\mathbf{S}_{i,i} = 1$ if any part of pixel i is inside $b^*(r)$ and 0 otherwise, \mathbf{I} is the identity matrix, and the \mathbf{B}_i denote the $N \times N_a$ and $N \times N_b$ matrices corresponding to evaluating $B_i(r)$ at the N gridpoints.

Finally, using $\mathbf{A} = \mathbf{G}\mathcal{D}(\mathbf{E})$ with (10) and (9) we have

$$\mathbf{y} = [\epsilon_0\mathbf{A}\mathbf{V}, \quad \frac{j}{\omega}\mathbf{A}\mathbf{V}]\mathbf{w} + \mathbf{n} \equiv \mathbf{K}\mathbf{w} + \mathbf{n}, \quad \mathbf{y} \in C^{M \times 1}. \quad (11)$$

Note that there is one such matrix-vector equation of this form for each frequency and each different incident angle. In the remainder of the paper, we assume n_1 frequencies and n_2 angles are used, and we use double superscripts i, k to denote the corresponding vectors and matrices at frequency ω_i and angle θ_k : single superscripts imply that that quantity depends only on ω_i . For example, $\mathbf{y}^{i,k}$ denotes the data obtained via (9) or (11) for frequency ω_i and incident angle θ_k

$$\mathbf{y}^{i,k} = \mathbf{G}^i \mathcal{D}(\mathbf{E}^{i,k}) \mathbf{g}^i + \mathbf{n}^{i,k} = \mathbf{K}^{i,k} \mathbf{w} + \mathbf{n}^{i,k}, \quad (12)$$

whereas \mathbf{g}^i denotes (10) at frequency ω_i .

3.2 Generating Scattered Field

To generate the data for a given frequency ω_i and incident angle θ_j , we need to compute the M -length vector $\mathbf{E}^{i,j}$ (i.e. we need to solve the forward scattering problem). To do this, we determine the scattered field by discretizing (4) using a finite difference scheme with PML boundary condition (details are described in [18]) to obtain the matrix equation $\mathbf{M}^i \mathbf{E}_s^{i,j} = -\omega_i^2 \mu_0 \mathbf{g}^i \mathbf{E}_{unp}^{i,j}$ where the unperturbed field $\mathbf{E}_{unp}^{i,k}$ is known and depends on incident angle and frequency, and \mathbf{M}^i is the matrix corresponding to the discretized operator in (4) with the PML boundary condition. The scattered field is determined by solving the matrix equation for $\mathbf{E}_s^{i,j}$, and finally, $\mathbf{E}^{i,j} = \mathbf{E}_{unp}^{i,j} + \mathbf{E}_s^{i,j}$.

3.3 Algorithm Description

Our algorithm seeks to find a good approximation to $g(\mathbf{r})$ by successively generating better and better approximations to $b^*(s)$, the boundary of the anomaly, and the coefficient vectors \mathbf{a} and \mathbf{c} .

In particular, we seek a minimum of the following cost function

$$J(b(s), \mathbf{a}, \mathbf{c}) = \sum_{i=1}^{n_1} \sum_{k=1}^{n_2} \|\mathbf{y}^{i,k} - \mathbf{G}^i \mathcal{D}(\mathbf{E}^{i,k}) \mathbf{g}^i\|_2^2 + \lambda_1 \Omega_1(b(s)) + \lambda_2 \Omega_2(b(s)). \quad (13)$$

where we emphasize that both $\mathbf{E}^{i,j}$ as well as \mathbf{g}^i depend implicitly on the geometry of the anomaly as well as the expansion vectors via the models developed in § 2 and § 3.1. The first term in (13) enforces fidelity to the data while the second and third play the role of regularizers.

Traditional regularization methods used to combat ill-posedness in an image restoration framework function by enforcing smoothness, or in some cases edge preservation, in the reconstruction. In our case, the $\Omega_i, i = 1, 2$ are used to influence the geometric structure of the recovered anomaly. Specifically, we define these functions as

$$\Omega_1(b(s)) = \sum_{i=0}^{K-2} (z_i - z^*)^2 \quad \text{and} \quad \Omega_2(b(s)) = \sum_{i=0}^{K-2} (x_i - x_{i+1})^2 + (z_i - z_{i+1})^2 \quad (14)$$

where K is the total number of control points, x_i and z_i are the coordinates of the i th control point, and z^* is a fixed z -value depending on the particular application. The first term attempts to penalize objects that are too deep while the second penalizes the total length between control points. In this way we dissuade the algorithm from choosing curves that are overly elongated and/or deep. Our justification for the first of these choices comes from our knowledge that significant depth information is not available in the measured data due to the loss in the soil and the positioning of detectors only above the interface. Moreover, in most applications one possesses some *a priori* information concerning the depth at which targets are likely to be buried. The values of λ_1, λ_2 tell how strongly we want to dissuade the algorithm from reconstruction of curves that are too elongated and/or deep. Finding near optimal regularization parameters is a very difficult problem and there is a whole body of literature dedicated to this issue (see, for instance, [2, 16]). In this work, we assume

that good parameters are known *a priori*: determining accurate parameter selection strategies will be the subject of future work.

We consider a greedy-type algorithm for minimizing (13). We begin with an initial estimate, $b^{(0)}(s)$, of $b^*(s)$ defined by a set of K knots and $K - 2$ distinct control points. From $b^{(0)}(s)$ the matrix \mathbf{S} is found. Suppose for the moment that initial estimates of \mathbf{a}, \mathbf{c} , denoted $\mathbf{a}^{(0)}, \mathbf{c}^{(0)}$, are known (we address the issue of generating these guesses below). Using (10), the \mathbf{g}^i are calculated for the different frequencies. We then use (13) to determine the cost of $(b^{(0)}, \mathbf{a}^{(0)}, \mathbf{c}^{(0)})$.

We update the estimate of $b^{(0)}(s)$ by systematically perturbing each control point from its original position by a fixed amount \bar{h} in the horizontal, vertical, and diagonal directions, respectively, for a total of 8 different moves per point. Note that each of these 8 moves corresponds to a new curve by definition in (7). Since there are $K - 2$ unique points and 8 moves for each point, this corresponds to $8(K - 2)$ different possible new curves. For each possible curve, we first estimate new values for \mathbf{a} and \mathbf{c} and then evaluate the cost associated with the current curve and these estimates using (13). Finally, $b^{(1)}(s)$, $\mathbf{a}^{(1)}$, and $\mathbf{c}^{(1)}$, our new estimates, are taken as that triple giving minimum cost, provided that cost is less than (or equal to) the cost associated with $(b^{(0)}(s), \mathbf{a}^{(0)}, \mathbf{c}^{(0)})$. The process is repeated as many times as is necessary.

In principle, determination of \mathbf{a} and \mathbf{c} requires the solution of a low dimensional inverse scattering problem. That is, every time we want to determine the cost for a candidate curve, we need to solve a non-linear optimization problem. To simplify this procedure, we consider an alternative approach. At the end of the k th stage, we compute the internal fields associated with the current estimate of the anomaly. This requires the solution of $n_1 n_2$ forward scattering problems of size N . From (9), the data $\mathbf{y}^{i,l}$ depend on the anomaly through \mathbf{g}^i and the internal fields. At stage $k + 1$, we assume that as we move the control points to generate new candidate structures, the changes in the internal fields are negligible compared to the changes in the \mathbf{g}^i . Thus, for each of these candidates,

we hold the internal fields fixed at the values computed at the end of the previous stage.

This approach is useful for two reasons. First, it reduces the number of forward solves associated with each iteration from $8(K-2)n_1n_2$ to just n_1n_2 . Second, by holding the internal fields fixed, (9) and (11) indicate that the unknown expansion coefficients are locally *linearly* related to the data. Thus, estimates of these quantities can be obtained via a linear least squares procedure that is far less demanding than a nonlinear optimization problem. In particular, we have

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \|\mathbf{K}\mathbf{w} - \mathbf{y}\|_2^2 = (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\mathbf{y} \quad (15)$$

where \mathbf{K} is the matrix obtained by stacking all of the real and imaginary parts of $\mathbf{K}^{i,k}$ defined in (11) and (12) and (with a slight abuse of notation) \mathbf{y} is the corresponding stacked real and imaginary parts of $\mathbf{y}^{i,k}$.

Now we address the choice of $\mathbf{a}^{(0)}, \mathbf{c}^{(0)}$. Solving (15) requires \mathbf{K} , which by (11) requires $\mathbf{A}^{i,k}$, which in turn requires $\mathbf{E}^{i,k}$. But computing $\mathbf{E}^{i,k}$ according to §3.2 would require prior knowledge of the unknown \mathbf{g}^i . Thus, to get initial estimates of the expansion coefficients, we set $\mathbf{E}^{i,k}$ to $\mathbf{E}_{unp}^{i,k}$; in other words, we use the Born approximation to determine initial guesses of $\mathbf{a}^{(0)}, \mathbf{c}^{(0)}$.

The algorithm is initialized using an object of size larger than any target of interest located in the vicinity of the true object. There are many possible methods for determining such an initial guess. For example, there are methods that determine a “best fit” disk as an initialization: one based on array processing is described in [28] while another based on statistical hypothesis testing is detailed in [24]. Here, we always assume that the initial object is given and refer the reader to the literature for specific techniques to determine this guess.

The overall algorithm is sketched in Fig. 2. One advantage of this approach is that it is quite easy to implement and in principle, all candidates at any given stage can be generated in parallel. Thus it is computationally attractive. It is not difficult to prove that if we were to recompute the internal fields for each possible control point move (rather than leaving it fixed at the previous

estimate), then we would be guaranteed that at every outer iteration of our algorithm the cost would be non-increasing. By insuring old curves cannot be regenerated¹, this would imply convergence of the algorithm to a minimum of (13) taken over the collection of anomalies that can be generated by these moves. However, since we hold the internal fields fixed as we loop over control point moves and update them only after a move is chosen, monotonicity of the cost cannot be guaranteed. Generally, we have noticed through extensive simulation that as long as the parameters λ_1 and λ_2 are well chosen, the cost does in fact decrease as the iterations progress.

4 Numerical Examples

In this section we present several numerical examples that illustrate the effectiveness of our algorithm. All experiments were done in Matlab using double precision arithmetic. Creation and manipulation of the B-spline curves was achieved with Matlab's Spline Toolbox.

In all of the numerical examples, we used three frequencies (500, 700, and 900 MHz) and three incident angles ($0, \pi/4, -\pi/4$). The region of interest for which we would like to obtain an image was 40cm across (-20cm to 20cm) and 20 cm deep (0 to 20cm). We discretized this region into pixels of size 1cm-by-1cm. Data were collected at 19 receivers located at -18cm to 18cm in 2cm increments. To define the entries of the Green's function matrix G , we set $\epsilon_1 = 2.5\epsilon_0, \sigma_1 = 3E - 3$, which corresponds to assuming that the upper halfspace is air and the lower halfspace is sand. Depending on the example, we fill the $B_i(\mathbf{r})$ matrices of (5) and (6) with monomials of at most degree 2. Finally, Table 2 summarizes the parameters for each of the experiments. The values for the λ_i were chosen by trial and error. In all experiments, the value of the step size, \bar{h} , was taken to be one centimeter. Finally, the contours of all the true objects were generated using a B-spline with $K = 6$ knots while the reconstructions were generated for a $K = 5$ knot contour.

¹It turns out that our approach to updating the internal electric field can create the situation in which distinct curves yield the same cost. Thus we implement extra logic in our implementation to ensure that we never enter a loop.

To produce noisy data, independent additive Gaussian noise was added to both the real and imaginary parts of the noise free data. Letting $\tilde{\mathbf{y}}$ denote the vector obtained by stacking $\mathbf{K}^{i,j}\mathbf{w}$ for all frequencies and incident angles we have $\mathbf{y} = \tilde{\mathbf{y}} + \mathbf{n}$ where $\mathbf{n} = \zeta_r \mathbf{n}_r + j\zeta_i \mathbf{n}_i$ and $\mathbf{n}_r, \mathbf{n}_i$ were generated using Matlab's `randn` function. The constants ζ_r, ζ_i were determined so that the signal-to-noise ratio (SNR) with respect to the real and imaginary parts of $\tilde{\mathbf{y}}$ were the same:

$$SNR = 10 \log_{10} \frac{\text{real}(\tilde{\mathbf{y}})^T \text{real}(\tilde{\mathbf{y}})}{\zeta_r N} = 10 \log_{10} \frac{\text{imag}(\tilde{\mathbf{y}})^T \text{imag}(\tilde{\mathbf{y}})}{\zeta_i N}.$$

We used two measures of success in comparing our reconstructions with the true images. The first is a relative error measure over all pixels in the union of the true curve and the reconstructed curve, which we index by the index set \mathcal{I} :

$$SE = \frac{\|\epsilon_{true}(\mathcal{I}) - \epsilon_p(\mathcal{I})\|^2}{\|\epsilon_{true}(\mathcal{I})\|^2} + \frac{\|\sigma_{true}(\mathcal{I}) - \sigma_p(\mathcal{I})\|^2}{\|\sigma_{true}(\mathcal{I})\|^2}$$

For each example, we also recorded the maximum pointwise relative error in the real (imaginary) part of the solution over the pixels in the intersection of the true and reconstructed curves. The values for these measures for the 5 examples described below are in Table 4.

4.1 Examples 1 and 2: Homogeneous Perturbations

As a first example, we consider the problem with a homogeneous object of unknown contrast embedded in a homogeneous half-space at an SNR of 20 dB, Fig. 3(a)–(b). While the inversion scheme knows the correct basis function to use for the object and background (i.e. $N_a = 1 = N_b$ in (5) and (6) with $b_{1,1} = 1 = b_{2,2}$ in both cases) it still does not know the values of the expansion coefficients. In Fig. 3(e) we display the true boundary of the object along with the initial guess of this quantity and the final boundary estimate produced by our algorithm. The true and estimated images of the ϵ_p and σ_p for this problem are shown in Fig. 3(c)–(d) where we observe that the unknown contrasts are estimated quite accurately. This conclusion is supported by the numerical values for our figures of merit given in Table 4.

For comparison sake, we have also implemented a more traditional Born iterative method (BIM)

[31] inversion procedure using a TSVD (truncated singular value decomposition) regularization technique to stabilize the linear system that must be solved at each iteration. After optimizing the regularization parameter for this procedure, the reconstructions of the real and imaginary parts of \mathbf{g} are shown in Figs. 4.

This example demonstrates that even under noisy circumstances, the approach we have proposed is able to capture accurately both the shape and the numerical values of the unknown permittivity and conductivity of the object. While the BIM clearly indicates the presence of an object in ϵ_p , almost nothing is seen in the σ_p image. Further, note that the anomaly boundary is much less well reconstructed compared to our new approach. Also, numerous artifacts appear in both parts of the reconstruction. By constraining the reconstruction as is done in our algorithm, we obtain a much more accurate representation of the true profile. Our reconstructed object differs from the true by only four pixels. Moreover, the amplitudes of ϵ_p and σ_p are quite close to their true values.

The performance of our approach is further verified in Fig. 5 where we display boundary curves and true and estimated images of ϵ_p and σ_p for an object rotated relative to the interface. The BIM plots are shown in Fig. 6. The same level of performance as was seen in the previous example is also seen here. Moreover, because we reconstruct the boundary explicitly, this approach allows us to easily characterize the orientation of the buried object, a potentially useful piece of information for later processing stages concerned with identification and classification.

4.2 Cramer-Rao Bounds for Examples 1 and 2

We next turn our attention to more challenging problems in which the electrical properties of both the background and the object can vary. A first issue of concern is construction of the B matrices used to model these variations. Intuitively, we expect that that as the size of the object falls, it will be increasingly difficult to recover higher order information about the target structure when the profile of the much larger background region is also unknown. In fact, we can quantify

the validity of this idea using the notion of a Cramer-Rao bound (CRB).

As explained more fully in [30, § 2.4], the CRB provides a lower bound on the variance of any unbiased parameter estimator in a noisy data information extraction procedure of the type considered here. The CRB is a deterministic quantity reflecting the nature of the physical model, the parameterization of the problem, and statistics of the noise corrupting the data. While we make no claims here concerning the bias of our estimator, the CRB provides a useful gauge as to the maximum confidence we should have in the numerical values of the \mathbf{a} and \mathbf{c} estimates produced by our algorithm. Specifically, by examining the bounds for different configurations of object and background, we obtain insight into how we may want to structure our inversion algorithm.

Here we consider our ability to recover information concerning at most linear variations in both the background and the object, i.e. cases in which the real or imaginary parts of these quantities behave as $d_0 + d_1x + d_2z$ where the d_j can represent any of the $a_{i,j}$ or $c_{i,j}$ coefficients appearing in the model of § 2.1. This level of complexity is sufficient to understand the basic issues. In particular, we consider CRB information for objects of three sizes with linear contrast variations (shown in Fig. 7) embedded in backgrounds whose variations are also linear. Fig. 8 illustrates the configurations of interest for the medium size object. The SNR for all experiments is 30 dB.

The numerical values of the square roots of the CRBs (i.e. the lower bounds on the standard deviations) for the experiments are provided in Table 3. The first column refers to the coefficient in the model. For example, the row beginning “Object real: const.” holds the information regarding the true value and the bound on the constant coefficient in the real part of the object. Similarly, “Bkgnd real: x ” is the row for the coefficient governing the real part of the x variation in the background. In all cases, the bounds on the coefficients governing the background are a small fraction of the true values and vary little with the size of the embedded object. These small bounds imply small variance in the estimates of the background structure and thus indicate that we can

in fact determine these quantities rather accurately. This is true both for the constant term in the variation as well as the linear terms. In contrast, our ability to recover anything but the unknown constant term describing the object is more limited. The bounds on the x and z coefficients are at best on the order of the true values and grow at a much more rapid rate as the object size decreases making determination of these quantities a very delicate procedure.

With this in mind, in the remainder of this paper (with the exception of Example 5) we consider object models comprised of only unknown constants. As we show below, even when the true object is of a more complex structure, the use of this simple model still allows for accurate localization and the recovery of limited quantitative information. In light of the underlying objective of our approach, we view this tradeoff of accuracy in “pixel” space for accuracy in localization and geometric characterization as acceptable. In terms of specifying the model for the background, the situation is less clear. Extensive numerical experiments (not reported here) indicate that one needs a relatively accurate model to achieve good localization. That is, the order of the model should approximate well the true distribution of the volume inhomogeneity. As a rigorous solution to the model order determination problem is outside of the scope of this paper, here we consider only models whose order meets or exceeds that of the true distribution.

4.3 Examples 3-5

The next example we consider is illustrated in Fig. 9. Here we have an object with a piecewise constant profile in a background with linear variation. We invert using a model for a constant object and a quadratic background variation. The values for the object are intended to represent a dielectric scatterer with an air gap. It is important to note that this object can never be well represented using our low order polynomial model. The results of inverting with a model employing a constant object with a quadratic background are shown in Figs. 10(c)–(e). The strong localization performance demonstrated in these images in spite of this inherent model mismatch points to the

robustness of our approach to inversion ².

A similar experiment was repeated except that the values for the real and imaginary parts of the background were each corrupted by small amounts of appropriately scaled $[0, 1]$ additive uniform noise. Now neither the object nor the background can be exactly represented using the polynomial model. As shown in Fig. 10 we achieve highly accurate shape information along with useful information concerning the variations in the object as well as the background. Again, these results are indicative of the ability of a low order parameterization to withstand modeling inaccuracies. We note that for both of these examples, the results in Table 4 reflect the fact that we have sacrificed contrast accuracy for geometric fidelity. However, by modeling the contrasts as unknown constant, as expected we obtain estimated values that are close to the average values of the actual perturbations (which can be positive or negative) over the support of the anomaly. In Example 3 the computed real and imaginary perturbations are 8.9E-2 and -2.5E-2, respectively, compared to the means of the true perturbations to the real and imaginary parts, 8.3E-2 and -2.3E-3. Likewise, for Example 4 the calculated real and imaginary parts over the anomaly are 5.8E-3 and -7.01E-3 whereas the means of the true real and imaginary parts are 2E-2 and -2.4E-3.

Finally, in Fig. 11 we consider the problem of recovering a linearly varying object in a linearly varying background. Motivated by the CRB results, the object here is taken to be larger than those of the previous examples so that there will be sufficient signal to allow us to resolve the permittivity variations. As in the other cases, we again achieve strong localization. However there is some noticeable error in the estimates of the permittivity coefficients which is a reflection of the inherent difficulty in accurately obtaining this information. This example points to the need for further work in refining the permittivity estimates after having determined the boundary of the

²For this and the remaining experiments, we do not show the results of the BIM approach to inversion. Even after extensive fine tuning of the algorithm, it was impossible to obtain results comparable to those seen with the currently considered method.

object: model order determination and low order parameterizations should both play a role.

5 Conclusions and Future Work

We presented a new and potentially efficient technique for simultaneously solving the image formation and object characterization problems from scattered electric field data. The key idea was to formulate the perturbation in terms of a small number of parameters via a B-spline representation for the contour of the target. The examples illustrated that our technique can lead to good quality reconstructions: in particular, we found that it was possible to get good localization information even if the background is not homogeneous. A key issue associated with this strong performance is the ability to select good regularization parameters that balance the information content of the data with that of the constraints. In this work, we selected those parameters by hand. Clearly, an important area of future effort is the automation of this process.

Our experiments showed and the computed Cramer-Rao bounds confirmed that it is difficult, if not impossible, to get “linear” or even more complex information about the object unless it was sufficiently large, the noise sufficiently small, and the perturbation in the object was large relative to the perturbation in the background (i.e. the background needs to be nearly homogeneous).

One potential computational advantage to our technique is its inherent parallelism: cost evaluations can be done in parallel, making it computationally feasible to consider more complicated structures. For our reconstruction technique to capture more complex structures, however, we need to consider how to alter the complexity of the boundary representation by inserting and deleting basis elements in the B-spline representation. Solving this order determination problem in an efficient and close to optimal manner is far from trivial. In the future, we will also consider extensions of this work to the case of multiple objects. Finally, we hope to extend the work presented here to the 3D problem and to analyze its performance on real data.

References

- [1] T. S. Angell, Xinming Jiang, and R. E. Kleinman. A distributed source method for inverse acoustic scattering. *Inverse Problems*, (13):531–545, 1997.
- [2] Murat Belge, Eric Miller, and Misha Kilmer. Simultaneous multiple regularization parameter selection by means of the l-hypersurface with applications to linear inverse problems posed in the wavelet transform domain. In *SPIE International Symposium on Optical Science, Engineering, and Instrumentation: Bayesian Inference for Inverse Problems*. SPIE, July 1998.
- [3] M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data, II, Stability and regularisation. *Inverse Probl.*, 4:573–594, 1988.
- [4] Andrew Blake and Michael Isard. *Active Contours*. Springer, 1998.
- [5] W. C. Chew, G. P. Otto, W. H. Weedon, J. H. Lin, C. C. Lu, Y. M. Wang, and M. Moghaddam. Nonlinear diffraction tomography: The use of inverse scattering for imaging. *International Journal of Imaging Systems and Technology*, 7:16–24, 1996.
- [6] W.C. Chew and Y. M. Wang. Reconstruction of two-dimensional permittivity distribution using the distorted born iterative method. *IEEE Trans. Medical Imaging*, 9(2):218–225, June 1990.
- [7] Weng Cho Chew. *Waves and Fields in Inhomogeneous Media*. Van Nostrand Reinhold, New York, 1990.
- [8] Weng Cho Chew and G. Otto. Microwave imaging of multiple conducting cylinders using local shape function. *IEEE Microwave Guided Wave Letters*, 2:284–284, 1992.
- [9] Chien-Ching Chiu and Yean-Woei Kiang. Electromagnetic imaging for and imperfectly conducting cylinder. *IEEE Trans. Microwave Theory and Techniques*, 39(9):1632–1639, September 1991.
- [10] Chien-Ching Chiu and Yean-Woei Kiang. Inverse scattering of a buried conducting cylinder. *Inverse Problems*, 7:187–202, 1991.
- [11] Chien-Ching Chiu and Yean-Woei Kiang. Electromagnetic inverse scattering of a conducting cylinder buried in a lossy half-space. *IEEE Trans. on Antennas and Propagation*, 40(12), Augusts 1992.
- [12] Chien-Ching Chiu and Yean-Woei Kiang. Microwave imaging of multiple conducting cylinders. *IEEE Trans. on Antennas and Propagation*, 40(8), August 1992.
- [13] David Colton and Andreas Kirsch. A simple method for solving inverse scattering problems in the resonance region. *Inverse Problems*, 12:383–393, 1996.
- [14] David Colton and Peter Monk. The detection and monitoring of leukemia using electromagnetic waves: mathematical theory. *Inverse Problems*, 10:1235–1251, 1994.
- [15] David Colton, Michele Piana, and Roland Potthast. A simple method using Morozov’s discrepancy principle for solving inverse scattering problems. *Inverse Problems*, (13):1477–1493, 1997.

- [16] Per Christian Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580, December 1992.
- [17] R. F. Harrington. *Field Computations by Moment Methods*. Macmillan, New York, 1968.
- [18] Misha Kilmer, Eric Miller, and Carey Rappaport. Preconditioners for structured matrices arising in subsurface object detection. *Journal of Computational Physics*. in review.
- [19] Andreas Kirsch. Characterization of the shape of a scattering obstacle using the spectral data of the far field operator. *Inverse Problems*, 14:1489–1512, 1998.
- [20] Jin Au Kong. *Electromagnetic Wave Theory*. John Wiley and Sons, 1986.
- [21] G. Kristensson and C. R. Vogel. Inverse problems for acoustic waves using the penalized likelihood method. *Inverse Problems*, 2:461–479, 1986.
- [22] A. Litman, D. Lesselier, and F. Santosa. Reconstruction of a two-dimensional binary obstacle by controlled evolution of a level-set. *Inverse Problems*, 14:685–706, 1998.
- [23] Pierre Lobel, Laure Blanc-Feraud, Christian Pichot, and Michel Barlaud. New regularization scheme for inverse scattering. *Inverse Problems*, 13(2):403–410, April 1997.
- [24] Eric L. Miller. Statistically based methods for anomaly characterization in images from observations of scattered radiation. *IEEE Trans. on Image Processing*, 8(1):92–101, January 1999.
- [25] Eric L. Miller, Lena Nicolaides, and Andreas Mandelis. Nonlinear inverse scattering methods for thermal wave slice tomography: A wavelet domain approach. *Journal of the Optical Society of America (A)*, 1998.
- [26] Lena Nicolaides and Andreas Mandelis. Image-enhanced thermal-wave slice diffraction tomography with numerically simulated reconstructions. *Inverse Problems*, 13(5):1339–1412, 1997.
- [27] Christian Pichot, Pierre Lobel, Cedric Dourthe, Laure Blanc-Feraud, and Michel Barlaud. Microwave inverse scattering: Quantitative reconstruction of complex permittivity for different applications. *IEICE Transactions on Electronics*, E80-C:1343–1348, November 1997.
- [28] Adnan Sahin and Eric L. Miller. Electromagnetic scattering-based array processing methods for near-field object characterization. *Journal of Electromagnetic Waves and Applications*, 13:1209–1236, 1999.
- [29] L. Souriau, B. Duchene, D. Lesselier, and R. E. Kleinman. Modified gradient approach to inverse scattering for binary objects in stratified media. *Inverse Problems*, (12):463–481, 1996.
- [30] Harry L. Van Trees. *Detection, Estimation and Modulation Theory: Part I*. John Wiley, New York, 1968.
- [31] Y. M. Wang and W. C. Chew. An iterative solution to the two-dimensional electromagnetic inverse scattering problem. *International Journal of Imaging Systems and Technology*, 1(1):100–108, 1989.
- [32] William H. Weedon and Weng Cho Chew. Time-domain inverse scattering using the local shape function (LSF) method. *Inverse Problems*, 9:551–564, 1993.

K	total number of control points ($K - 2$ unique)
n_1	number of frequencies used to probe the earth
n_2	number of angles used to probe the earth
N	number of unknowns in subsurface discretization
M	number of data points for given θ, ω
NM	total number of data points
N_a	number columns in \mathbf{B}_1
N_b	number columns in \mathbf{B}_2

Table 1: Summary of index parameters. Note that the degrees of polynomials used to represent the object and the background are $N_a - 1$ and $N_b - 1$, respectively.

Example	SNR	True Obj.	True Back	Obj.	Back	λ_1	λ_2
1	20dB	0	0	0	0	3.28E-4	1.59E-4
2	20dB	0	0	0	0	6.0E-5	6.0E-5
3	30dB	PWC	1	0	2	1.55E-4	7.9E-5
4	20dB	PWC	1,rand	0	2	9.0E-5	6.0E-5
5	30dB	1	0	1	0	4.0762E-4	1.7015E-4

Table 2: Summary of experiments. Numbers in columns 3 to 6 indicate the degree of the polynomial used to generate the data (columns 3,4) or to reconstruct (columns 5,6). “PWC” indicates that the true object had piecewise constant perturbations in both the real and imaginary parts. “rand” indicates that random uniform perturbations were added to the real and imaginary parts of the background when generating data.

Coefficient		(True)	(Large)	(Medium)	(Small)
		Value	Bound	Bound	Bound
Object	real: const	4E-1	7.43E-2	3.42E-1	4.27
	real: x	1	1.02	2.56	4.57E+1
	real: z	8	9.17E-1	5.18	9.29E+1
	imag: const	-2.2E-3	3.09E-3	3.11E-2	8.57E-1
	imag: x	1E-2	4.22E-2	1.08E-1	2.05
	imag: z	4E-2	4.98E-2	7.09E-1	2.06E+1
Bkgnd	real: const	-1E-1	2.42E-2	2.14E-2	1.71E-2
	real: x	1	1.72E-1	1.58E-1	1.51E-1
	real: z	4	2.09E-1	2.00E-1	1.89E-1
	imag: const	-1E-3	8.93E-4	7.99E-4	7.32E-4
	imag: x	0	6.95E-3	6.15E-3	5.80E-3
	imag: z	1E-2	8.25E-3	7.45E-3	7.35E-3

Table 3: Values of the bounds for the case when the background and the object BOTH vary linearly (at 30dB). Note that the bounds for all 3 linear terms for the object dramatically increase as the size of the object decreases, whereas the bounds for the background decrease only slightly.

Ex.	SE	RE (real)	RE (imag)
1	0.335	0.110	0.607
2	0.593	0.367	0.616
3	1.051	1.059	0.172
4	6.182	1.191	2.565
5	1.447	3.185	0.085

Table 4: Measures of success in reconstruction. First column gives the experiment number and second column gives the error measure defined in the beginning of § 4. Last two columns give maximum pointwise relative error for the real and imaginary parts, respectively, over the pixels in the intersection between the true and reconstructed images.

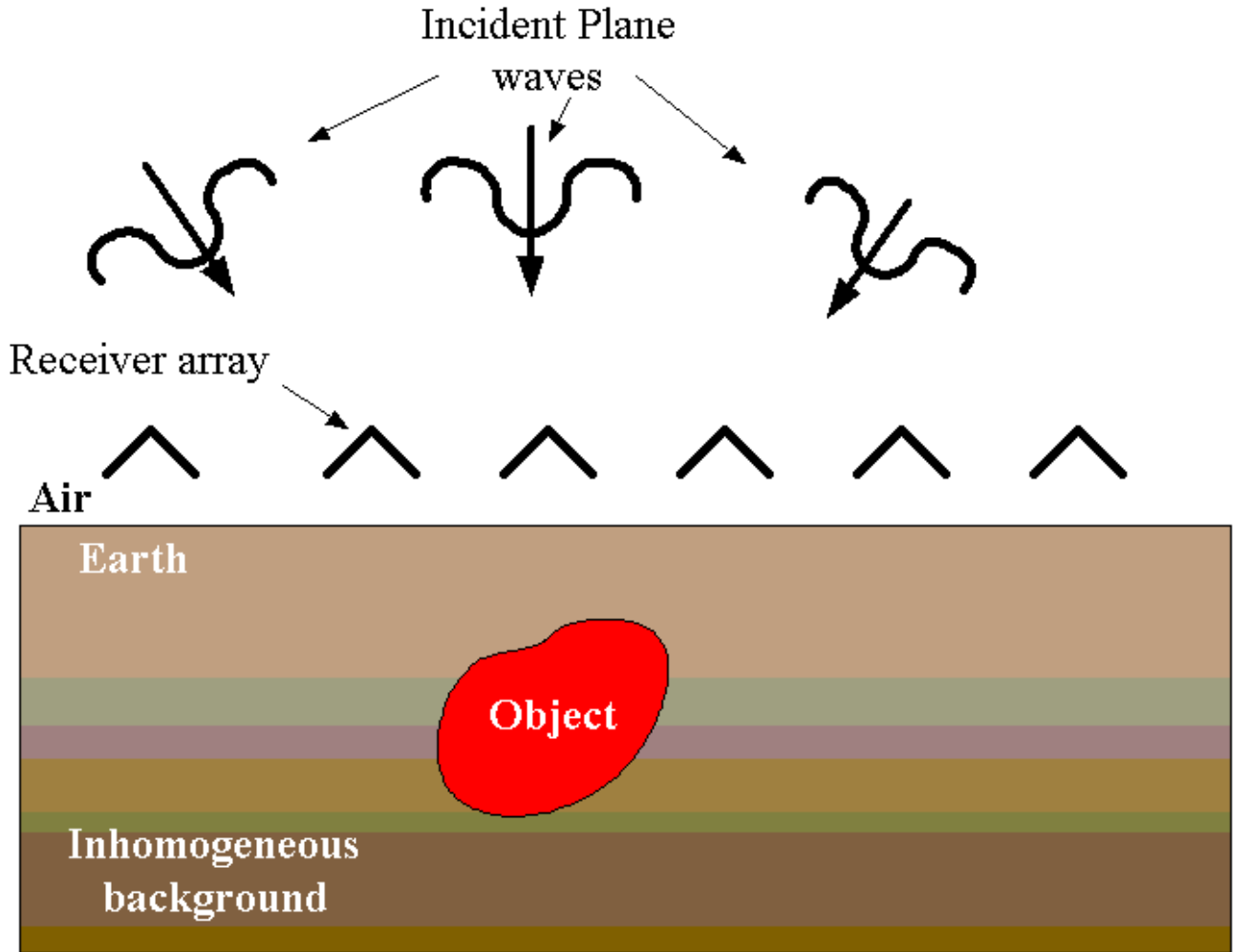


Figure 1: Experimental configuration for general problem of interest

Algorithm 1

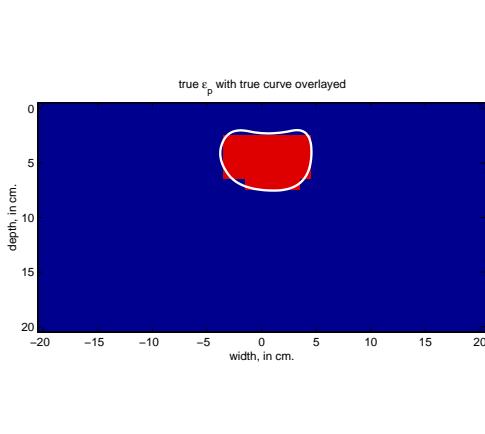
```

 $k := 1$ 
Given initial estimates  $[b^{(1)}(s), \mathbf{a}^{(1)}, \mathbf{c}^{(1)}]$  compute  $\mathbf{E}^{i,j}$  and  $\mathbf{A}^{i,j}$  for  $1 \leq i \leq n_1, 1 \leq j \leq n_2$ 

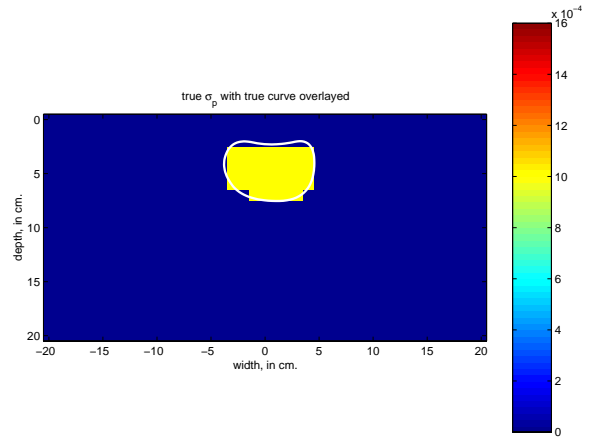
While (current-cost can still be reduced) do
    • For  $i = 0, \dots, K - 3$  do
        1. Select control point  $(x_i, z_i)$ 
        2. For each of 8 moves of  $(x_i, z_i)$  by  $\bar{h}$  do
            (a) Update  $(x_i, z_i)$  by a move
            (b) Form candidate contour,  $b(s)$ , from other control points and new version of  $(x_i, z_i)$ 
            (c) Generate estimates  $\hat{\mathbf{a}}, \hat{\mathbf{c}}$  using current  $\mathbf{A}^{i,j}$  (see (15))
            (d)  $\text{cost} := J(b(s), \hat{\mathbf{a}}, \hat{\mathbf{c}})$  (see (9) and (13))
            (e) If  $\text{cost} < \text{current-cost}$ 
                 $b^{(k)}(s) := b(s); \text{current-cost} := \text{cost}; [\mathbf{a}^{(k)}, \mathbf{c}^{(k)}] := [\hat{\mathbf{a}}, \hat{\mathbf{c}}]$ 
            (f) Elseif  $\text{cost} = \text{current-cost}$ 
                If  $b(s)$  is different from all previous  $b^{(k)}(s)$ ,
                 $b^{(k)}(s) := b(s); [\mathbf{a}^{(k)}, \mathbf{c}^{(k)}] := [\hat{\mathbf{a}}, \hat{\mathbf{c}}]$ 
        • From  $[b^{(k)}(s), \mathbf{a}^{(k)}, \mathbf{c}^{(k)}]$  compute  $\mathbf{E}^{i,j}$  and  $\mathbf{A}^{i,j}$  for  $1 \leq i \leq n_1, 1 \leq j \leq n_2$ 
        • Update  $[\mathbf{a}^{(k)}, \mathbf{c}^{(k)}]$  using the new  $\mathbf{A}^{i,j}$ 
        •  $\text{current-cost} := J(b^{(k)}(s), \mathbf{a}^{(k)}, \mathbf{c}^{(k)})$ 
        •  $k := k + 1$ 
end while

```

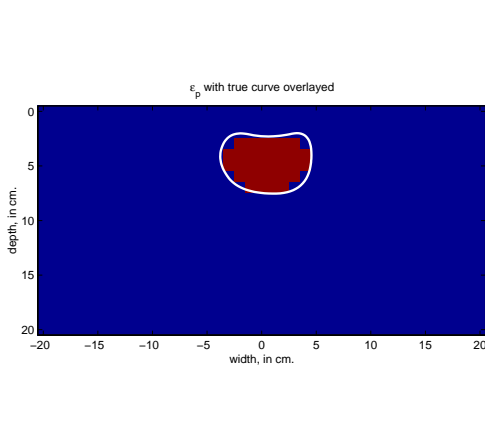
Figure 2: Anomaly Recovery Algorithm



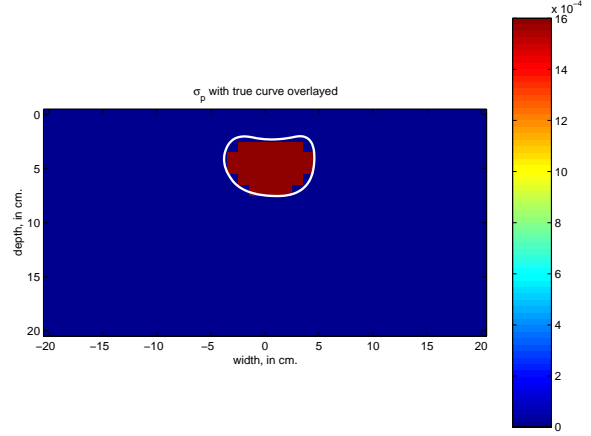
(a) True distribution for $\epsilon_p(\mathbf{r})$



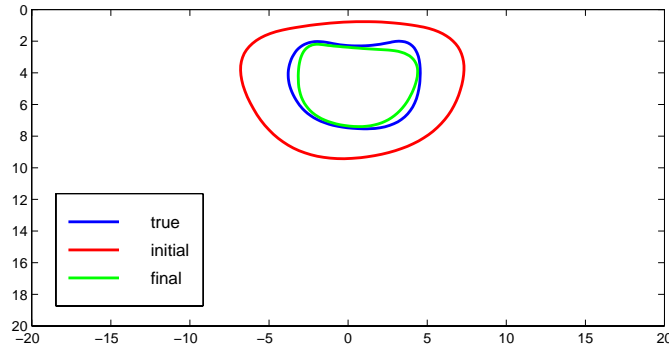
(b) True distribution for $\sigma_p(\mathbf{r})$



(c) Estimate of $\epsilon_p(\mathbf{r})$

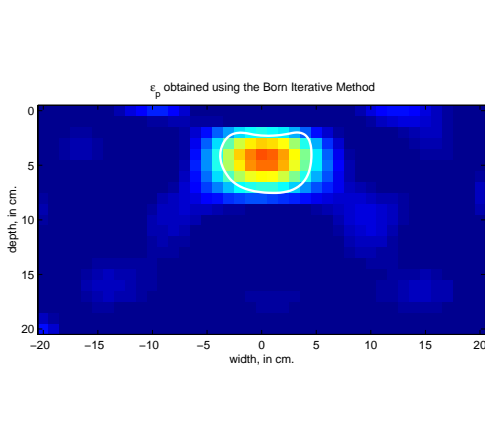


(d) Estimate of $\sigma_p(\mathbf{r})$

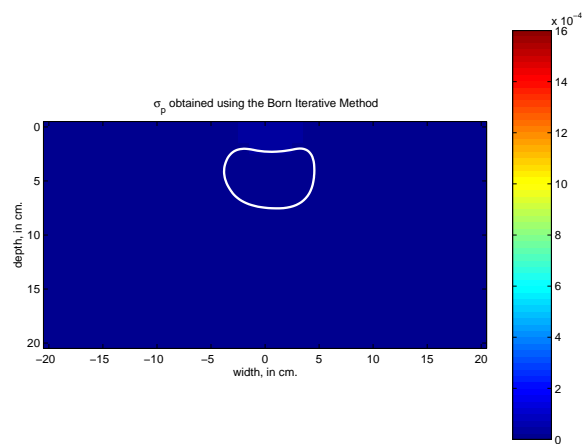


(e) Initial guess, final estimate, and true object boundary

Figure 3: True distributions, estimates, and boundary structures for example 1 (parameters given in Table 2).

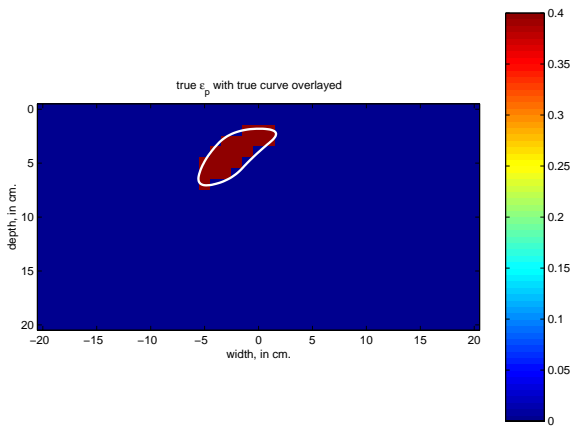


(a) Estimate of $\epsilon_p(\mathbf{r})$

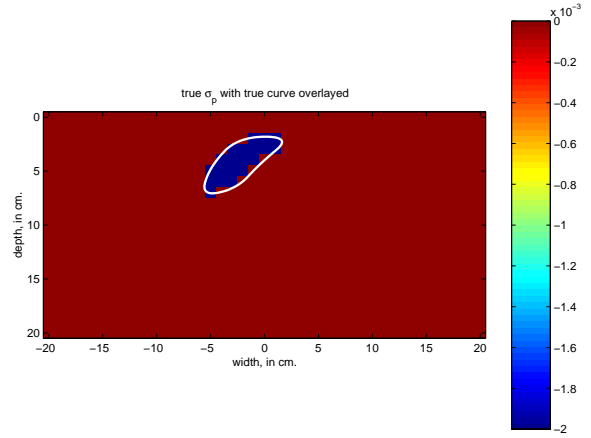


(b) Estimate of $\sigma_p(\mathbf{r})$

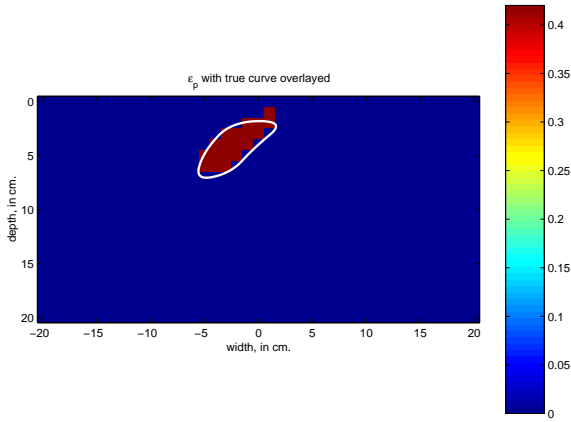
Figure 4: Born iterative results for example 1



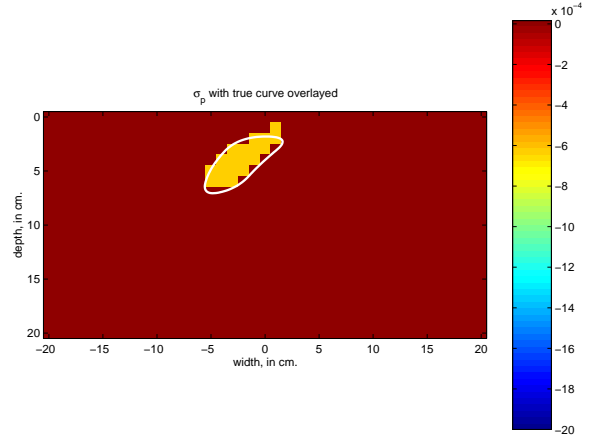
(a) True distribution for $\epsilon_p(\mathbf{r})$



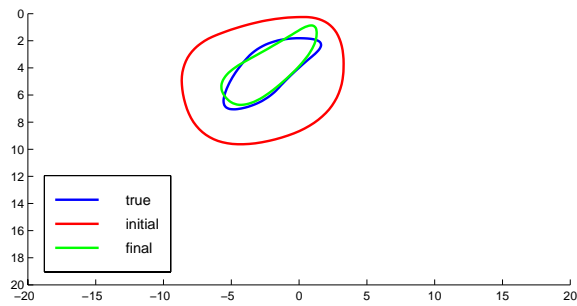
(b) True distribution for $\sigma_p(\mathbf{r})$



(c) Estimate of $\epsilon_p(\mathbf{r})$

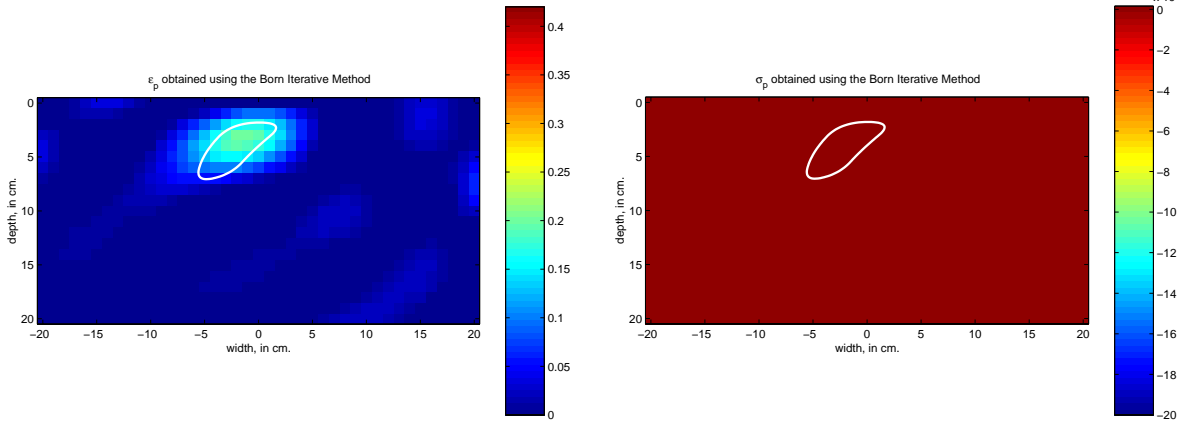


(d) Estimate of $\sigma_p(\mathbf{r})$



(e) Initial guess, final estimate, and true object boundary

Figure 5: True distributions, estimates, and boundary structures for example 2 (parameters given in Table 2).



(a) Estimate of $\epsilon_p(\mathbf{r})$

(b) Estimate of $\sigma_p(\mathbf{r})$

Figure 6: Born iterative results for example 2.

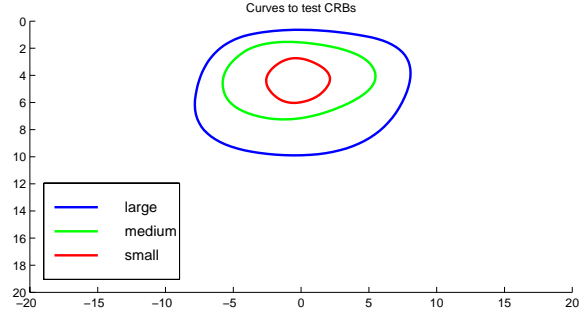
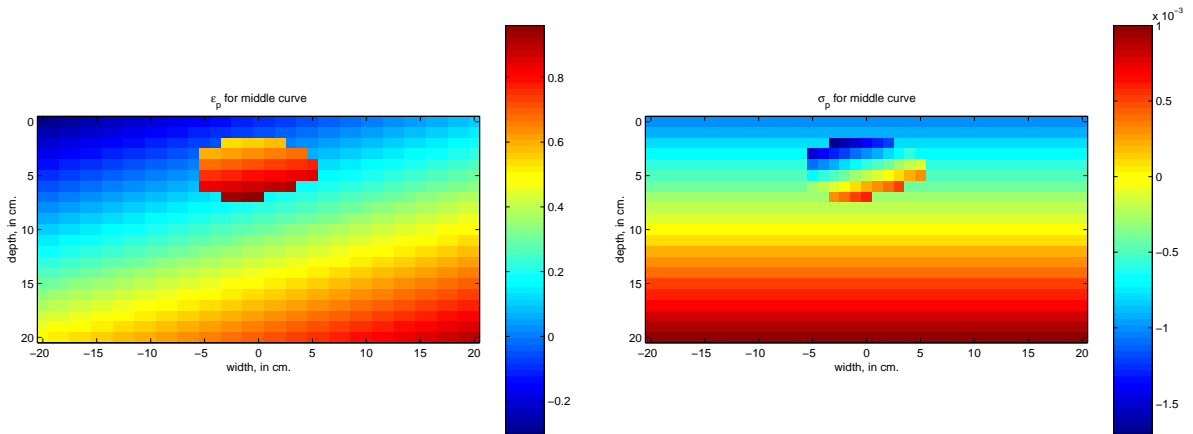


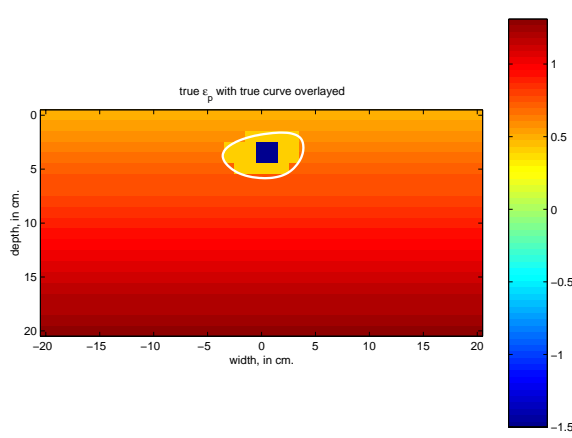
Figure 7: Three boundary curves used to evaluate Cramer Rao bound information



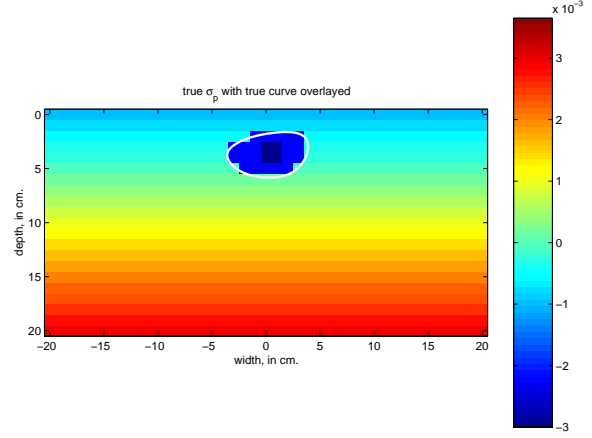
(a) $\epsilon_p(\mathbf{r})$ for linear object on linear background.

(b) $\sigma_p(\mathbf{r})$ for linear object on linear background.

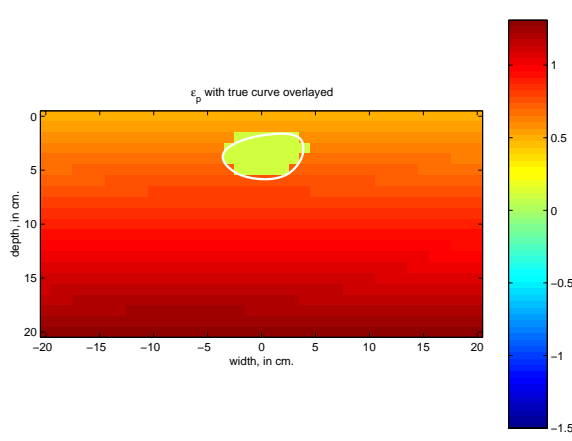
Figure 8: Permittivity distributions for medium size object CRB analysis.



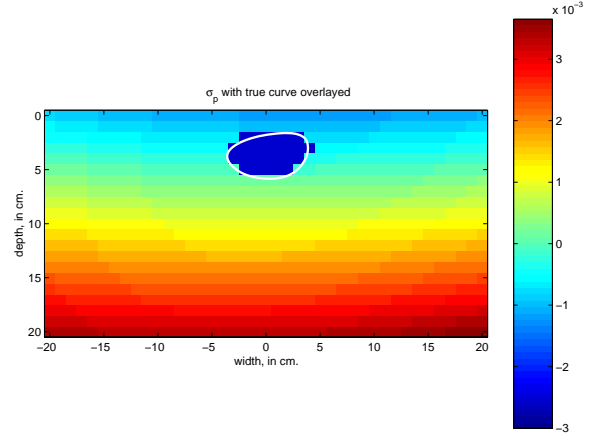
(a) True distribution for $\epsilon_p(\mathbf{r})$



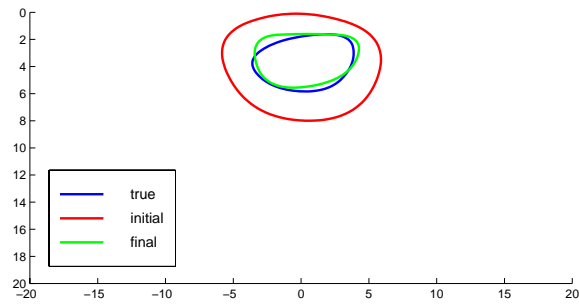
(b) True distribution for $\sigma_p(\mathbf{r})$



(c) Estimate of $\epsilon_p(\mathbf{r})$

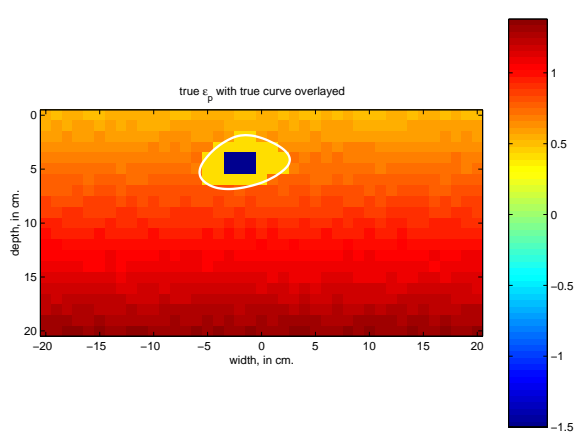


(d) Estimate of $\sigma_p(\mathbf{r})$

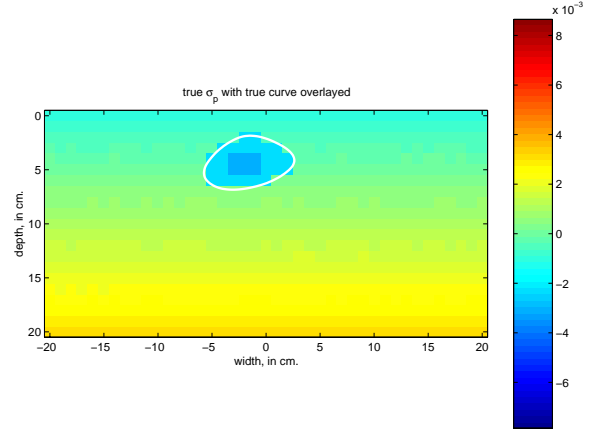


(e) Initial guess, final estimate, and true object boundary

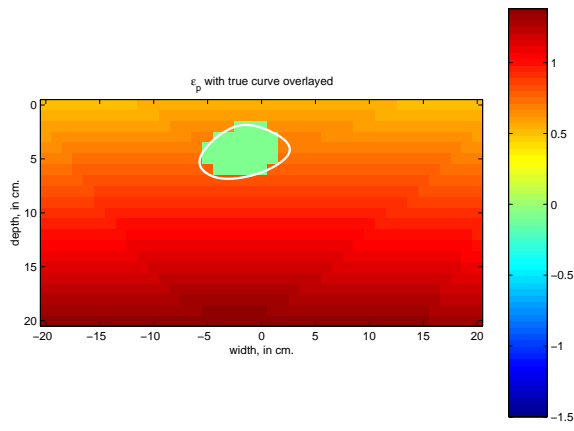
Figure 9: True distributions, estimates, and boundary structures for example 3 (parameters given in Table 2).



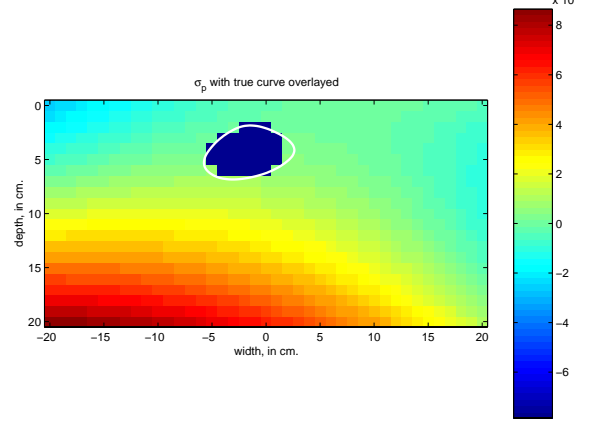
(a) True distribution for $\epsilon_p(\mathbf{r})$



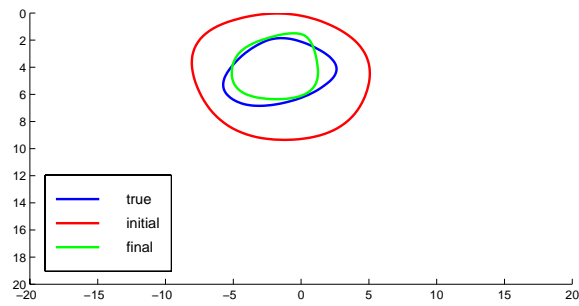
(b) True distribution for $\sigma_p(\mathbf{r})$



(c) Estimate of $\epsilon_p(\mathbf{r})$

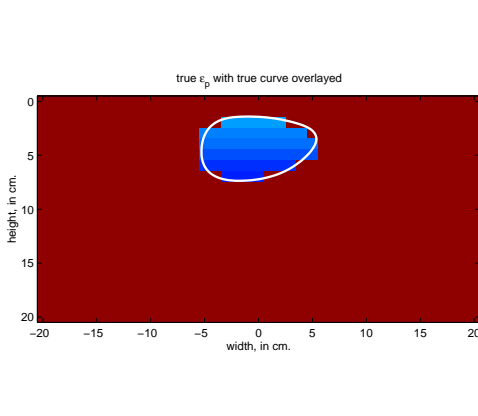


(d) Estimate of $\sigma_p(\mathbf{r})$

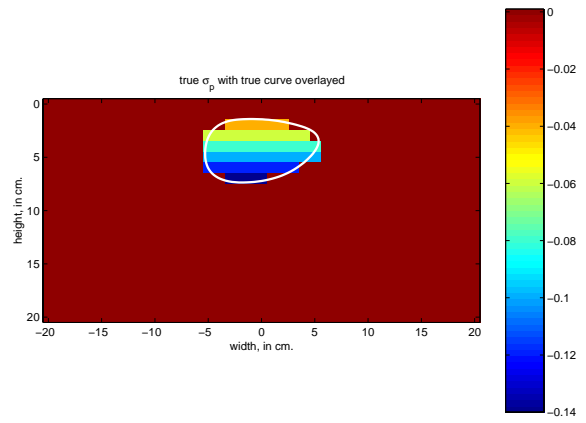


(e) Initial guess, final estimate, and true object boundary

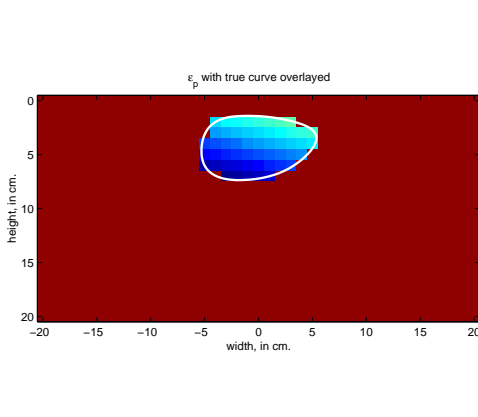
Figure 10: True distributions, estimates, and boundary structures for example 4 (parameters given in Table 2).



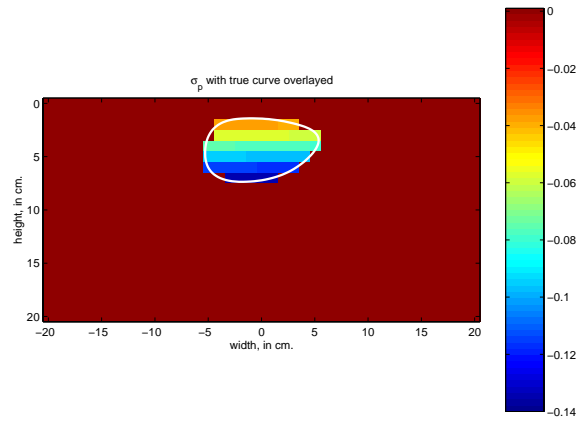
(a) True distribution for $\epsilon_p(\mathbf{r})$



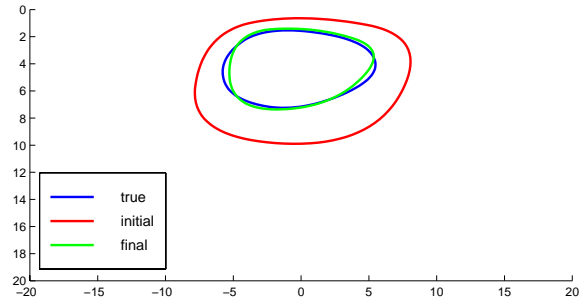
(b) True distribution for $\sigma_p(\mathbf{r})$



(c) Estimate of $\epsilon_p(\mathbf{r})$



(d) Estimate of $\sigma_p(\mathbf{r})$



(e) Initial guess, final estimate, and true object boundary

Figure 11: True distributions, estimates, and boundary structures for example 5 (parameters given in Table 2).

An adaptive, multiscale inverse scattering approach to photothermal depth profilometry

Eric L. Miller and Ibrahim Yavuz
Dept. of Electrical and Computer Engineering
Northeastern University,
235 Forsyth Building
Boston MA, 02115, USA
Tel: (617) 373-8386 Fax: (617) 373-8627
e-mail: elmiller@ece.neu.edu

Lena Nicolaidis and Andreas Mandelis
Photothermal and Optoelectronic Diagnostics Laboratories
Dept. of Mechanical Engineering, Univ. of Toronto
5 King's College Road, Toronto M5S 3G8, CANADA
Phone/Fax: (416) 978-5106
e-mail: mandelis@mie.utoronto.ca

November 3, 1999

Abstract

Photothermal depth profilometry is formulated as a non-linear inverse scattering problem. Starting with the one dimensional heat diffusion equation, we derive a mathematical model relating arbitrary variation in the depth-dependent thermal conductivity to observed thermal wavefields at the surface of a material sample. The form of the model is particularly convenient for incorporation into a non-linear optimization framework for recovering the conductivity based on thermal wave data obtained at multiple frequencies. We develop an adaptive, multi-scale algorithm for solving this highly ill-posed inverse problem. The algorithm is designed to produce an accurate, low order representation of the thermal conductivity by automatically controlling the level of detail in the reconstruction. This control is designed to reflect both (1) the nature of the underlying physics which says scale should decrease with depth and (2) the particular structure of the conductivity profile which may require a sparse collection of fine scale components to adequately represent significant features such as a layering structure. The approach is demonstrated in a variety of synthetic examples representative of non-destructive evaluation problems seen in the steel industry.

Keywords: Adaptive signal representation, multiscale methods, B-splines, non-linear inverse scattering, photothermal depth profiling.

Running title: Multiscale adaptive inverse scattering

1 Introduction

One of the most promising applications of photothermal non-destructive evaluation (NDE) technology is in the area of *depth profilometry* (DP) [4, 8–13]. This technique uses a modulated laser beam to induce thermal waves in a material sample with the resulting wavefield amplitude and phase measured at the surface. The experiment is repeated for a number of modulation frequencies yielding a spectrum of observations. Assuming that the structure varies only with depth, the problem is to convert the measured spectrum into a depth-varying profile of a relevant physical property such as thermal conductivity [8], thermal effusivity [9], or thermal diffusivity [10, 12, 13].

Considerable work has been done in developing physical models and algorithms based on these models for reconstructing profiles. All models reported in the literature are based on the one dimensional, heat diffusion equation (HDE). In much of this work, the material is considered to be composed of layers where the physical properties assume simple functional forms. For piecewise constant layers, the HDE can be solved in terms of a recursively defined generalized reflection coefficient [4, 9]. In the limit of infinitesimally thin layers, the recursion converges to a non-linear Riccati equation. For finitely thick, piecewise linear layers the solution is given via a different recursively defined reflection coefficient [8].

An alternate approach based on the notion of a thermal harmonic oscillator (THO) has been pioneered by Mandelis and collaborators [11, 12]. Making a WKB approximation for variations of the thermal effusivity yields solutions to the HDE in terms of integrals of the thermal diffusivity. Using a particular functional form for this profile, analytic expressions have been reported for a semi-infinite medium and a finite depth inhomogeneity backed by a semi-infinite substrate [10].

Each of these physical models has been used as the basis for an inversion algorithm. Under suitable approximations (most notably, ignoring the non-linear term) the Riccati equation model reduces to a Laplace transform type relationship between the derivative of the log-effusivity and

the data [4, 9]. Tikhonov regularized, linear inversion methods are then used to determine the log effusivity [9]. While the details of the inversion methods for the piecewise constant and THO models are different, they both share a common structure in which the data are processed progressively from high to low modulation frequencies. The higher frequencies are used to estimate the profile structure in slices of the material closest to the surface while the lower frequency data provide information concerning the deeper variations. For the piecewise constant profile method, the width of these slices is determined *a priori* [8]. The THO method locally fits profiles of a predefined structure to slices whose widths are determined adaptively as the algorithm progresses [10, 13].

We pursue a different approach to the DP problem. Rather than considering solutions to the HDE for specific profiles or approximations to the physics, we employ a numerical implementation of the HDE which allows for exact solutions (up to discretization error) for arbitrary profiles. In particular, it is not necessary to assume that the thermal conductivity is slowly varying. The price for this flexibility is increased computational complexity. Thus, we also introduce a highly efficient, non-linear approximation to the exact solution. For the problem of interest in this paper, we show that this method, which is similar in nature to the Extended Born Approximation [5, 18], is about 20 times faster than the exact model with a loss in accuracy close to machine precision.

The use of this physical model allows us to formulate the profile reconstruction process as a non-linear inverse scattering problem in a highly lossy medium. Problems of this type are known to be ill-posed in that small perturbations in the data can lead to large amplitude, non-physical artifacts in a reconstruction. To stabilize the inversion process, one typically uses a *regularization* procedure. Our previous efforts in this field have included the use of wavelet-based regularization techniques for two dimensional, single frequency, non-linear inverse scattering problems [14, 15]. The use of wavelets was motivated by analysis which indicated that a given level of accuracy in a reconstruction required that the resolution in the estimated profile be space varying. The

data and the physics supported fine scale estimates in regions close to the detectors while coarser scale information could only be recovered further into the material sample. Representation of the unknown image using an orthonormal wavelet basis provided a convenient means of enforcing this variable resolution. Regularization was achieved in part by concentrating the information in the data on the recovery of a low order representation of the unknown comprised of relatively few fine scale wavelet coefficients supplemented by a small number of coarse scale coefficients

For the DP problem we build on this work using an alternate multiscale representation of the unknown profile. Rather than an orthonormal basis of wavelets, we use a collection of spline functions possessing a natural multiscale structure. Specifically, coarse scale elements can be expressed as linear combinations of finer scale splines [17]. By surrendering the orthonormality property of wavelets, we have a more flexible method for modeling the unknown profile.

As in [14] our inversion approach here is designed to produce a low order reconstruction in which the distribution of fine scale detail is determined in an automatic, controlled, and rational manner. The precise method for accomplishing this, however, is substantially different from our previous wavelet related efforts. Starting with a coarse scale set of functions, we iteratively refine the reconstruction to (a) add detail and then (b) prune away unnecessary degrees of freedom to obtain a more parsimonious description of the profile. The final estimate is comprised of functions at many scales whose spatial distribution is dictated adaptively by the data and the physical model. Like the results in e.g. [10] we tend to see fine scale detail close to the surface with coarser scale elements used deeper into the material. However, our approach involves no decomposition of the material into slices (predefined or virtual). Unlike the techniques in [8, 10] where low frequency data play no role in the structure of the reconstruction near the surface, here the data from all frequencies impact the entire structure of the estimated profile.

In §4 we show that limiting the number of degrees of freedom in the reconstruction is useful

for a number of reasons. First, the low dimensionality can lead to computational efficiencies. We demonstrate via numerical experiments that the adaptive algorithm can be substantially faster than a brute-force, fine scale reconstruction with little loss in accuracy. Additionally, our approach can lead to better reconstructions. As described in greater detail in §3.1, inversion requires the solution of a non-linear optimization problem and is thus prone to convergence to local minima of an associated cost function. By constraining the reconstruction using our adaptive approach, we show that it is also possible to converge to a lower cost point in solution space than is the case for a high dimensional, fine scale approach to inversion. Currently, we have observed these advantages only through simulation. An area of work is in the development of a more rigorous theory to understand the conditions where we might expect improved computational and/or reconstruction performance.

The remainder of this paper is organized as follows. In § 2, the physical model is developed. Our approach to inversion is provided in § 3. Here we formulate the basic problem in an inverse scattering context, define the spline functions, and describe the details of the adaptive algorithm. In § 4 we present examples of this method. Conclusions and future work are the subject of § 5.

2 Formulation of the Depth Profilometry Problem

2.1 Continuous Formulation

The DP problem of interest in this paper is the reconstruction of the thermal diffusivity profile based on observations of thermal waves obtained from illumination by a modulated laser source. The starting point is the one dimensional HDE with boundary conditions stated as follows:

$$\frac{d}{dz}\kappa(z)\frac{d}{dz}T(z) - i\omega\rho(z)c(z)T(z) = 0 \tag{1}$$

$$-\kappa(z)\frac{d}{dz}T(z) = \frac{1}{2}Q \quad z = 0 \tag{2}$$

$$T(z) = 0 \quad z \rightarrow \infty \tag{3}$$

where κ is the thermal conductivity, ρ the density and c the specific heat. $T(z)$ is the thermal wavefield, ω the angular modulation frequency, and Q the intensity of the laser light incident at the surface which is taken to be at $z = 0$. An $e^{i\omega t}$ time dependence for the source and T is assumed.

To produce a form of the model useful in an inverse scattering context we start by decomposing $\kappa(z)$ into the sum of a background component, $\kappa_b(z)$, and a perturbation of arbitrary magnitude, $\kappa_p(z)$. Generally the background is taken to be a nominal profile such as a halfspace as in [8, 12] or a well characterized material sample backed by air [10]. The goal of the inverse problem is the reconstruction of $\kappa_p(z)$ ¹. Inserting $\kappa = \kappa_b + \kappa_p$ into (1), we write the overall model as

$$(\mathcal{F}_b + \mathcal{F}_p)T = v \quad (4)$$

where we have defined the matrices of linear operators, \mathcal{F}_b and \mathcal{F}_p , and the source vector, v , as

$$\mathcal{F}_b = \begin{bmatrix} \frac{d}{dz}\kappa_b(z)\frac{d}{dz} - i\omega\rho(z)c(z) \\ b_0 \\ b_\infty \end{bmatrix} \quad \mathcal{F}_p = \begin{bmatrix} \frac{d}{dz}\kappa_p(z)\frac{d}{dz} \\ 0 \\ 0 \end{bmatrix} \quad v = \begin{bmatrix} 0 \\ \frac{1}{2}Q \\ 0 \end{bmatrix}. \quad (5)$$

The linear operator b_0 enforces the boundary condition, (2), and b_∞ implements (3).

We manipulate (4) to obtain a solution of the HDE in the form

$$T = (I + \mathcal{G}_b\mathcal{F}_p)^{-1}\mathcal{G}_bv \quad (6)$$

with $\mathcal{G}_b \equiv \mathcal{F}_b^{-1}$ the Green's operator for the boundary value problem (1)–(3) with $\kappa = \kappa_b$. Borrowing from the inverse scattering literature, the field computed with $\kappa_p = 0$ is termed the background field and is given as $T_b \equiv \mathcal{G}_bv$. The scattered field $T_s \equiv T - T_b$ is then

$$T_s = \mathcal{G}_b\mathcal{F}_pT = \mathcal{G}_b\frac{d}{dz}\kappa_p\frac{d}{dz}T \quad (7)$$

$$= \mathcal{G}_b\mathcal{F}_p(I + \mathcal{G}_b\mathcal{F}_p)^{-1}\mathcal{G}_bv. \quad (8)$$

where (7) follows from the definition of \mathcal{F}_p in (5) and (6) is used to obtain (8).

¹In theory, the specification of $\kappa_b(z)$ is somewhat arbitrary as the algorithm is designed to produce a κ_p such that $\kappa_b + \kappa_p$ is as accurate as possible. However, making κ_p “small” through as precise a description of κ_b as is possible does practically improve the convergence of the inversion.

The goal of the DP problem is to recover κ_p from observations of T_s taken at $z = 0$ for multiple modulation frequencies. In preparation for a description of the inversion algorithm in § 3, (7) indicates that $T_s(0)$ can be written as a linear functional of κ_p in which the functional explicitly depends on \mathcal{G}_b and dT/dz . Because $T(z)$ itself is dependent on κ_p , the functional also implicitly depends on the thermal conductivity. Thus letting y_k be the complex value datum taken at the k th frequency ω_k , we write

$$y_k = c_k(\kappa_p)\kappa_p \quad (9)$$

where c_k is the linear functional obtained from (7). In § 3.2 we provide an explicit formula for c_k in the context of a discretized model. The data vector comprised of y_k for $k = 1, 2, \dots, N_y$ is

$$y = \mathcal{C}(\kappa_p)\kappa_p \quad (10)$$

with \mathcal{C} being the column vector of c_k 's. The DP problem of interest in this work then is the recovery of κ_p from knowledge of y as well as the HDE model.

2.2 Discretization

To implement a solution to the DP problem requires a discretized form of the physical model. Here we use a standard first-order finite difference scheme for the HDE in (1)–(3). The field $T(z)$, and the physical properties $\kappa(z)$, $\rho(z)$ and $c(z)$ are sampled on a uniform grid of spacing h . First order differentiation w.r.t z then is represented using a bi-diagonal matrix, D_z , with $1/h$ on the main diagonal and $-1/h$ on the first super diagonal. For the homogeneous Neumann boundary condition, (3), we set to zero the last element in the vector of samples for the field T where we ensure that the boundary is taken far enough from the surface so that numerical artifacts are negligible.

To construct, the discrete form of the c_k in (9), consider the expression for T_s in (7). Under the discrete model we have

$$T_{s,k} = G_{b,k} D_z \mathcal{D}(\kappa_p) D_z T_k \quad (11)$$

with $G_{b,k}$ the matrix form of Green's operator for the k -th frequency, T_k the vector of samples of

the thermal field and similarly for κ_p and $T_{s,k}$. For a vector x , $\mathcal{D}(x)$ is the diagonal matrix whose (ii) th element is x_i . After some tedious but elementary algebra, it is not hard to show that from (11) we can write the surface value of the scattered wavefield under the discretized model as

$$T_s(z=0) = G_{b,k}(1, :)\mathcal{D}(T_k)D_z\mathcal{D}(D_zT_k)\kappa_p \equiv c_k(\kappa_p)\kappa_p \quad (12)$$

with $G_{b,k}(1, :)$ being the first row of the matrix G_b . From (12) we conclude that the discrete representation for the linear functional $c_k(\kappa_p)$ in (9) is the row vector $G_{b,k}(1, :)\mathcal{D}(T_k)D_z\mathcal{D}(D_zT_k)$.

The most computationally intensive part of this forward model is determination of the vector T_k . From (6), the discretized form of this calculation requires the solution of the linear system

$$(I + G_{b,k}F_p)T_k = G_{b,k}v. \quad (13)$$

Because the inversion algorithm requires systems of this form be solved hundreds if not thousands of times, there is significant motivation for seeking accurate approximations to the determination of T_k . The similarity of the DP problem to inverse electrical conductivity problems found in the geophysical community, lead us to consider a so-called Extended Born Approximation (EBA) to (13). A detailed theoretical treatment of this method may be found in [5, 18]. The practical implementation of the EBA for our problem amount to replacing the full matrix $I + G_{b,k}F_p$ by only its diagonal elements thereby greatly reducing the number of operations need to determine T_k . In § 4, we provide experimental verification of both the accuracy and efficiency of this method.

3 A Scale-Adaptive Algorithm for the DP Problem

We start by considering the fixed-scale solution to the DP problem in which we seek to recover κ_p directly from the data. The machinery developed here then forms the basis for the adaptive method developed later in this section.

3.1 The Fixed-Scale Solution

As is typically done for an inverse problem of the type considered here [15, 16, 18], we use our discretized model to define the estimate of κ_p as the solution to the following regularized least

squares cost function:

$$\hat{\kappa}_p = \arg \min_{\kappa_p} \|y - C(\kappa_p)\kappa_p\|_2^2 + \lambda^2 \|D_z \kappa_p\|_p^p. \quad (14)$$

The first term in (14) ensures that the estimate is consistent with the data. The second term plays the role of a regularizer and is used to help combat the ill-posedness. By varying p in the range 1 to 2 one can control the smoothness in the reconstruction. For $p = 2$, one obtains a traditional smoothness regularizer. As p approaches 1, the regularizer is more encouraging of forming profiles which have edges or other sharp discontinuities as might be found in certain NDE applications. With $p = 1$, one has a total variation (TV) regularization scheme [7, 19].

The regularization parameter, λ , in (14) determines the relative importance of the two terms on $\hat{\kappa}_p$. As $\lambda \rightarrow 0$, we demand that $\hat{\kappa}_p$ just fit the data. As mentioned in § 1, the resulting estimate tends to display high frequency, large amplitude artifacts. On the other hand, as $\lambda \rightarrow \infty$, the data play a limited role in influencing $\hat{\kappa}_p$ and we obtain overly smooth estimates. Proper selection of this parameter is a non-trivial problem [1, 6]. In this paper, we set λ by trial and error.

To solve (14) we start by writing $\|D_z \kappa_p\|_p^p = g(\kappa_p)^T g(\kappa_p)$ with g the vector whose i th element is $[D_z \kappa_p]_i^{p/2}$. Thus (14) is formulated as a non-linear least squares optimization problem:

$$\hat{\kappa}_p = \arg \min_{\kappa_p} e^T(\kappa_p) e(\kappa_p) \quad (15)$$

$$e(\kappa_p) = \begin{bmatrix} y - C(\kappa_p)\kappa_p \\ \lambda g(\kappa_p) \end{bmatrix} \quad (16)$$

Eq. (15) is solved using the `leastsq` routine provided in Matlab's Optimization Toolbox. For a given κ_p this program requires the ability to evaluate the RHS in (15) and a method for computing the Jacobian of the vector e in (16) with respect to each element of κ_p . The cost evaluation is done using the equations developed in the previous section. We approximate the Jacobian, J , as

$$J = \begin{bmatrix} -C(\kappa_p) \\ \frac{\partial g}{\partial \kappa_p} \end{bmatrix}. \quad (17)$$

After some straightforward calculus, it is not hard to show that

$$\frac{\partial g}{\partial \kappa_p} = \frac{p}{2} \mathcal{D}(g)^{-p/2} \mathcal{D}(\text{sign}(g)) D_z$$

with $\text{sign}(g)$ the vector whose i th element is the signum of g_i . Note that in (17) we ignore the dependence of C on κ_p . This is the primary motivation for formulating the physical problem in the manner of §2. This approximation is closely related to the Born Iterative Method [2] for problems in which the ρc product in (1) is to be recovered rather than κ .

3.2 Adaptive Multiscale Inversion

As discussed in §1, a key issue of interest in this paper is the determination of a low order, multiscale representation for the vector κ_p as a means of reducing the complexity of the inversion process and improving the quality of the ultimate reconstruction by “appropriately” distributing fine scale information. To achieve this, we constrain κ_p to live in the linear span of a set of vectors where we adaptively determine both the vectors and the required expansion coefficients. Formally, we consider representation of κ_p of the form

$$\kappa_p = \sum_{i=1}^{N_b} b_i a_i = B a \quad (18)$$

where b_i is the i th vector, a_i the expansion coefficient, B the matrix whose i th column is b_i , and a the vector of a_i . Note that a is of length N_b .

Using (18) and (15), the determination of $\hat{\kappa}_p$ now reduces to the estimation of the a_i via

$$\hat{a} = \arg \min_a e^T(Ba) e(Ba) = \arg \min_a \|e(Ba)\|_2^2 \quad (19)$$

where $e(Ba)$ is (16) evaluated at the vector $\kappa_p = Ba$. Letting N be the length of κ_p , then typically we have $N_b \ll N$ so that (19) is a far smaller optimization problem than (15). In the remainder of this section, we describe the family of vectors used in our approach as well as the methods we have developed to adaptively determine those b_i to include in $\hat{\kappa}_p$.

The b_i vectors are sampled versions of a family of continuous functions which satisfy the following

two scale dilation equation [17]

$$b(z) = \sum_{k=0}^{N_k-1} \alpha_k b(2z - k). \quad (20)$$

In words, (20) says that the function b can be expressed as a linear combination of dilated and translated versions, $b(2z-k)$, of itself. While specification of the coefficients α_i can lead to many functions which satisfy (20), including wavelets, here we are concerned with B-splines [17, Chap. 7.4]. For example, with $\alpha_0 = \alpha_1 = 1$, the $b(z)$ satisfying (20) is nothing more than the box function from $z = 0$ to $z = 1$. Such zeroth order splines form a basis for piecewise constant functions. With $\alpha_0 = \alpha_2 = 1$ and $\alpha_1 = 2$, we obtain a hat function, or first order B-spline, which is piecewise linear.

Here we employ piecewise quadratic B-splines generated using $\alpha_0 = \alpha_3 = 1$ and $\alpha_1 = \alpha_2 = 3$. In Figure 1, we illustrate these functions on two spatial scales: a coarse scale at the top and a finer scale at the bottom. Note that the relative spacing of the spline functions is proportional to their width. Additionally, because these functions satisfy (20), wide functions on one scale can be expressed as linear combinations of narrow functions on the next scale. For example, the highlighted element shown at the top of Figure 1 can be decomposed via (20) in terms of the fine scale functions in the bottom panel. We represent the relationship among these functions using a graph as shown in Figure 2. The nodes in the graph represent the functions (or associated expansion coefficients) and the edges indicate the links from one scale to the next.

We make use of the non-linear least squares formulation of § 3.1 and the functions defined above to obtain a reconstruction of κ_p using the algorithm outlined in Fig. 3. We start with a low order, coarse scale collection of quadratic B-splines. With these b_i , we estimate a set of expansion coefficients to get a rough $\hat{\kappa}_p$. The remainder of the algorithm is a loop where the current collection of vectors is alternately refined and then unneeded detail is pruned away. The goal is to provide sufficient flexibility to add arbitrary detail to $\hat{\kappa}_p$ followed by a stage in which we determine which, if any, degrees of freedom associated with this detail were warranted. Those not needed are removed.

While one might consider many methods to accomplish this objective, in the following paragraphs, we detail an approach which we have found to work well at least for the DP problem.

In the refinement stage, we replace all b_i in our representation of $\hat{\kappa}_p$ by their finer scale children². This allows for the recovery of finer scale information at the expense of a higher order representation. Once the new collection of vectors is constructed, we solve for a new vector of expansion coefficients to produce a finer scale estimate of κ_p . While this requires the solution of a new non-linear least squares problem, the coarse scale estimate is used to initialize the finer scale reconstruction procedure to speed convergence. Specifically, the a_i for each child is initialized using the average of the previously generated parent estimates. From Figure 2, assume that vectors 3-8 had just been replaced by 9-22. Then, e.g., the initial values for a_{13} and a_{14} would both be $\frac{a_4+a_5}{2}$.

After refining the estimate, we next remove unnecessary detail. The goal here is to reduce the complexity of the model by

- **Coarsening** Replacing fine scale children with their parents in regions where such detail is not warranted
- **Pruning** Removing from the representation vectors which contribute little to $\hat{\kappa}_p$

Coarsening is done in a sequence of steps in which limited detail is removed at each stage until we can no longer represent the current estimate of κ_p to within some predetermined tolerance. A similar procedure is carried out for pruning except that we remove from the estimate in a one-by-one manner vectors whose $|a_i|$ are small. The result is a lower order model which is a “small” perturbation of the optimal estimate computed from the previous optimization step.

As illustrated in Fig. 4, we begin coarsening by looking for all collections vectors in our current B matrix which represent a complete set of children for a given parent. For example, in Figure 2, vectors 13-16 are a complete set of children for 5. For each such children-parent set, we consider

²Obviously, those b_i already at the finest scale are not refined.

a new collection of vectors obtained by replacing those fine scale children with *all* of their coarse scale parents keeping the remaining members of the current collection the same. Again, referring to Figure 2 suppose that the current B matrix was comprised of vectors 9-22. One set of vectors considered by the pruning process would be 9-12, 4-6, and 17-22. Here 13-16 are removed and replaced by all relevant parents, in this case 4, 5, and 6. Another possible configuration is 9, 10, 3-5, and 15-22 where 11-14 are replaced by 3-5.

This replacement strategy yields some finite number of possible sets of vectors for representing κ_p with each set having fewer members than the current one. To determine which, if any collection is selected, we project $\hat{\kappa}_p$ onto the linear span of the vectors in each set. Let B_i be the matrix constructed from one such coarse set of vectors. The projection, $\hat{\kappa}_{p,i}$, is computed as $\hat{\kappa}_{p,i} = B_i B_i^\dagger \hat{\kappa}_p$ where B_i^\dagger is the pseudo-inverse of B_i [3, p. 243]³. As explained in Appendix 6, to evaluate the utility of $\hat{\kappa}_{p,i}$ it is useful to compute a weighted error, $\|J(\hat{\kappa}_p - \hat{\kappa}_{p,i})\|$ where J is given in (17) and is evaluated at $\hat{\kappa}_p$. If the minimum weighted error over all i is less than some threshold, we replace the fine scale collection with the elements from the new, coarsened set, otherwise we stop. If we do not stop, then the pruning process is repeated. Specifically, if B_j^* denotes the selected set, we replace children in B_j^* by parents again using the relative norm criteria just described. If the minimum error still is not too small, B_j^* is replaced by a new, smaller set and the process continues.

To do pruning, we start by removing from the final set of vectors produced by coarsening that vector whose $|a_i|$ is smallest. As is done in pruning, we look at the weighted error between $\hat{\kappa}_p$ and its projection onto the remaining vectors. If this error exceeds a threshold, then we stop, otherwise, the procedure continues using the expansion coefficients associated with this projection.

After pruning is finished, we pass the new collection of expansion vectors into the non-linear

³Because we allow B_i to contain B-spline vectors at multiple scales and arbitrary dyadic shifts, it is possible that elements of B_j will be linearly dependent hence the need to use a pseudo-inverse. Both theoretically and practically this linear dependence causes no difficulty for the algorithm in this paper.

least squares solver to obtain the best estimate of the conductivity using this coarsened set. Again, another optimization is performed. As described in the previous paragraphs however, the pruning process yields a low order collection of vector which still supports an accurate approximation of $\hat{\kappa}_p$. Thus, it is reasonable to assume that the set of optimal expansion coefficients should be close to the coefficients obtained by projecting $\hat{\kappa}_p$ onto the coarse set. Throughout our numerical experiments, we have found that initializing in this way does in fact lead to rapid convergence of the non-linear least squares routine for this stage of the algorithm.

This refinement-pruning loop is continued until the relative difference of the two-norm between estimates produced at the end of two successive pruning stages is below some threshold. We note that this approach does not guarantee a monotonic decrease in the cost function as we change the vectors in B . Indeed, we have seen in our experiments that the optimal κ_p obtained after a coarsening/pruning step can have a higher cost than the previous $\hat{\kappa}_p$ comprised of a larger number of expansion vectors. As examined in Appendix 6 however, the choice of weighted error criteria for accepting a coarsened or pruned collection of vectors is constructed in a way to limit the resulting rise in cost. Thus, while the cost may increase, we guarantee that this increase is bounded.

4 Examples

Here we consider the performance of our inversion approach on a variety of profiles. The background is meant to model case hardened steel as in [8] so $\kappa_b = 0.45 \text{ W cm}^{-1} \text{ K}^{-1}$, $\rho = 7.7 \text{ g cm}^{-3}$, and $c = 0.48 \text{ J g}^{-1} \text{ K}^{-1}$. In all cases depth profiles are constructed for z between 0 and 400 μm from the surface based on thermal wave data from 16 logarithmically spaced frequencies in the range 10 to 10000 s^{-1} . The signal to noise ratio on T_s is taken to be 30 dB.

Over the 400 μm range of interest, we discretize the forward model on a grid of 256 equally spaced points corresponding to an interpoint spacing of about 1.57 μm . All data vectors are

generated using the exact, discretized forward model while the inversion algorithm makes use of the EBA-type model described at the end of §2.2. For all examples, ϵ , the tolerance parameters used in the coarsening and pruning stages was $1e - 4$. The threshold used to terminate the refinement-pruning loop was 5%. In all cases, $\lambda = 1$ (this value was determined by trial and error) and the reconstruction algorithms were initialized with zero for the perturbation. Finally, there were a total of five scales worth of B-spline functions that could be used in the construction of $\hat{\kappa}_p$.

A Monte-Carlo analysis was used to evaluate the Extended Born Approximation to the true scattering model. We started by generating 100 random κ_p profiles via

$$\kappa_p = \sum_{i=1}^{17} a_i b_i$$

where the b_i are a collection “mid-scale” B-splines shown in Figure 5(a) and the a_i were taken to be zero mean, independent Gaussian random variables with standard deviation 20^4 . The goal here was to produce profiles which were not merely small perturbations on the background. Some samples are shown in Figure 5(b). For each profile, we solved the forward problem using the exact method and the EBA. Over the 100 trials, the average relative error between the exact and approximate data vectors was about 10^{-13} while the EBA required about 18 times fewer floating point operations than the exact model. Thus, for the problems of interest here, the EBA represents a highly efficient, essentially exact solver for the HDE.

We start by examining the performance of our approach for a profile with a layer-like inclusion shown as the solid line in Figure 6(a). To gauge the average performance in the presence of noise, in this and the next example, we performed the inversion ten times for ten difference realizations of the additive disturbance. In Figure 6(a) the dashed line is a plot of the average reconstruction for the adaptive method while the results of inverting using the finest scale grid of 256 points are given by the dashed-dot line. Here we set $p = 1.1$ in (14) to obtain an edge-preserving type of regularization.

⁴Too keep κ positive, we set to zero any values of $\kappa_p < 45$, where 45 is the background thermal conductivity.

In practice it would be necessary to also develop a method for choosing this quantity; however for the sake of simplicity, as with λ , we assume that p is known *a priori*.

Figure 6(a) shows that, on average, the results of using the adaptive approach are comparable to those of a fixed, fine scale inversion. For the adaptive method, the peak value of the reconstruction is slightly closer to that of the true profile. However, for the fixed scale method, the average reconstruction is slightly flatter over the region of the layer. The average mean square error between the two methods differed by about 1% with the fine scale case having smaller average error.

As stated in the Introduction, one of the potential advantages of the adaptive method over a fixed scale inversion is computational. To explore this issue, we “instrumented” the `leastsq` procedure to collect statistics of the inversion process. For one of the ten runs in Figure 6(b) and (c), we plot the value of the cost function and the value of the error against the number of times the inversion routine made a call to evaluate the cost. Since cost evaluation requires the evaluation of the HDE model which is the most computationally intensive step in the process, the number of times the cost function is called is a valid measure of the complexity of the inversion procedure. Also, for the adaptive approach, the plot displays the cost and error for *all* refinement-pruning stages of the algorithm and thus is an accurate portrayal of the complexity of the entire routine.

We observe from Figure 6(b) and (c) that the adaptive method converges at a faster rate than the fine scale method. Thus, while the two methods converge essentially to the same profile, by constraining the inversion, we get closer to the final result faster. We note that this behavior was true in general for all ten runs. Moreover, because the adaptive approach required fewer unknowns, for this problem, significantly fewer floating point operations were needed. The average number of unknowns in the final reconstruction over the ten runs was about 19 for the adaptive approach versus 256 for the fine scale method. Reduction in the size of the inverse problem translated into a floating point count that was about 2.86 times smaller for the adaptive method than the fixed scale

inversion. Thus, for this problem, the adaptive method produces essentially the same reconstruction as a fixed scale technique at a significantly reduced level of complexity.

In addition to this quantitative advantage, the adaptive inversion also yields interesting qualitative results. By examining *which* vectors are used in the reconstruction, we see that the adaptive approach placed limited detail precisely where it belongs in a reconstruction using fewer, coarser scale vectors where κ_p varies more slowly. Indeed, Figure 6(a) indicates that for both fixed and adaptive inversions, the edge closer to the surface is recovered more accurately than the edge deeper into the medium. For the same run used to produce the graphs in Figure 6, we show in Figure 7(a) the individual vectors which contribute to the final adaptive reconstruction. We see in particular that most of the fine scale detail is distributed near the leading edge with wider, coarser scale vectors used for much of the remainder of the reconstruction. This point is further highlighted in Figure 7(b) where we plot as a function of depth, the finest scale in the reconstruction. Specifically, at each point, z , there may be a number of vectors whose support includes this point. Figure 7(b) then shows the scale of the finest vector having support at each position. Larger numbers imply finer scale and for these examples, there are 5 scales of B-splines which may be used.

Typically, for a diffusive inverse problem of the type considered here, physical intuition dictates that the level of detail decreases as a function of increasing depth due to the heavy damping of the thermal waves. While this may be true in general, it also ignores the fact that one may want limited fine scale detail to support the reconstruction of important sharp features (e.g. edges) which may be present in the profile. The results in Figure 7 demonstrate that the algorithm described in this paper is capable of adapting the detail in the reconstruction in a way which reflects both the underlying physics as well as the structure of the particular profile. The general trend is a decrease in resolution as a function of depth; however, we see an increase in the level of detail specifically in the neighborhood of the edge. Thus, to obtain the shaper profile, the algorithm automatically

chooses to distribute some fine scale degrees of freedom near the feature of interest.

In Figs. 8 and 9, we examine the results of the method represented in this paper for a smoother profile. For this analysis, we set $p = 2$ in (14) to obtain a more standard smoothness type of regularizer. The results here indicate a different advantage the adaptive method may have with respect to a fixed, fine scale inversion. Figure 8 shows that over the ten runs used to evaluate the performance of the algorithm, on average the fixed scale approach converges to a far from global minimum of the cost function. Alternatively, by constraining the search space, the adaptive inversion routine essentially settles into a lower cost local minimum which is also of much lower error. The convergence curves for one of the runs shown in Figure 8(b) and (c) provide further evidence of this behavior. For this problem, despite the fact that the adaptive approach uses on average about 11 vectors, approximately 16% more floating point operations are required versus the finest scale inversion. Thus, in contrast to the first example, here we see a slightly higher computational cost producing a significantly improved reconstruction.

Finally, in Figure 9, the constituent components of the one of the ten final reconstructions are displayed. We see here that the absence of an edge or other sharp feature in κ_p has resulted in coarser scale vector being used here relative to those in Figure 7. Generally, the trend is one of decreasing resolution as a function of depth.

5 Conclusions and Future Work

We have described an inverse scattering type of approach to the photothermal depth profilometry problem. Using the heat diffusion equation as the basic physical model, a measurement model is derived based on a decomposition of the thermal conductivity into the sum of a background component and an arbitrary sized perturbation. The physical model is cast in a form appropriate for use in a regularized, iterative inversion scheme. We have presented an accurate, low complexity

approximation to the exact HDE model having the form of an Extended Born Approximation.

We have introduced and explored the performance of an inversion technique designed to focus the information in the data on a small number of carefully selected degrees of freedom. Specifically, we have made use of multiscale, B-spline functions to construct an adaptive algorithm in which the presence of fine scale detail in the reconstruction is carefully controlled. The algorithm is based on an iterative loop in which the current set of vectors is first refined, to allow more detail, and subsequently pruned, to remove unwarranted degrees of freedom. The utility of this approach has been demonstrated on a number of simulated examples representative of applications of this technology to the non-destructive evaluation of case hardened steel. We plan on testing this method on experimental data in the near future.

In comparing the adaptive method to a fixed, fine scale inversion we have seen two types of behavior. For one problem the two approaches produce essentially the same reconstruction but the computational cost of the adaptive method is far smaller than that of the fixed scale technique. Alternatively, for a second problem, the adaptive method was shown to converge to a far lower cost of the optimization function; however the price paid for this improved performance was about a 20% increase in complexity. One important objective in the future is to develop a theoretical understanding of the circumstances (classes of profiles, noise conditions, sensing system parameters, etc.) under which one might expect one of these two types of performance.

6 Appendix 1

Because the basis coarsening and pruning steps remove degrees of freedom from the representation of $\hat{\kappa}_p$, it is possible value of the cost function will rise after these procedures. The objective of this appendix is to derive a method for accepting a coarsened set of vectors which guarantees that the relative rise in cost is no larger than some predefined threshold.

Referring to (15) we define $F(\kappa_p)$ as $e^T(\kappa_p)e(\kappa_p)$. Let us assume that we have a fine scale estimate, $\hat{\kappa}_p$ which we wish to coarsen. Because $\hat{\kappa}_p$ is by definition a solution to the non-linear least squares optimization problem for some B , a second order Taylor expansion of F about $\hat{\kappa}_p$ yields []

$$F(\hat{\kappa}_p + \delta) \approx F(\hat{\kappa}_p) + 1/2\delta^T H \delta \quad (21)$$

where H is the Hessian of the cost function. Since δ is small, the Hessian is well approximated by $J^T J$ [] with J the Jacobian of e evaluated at $\hat{\kappa}_p$. Thus, after some algebra we write the relative change in cost due to a perturbation δ from $\hat{\kappa}_p$ as

$$\frac{F(\hat{\kappa}_p + \delta) - F(\hat{\kappa}_p)}{F(\hat{\kappa}_p)} \approx \frac{\|J\delta\|_2^2}{2F(\hat{\kappa}_p)}. \quad (22)$$

From § 3.2, for the coarsening problem, $\delta = \hat{\kappa}_p - \hat{\kappa}_{p,i}$. Thus, to ensure that the relative change in the cost rises by at most ϵ , we require that $\|J(\hat{\kappa}_p - \hat{\kappa}_{p,i})\|_2^2 < 2\epsilon F(\hat{\kappa}_p)$.

7 Acknowledgments

The work of the first and second authors was supported by a CAREER Award from the National Science Foundation MIP-9623721, an ODDR&E MURI under Air Force Office of Scientific Research contract F49620-96-1-0028, and the Army Research Office Demining MURI under Grant DAAG55-97-1-0013. The work of the other authors was supported by a research contract from Material and Manufacturing Ontario (MMO). The first author would like to thank Dr. Oliver Dorn of Northeastern University and Prof. Misha Kilmer of Tufts University for some very helpful discussions which greatly improved the material in this manuscript.

References

- [1] Murat Belge, Eric Miller, and Misha Kilmer. Simultaneous multiple regularization parameter selection by means of the l-hypersurface with applications to linear inverse problems posed in the wavelet transform domain. In *SPIE International Symposium on Optical Science, Engineering, and Instrumentation: Bayesian Inference for Inverse Problems*. SPIE, July 1998.
- [2] Weng Cho Chew. *Waves and Fields in Inhomogeneous Media*. Van Nostrand Reinhold, New York, 1990.
- [3] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins Univ. Press,

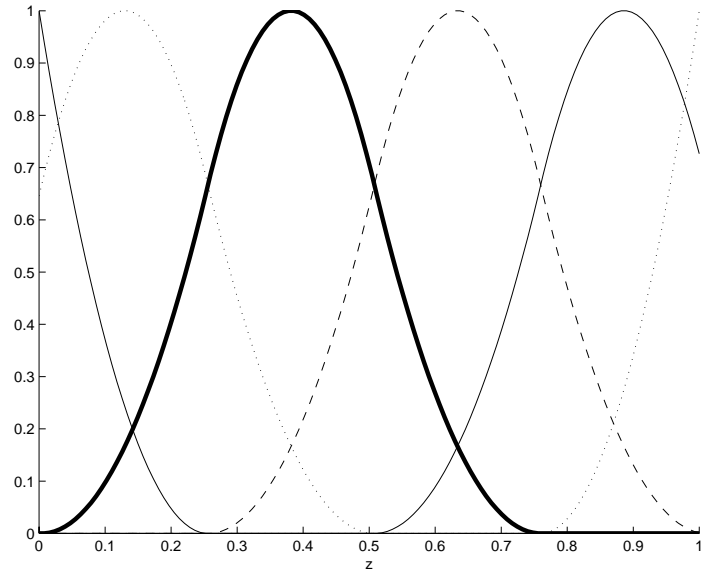
2nd edition, 1989.

- [4] P. Grossel and F. Depasse. Alternating heat diffusion in thermophysical depth profile: multi-layer and continuous descriptions. *J. Phys. D: Appl. Phys.*, 31:216–223, 1998.
- [5] Tarek M. Habashy, Ross W. Groom, and Brian R. Spies. Beyond the Born and Rytov approximations: A nonlinear approach to electromagnetic scattering. *J. Geophys. Res.*, 98(B2):1759–1775, February 1993.
- [6] Per Christian Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580, December 1992.
- [7] S. Osher L. I. Rudin and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60:259–268, 1992.
- [8] T. T. N. Lan, U. Seidel, and H. G. Walther. Theory of microstructure depth profiling by photothermal measurements. *J. Appl. Phys.*, 77(9):4739–4745, 1 May 1995.
- [9] Roberto LiVoti, Mario Bertolotti, and Concita Sibilìa. Thermal conductivity and diffusivity depth profiles by photothermal technique: The direct and inverse problem. In Xavier Maldague, editor, *III International Workshop -Advances in Signal Processing for NDE of Materials*, volume 3 of *Topics in Nondestructive Evaluation Series*, pages 379–386. The American Society for Nondestructive Testing, Inc., 1998.
- [10] Andreas Mandelis, Frank Funak, and Mahendra Munidasa. Generalized methodology for thermal diffusivity depth profile reconstruction in semi-infinite and finitely thick inhomogeneous solids. *J. Appl. Phys.*, 80(10):5570–5578, 15 Nov. 1996.
- [11] Andreas Mandelis. Hamilton-Jacobi formulation and quantum theory of thermal wave propagation in the solid state. *J. Appl. Phys.*, 26(10):2676–2683, October 1985.
- [12] Andreas Mandelis and Samuel B. peralta. Photoacoustic frequency domain depth profiling of continuously inhomogeneous condensed phases: Theory and simulations for the inverse problem. *J. Appl. Phys.*, 70(3):1761–1770, 1 August 1991.
- [13] Andreas Mandelis, Els Schoubbs, Samuel B. Peralta, and Jan Thoen. Quantitative photoacoustic depth profilometry of magnetic field-induced thermal diffusivity inhomogeneity in the liquid crystal octylcyanobiphenyl. *J. Appl. Phys.*, 70(3):1771–1777, 1 August 1991.
- [14] Eric L. Miller, Lena Nicolaides, and Andreas Mandelis. Nonlinear inverse scattering methods for thermal wave slice tomography: A wavelet domain approach. *Journal of the Optical Society of America (A)*, 15(6):1545–1556, June 1998.
- [15] Eric L. Miller and Alan S. Willsky. Wavelet-based methods for the nonlinear inverse scattering problem using the Extended Born Approximation. *Radio Sci.*, 31(1):51–67, Jan–Feb 1996.
- [16] M. Moghaddam, W.C. Chew, and M. Oristaglio. Comparison of the Born iterative method and Tarantola’s method for an electromagnetic time-domain inverse problem. *International Journal of Imaging Systems and Technology*, 3:318–333, 1991.
- [17] Gilbert Strang and Truong Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, 1996.

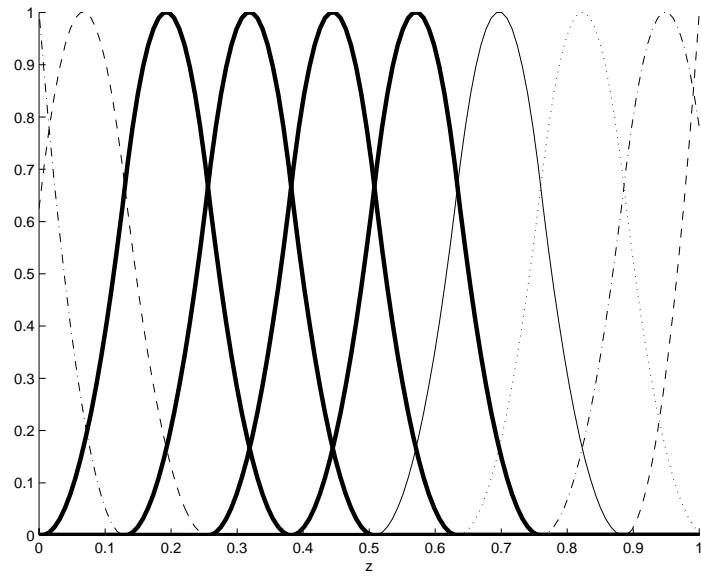
- [18] Carlos Torres-Verdín and Tarek M. Habashy. Rapid 2.5-D forward modeling and inversion via a new nonlinear scattering approximation. *Radio Sci.*, 29(4):1051–1079, July-August 1994.
- [19] C. R. Vogel and M. E. Oman. Fast, robust total variation-based reconstruction of noisy, blurred images. *IEEE Trans. Image Process.*, 7(7):813–824, July 1998.

8 Figure Captions

- **Fig. 1:** Examples of quadratic B-spline functions on two spatial scales
- **Fig. 2:** Graphical representation of multiscale B-spline vectors. Nodes on one horizontal level represent vectors or their associated expansion coefficients at a given scale. Links from one node to the next indicate cross-scale dependencies induced by the underlying two scale dilation equation, (20). Coarse scales are at the top with finer scales at the bottom.
- **Fig. 5:** In (a) we plot the B-spline vectors used to generate profiles for the testing of the EBA. A few samples of the resulting profiles are shown in (b)
- **Fig. 6:** Reconstruction performance and convergence behavior for adaptive method and fixed, finest scale approach for a layer-like profile. In (a), the average reconstructions taken over 10 noise realizations at 30 dB is compared to the true profile. For one of these runs, (b) and (c) demonstrate the improved convergence behavior typically exhibited by the adaptive method.
- **Fig. 7:** Reconstruction (a) and finest scale distribution (b) for the adaptive inversion method for layer-like profile.
- **Fig. 8:** Reconstruction performance and convergence behavior for adaptive method and fixed, finest scale approach for a smoothly varying profile. In (a), the average reconstruction taken over 10 noise realizations at 30 dB is compared to the true profile. For one of these runs, (b) and (c) demonstrate the convergence behavior of the two methods.
- **Fig. 9:** Reconstruction (a) and finest scale distribution (b) for the adaptive inversion method for smooth profile.



(a) Coarse scale vectors



(b) Fine scale vectors

Figure 1: Examples of quadratic B-spline vectors on two spatial scales

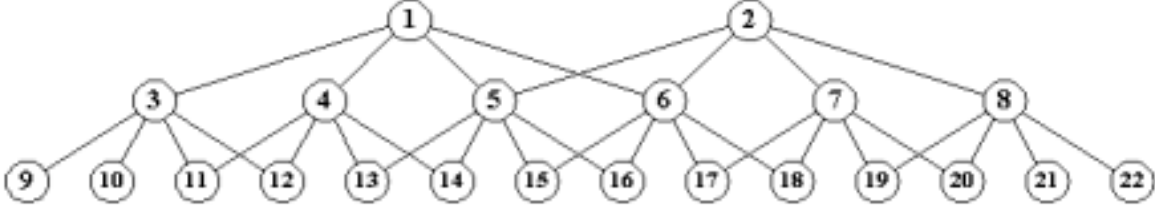


Figure 2: Graphical representation of multiscale B-spline vectors. Nodes on one horizontal level represent vectors or their associated expansion coefficients at a given scale. Links from one node to the next indicate cross-scale dependencies induced by the underlying two scale dilation equation, (20). Coarse scales are at the top with finer scales at the bottom.

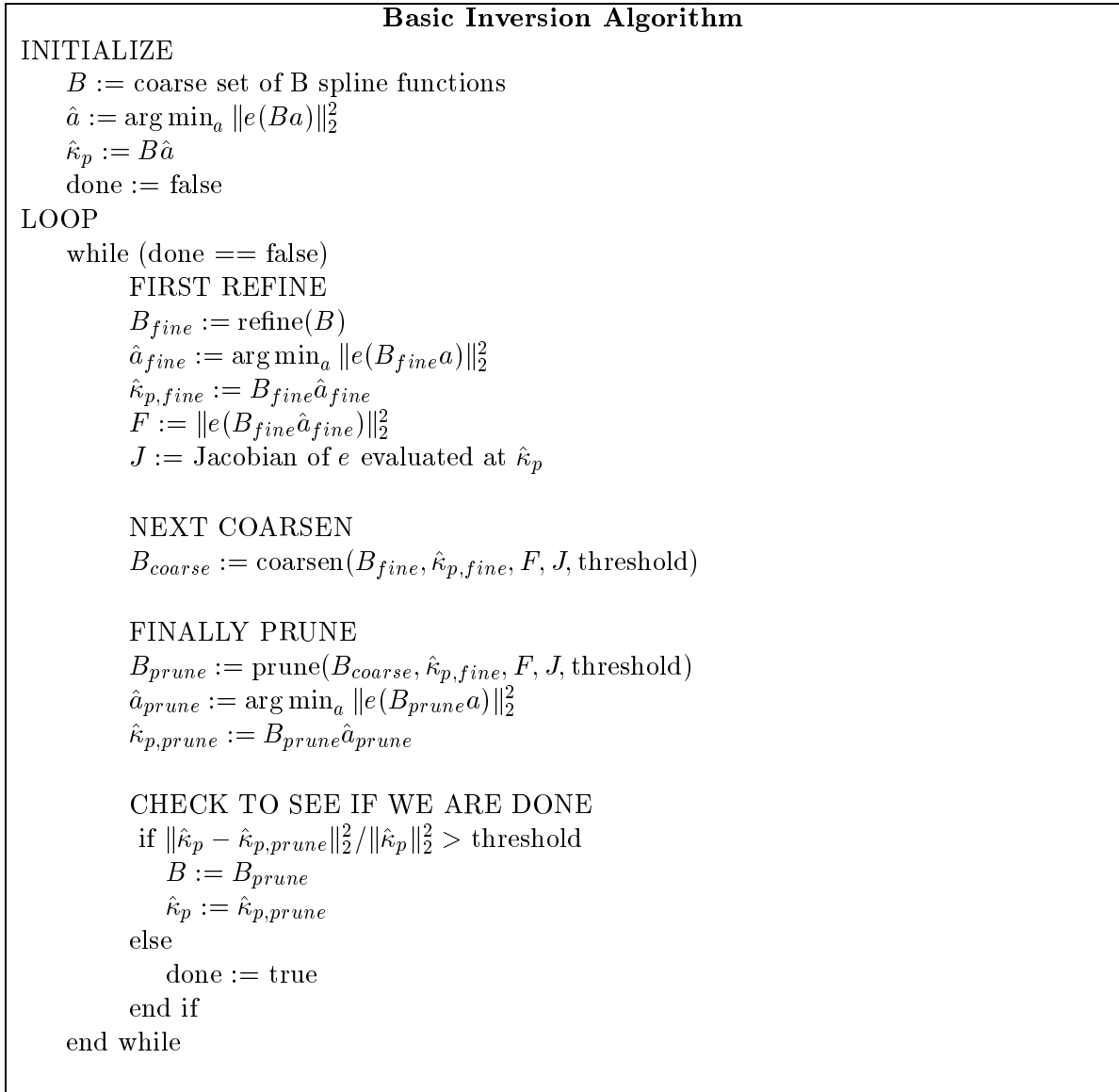


Figure 3: Pseudo-code for Inversion Algorithm

Coarsening Algorithm
<pre> $B := \text{prune}(B_{in}, \hat{\kappa}_p, F, J, \text{threshold})$ INITIALIZE $B := B_{in}$ done := false LOOP while (done == false) for $i = 1, 2, \dots$ number of ways to generate coarsened versions of B $B_i := i\text{th valid coarsened version of } B$ $\hat{\kappa}_{p,i} := B_i B_i^\dagger \hat{\kappa}_p$ $\text{error}_i := \ J(\hat{\kappa}_{p,i} - \hat{\kappa}_p)\ _2^2$ if $\min_i \text{error}_i < 2F \text{ threshold}$ $i^* := \arg \min_i \text{error}_i$ $B := B_{i^*}$ else done := true end if end for end while </pre>

Figure 4: Pseudo-code for Coarsening Algorithm

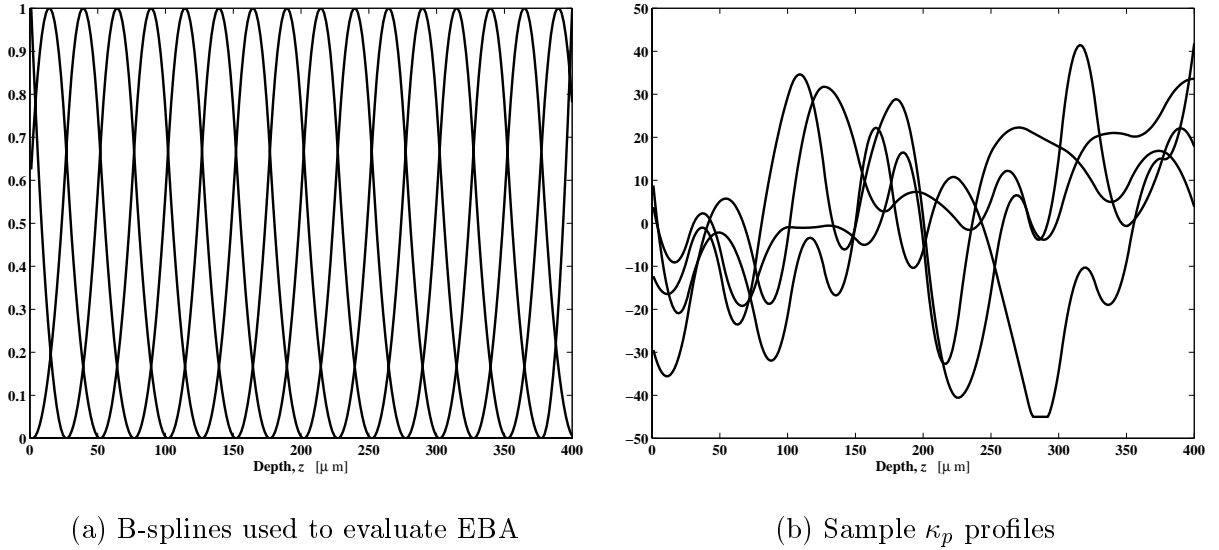
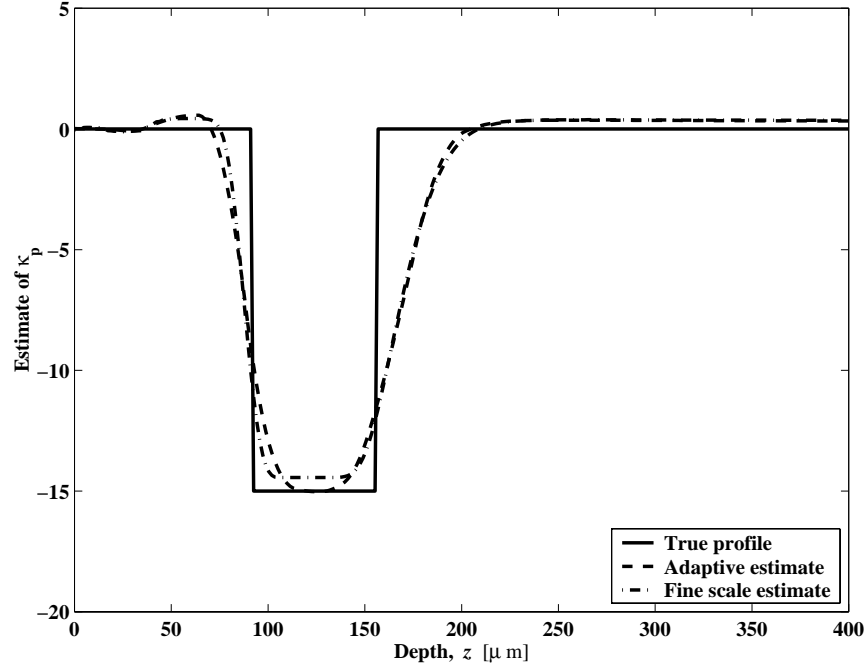
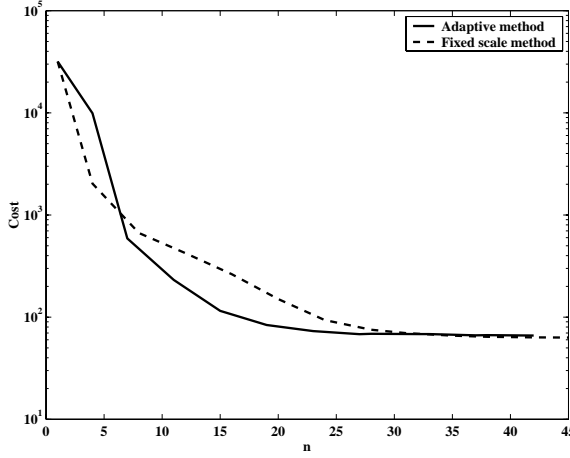


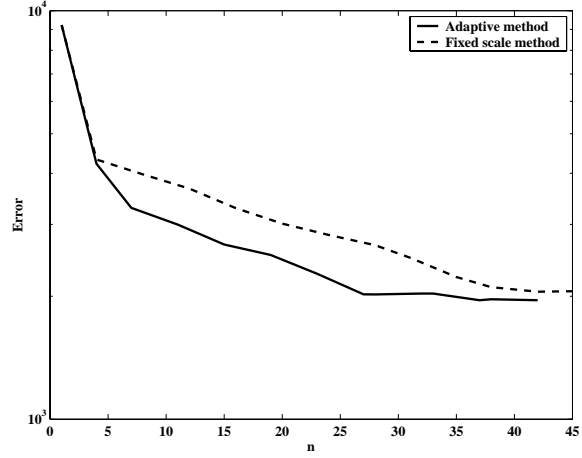
Figure 5: In (a) we plot the B-spline vectors used to generate profiles for the testing of the EBA. A few samples of the resulting profiles are shown in (b)



(a) Average reconstruction

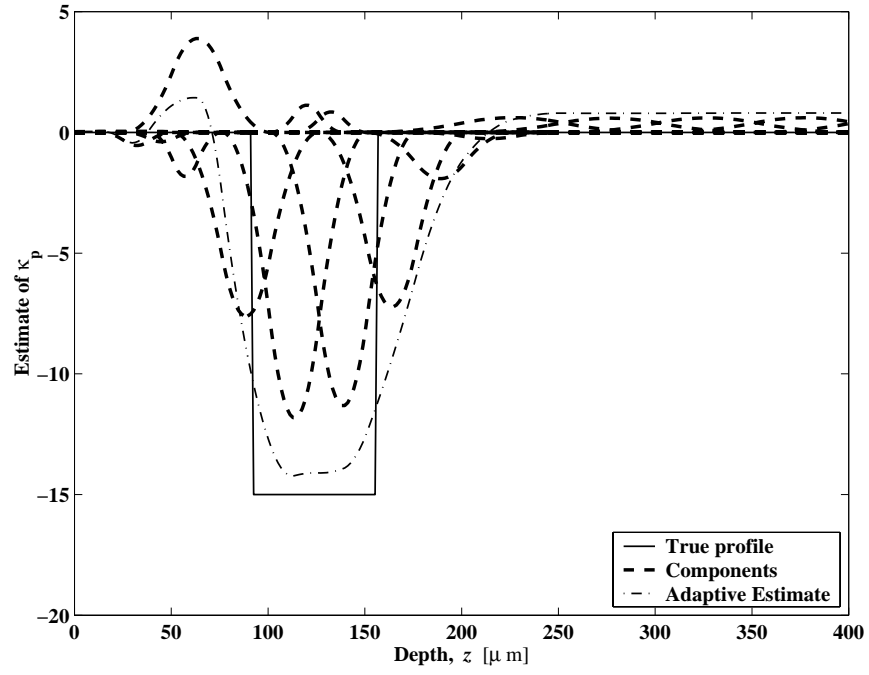


(b) Value of cost vector vs. iteration for one run

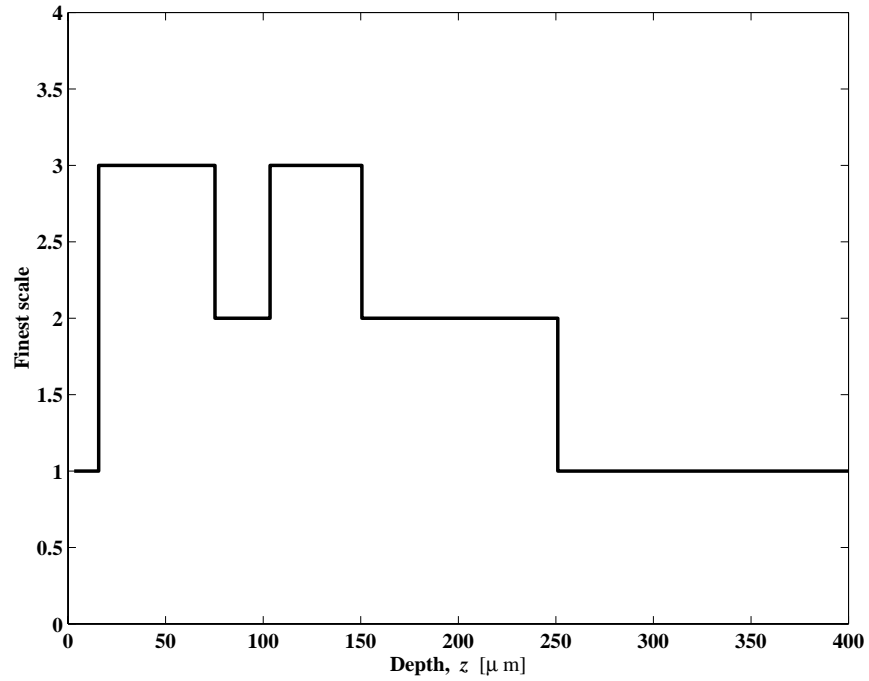


(c) Error vs. iteration for one run

Figure 6: Reconstruction performance and convergence behavior for adaptive method and fixed, finest scale approach for a layer-like profile. In (a), the average reconstructions taken over 10 noise realizations at 30 dB is compared to the true profile. For one of these runs, (b) and (c) demonstrate the improved convergence behavior typically exhibited by the adaptive method.

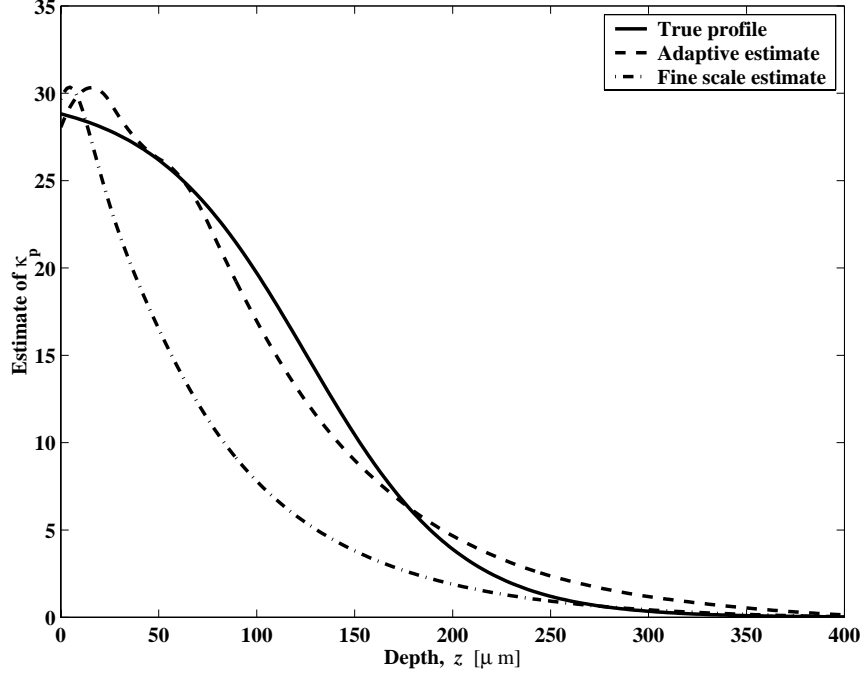


(a) Final estimate and constituent components

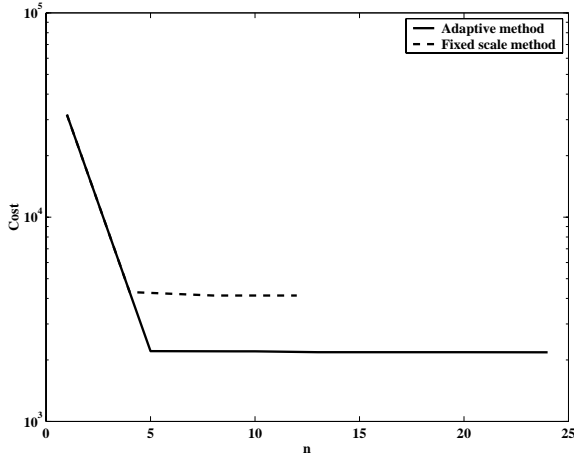


(b) Finest scale distribution

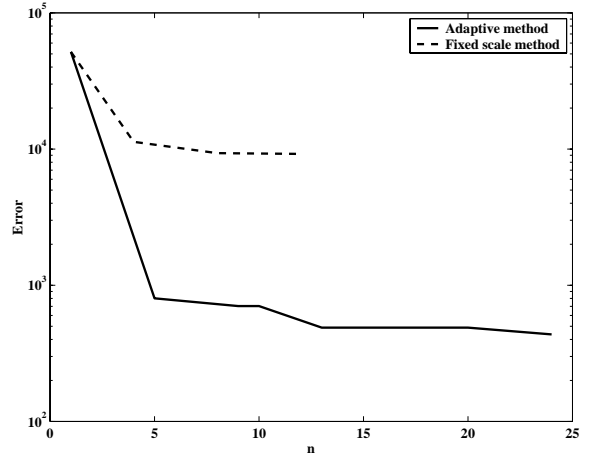
Figure 7: Reconstruction (a) and finest scale distribution (b) for the adaptive inversion method.



(a) Average reconstruction

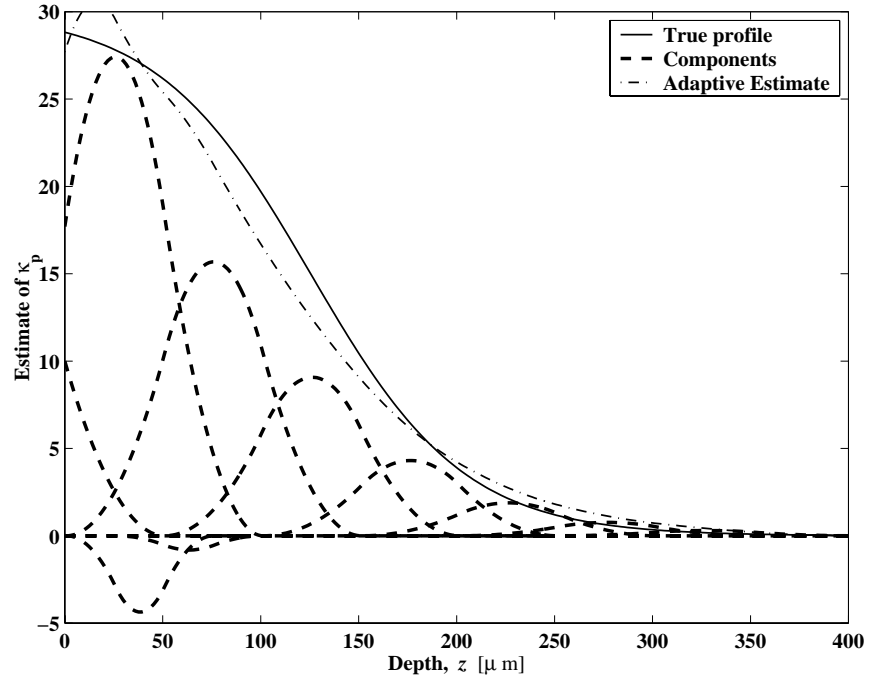


(b) Value of cost function vs. iteration for one run

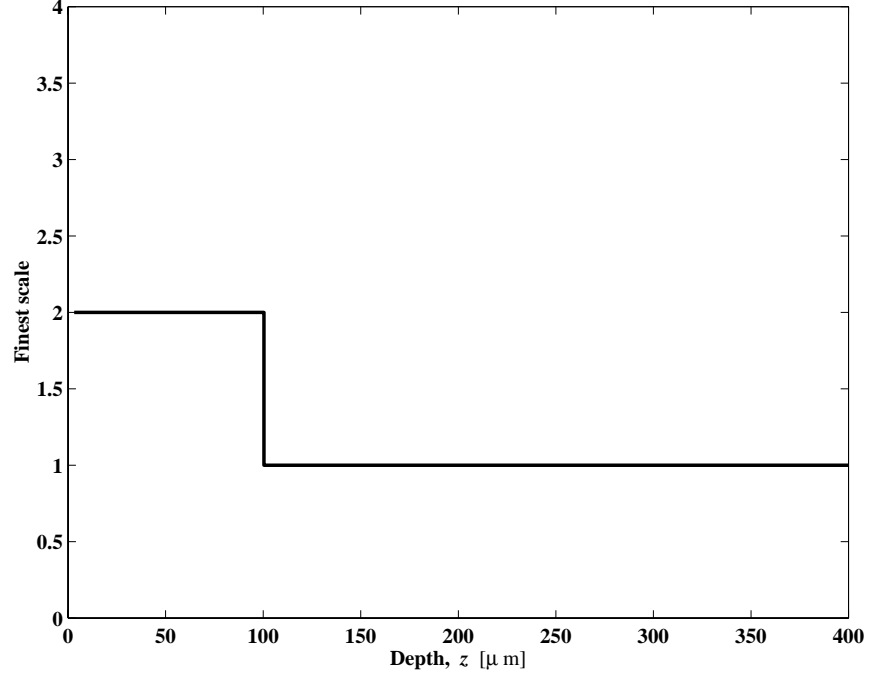


(c) Error vs. iteration for one run

Figure 8: Reconstruction performance and convergence behavior for adaptive method and fixed, finest scale approach for a smoothly varying profile. In (a), the average reconstruction taken over 10 noise realizations at 30 dB is compared to the true profile. For one of these runs, (b) and (c) demonstrate the convergence behavior of the two methods.



(a) Final estimate and constituent components



(b) Finest scale distribution

Figure 9: Reconstruction (a) and finest scale distribution (b) in adaptive inversion method.

A shape reconstruction method for electromagnetic tomography using adjoint fields and level sets

Oliver Dorn* Eric L. Miller* Carey M. Rappaport *

March 20, 2000

Abstract

A two-step shape reconstruction method for electromagnetic (EM) tomography is presented which uses adjoint fields and level sets. The inhomogeneous background permittivity distribution and the values of the permittivities in some penetrable obstacles are assumed to be known, and the number, sizes, shapes, and locations of these obstacles have to be reconstructed given noisy limited-view EM data. The main application we address in the paper is the imaging and monitoring of pollutant plumes in environmental cleanup sites based on cross-borehole EM data. The first step of the reconstruction scheme makes use of an inverse scattering solver which first recovers equivalent scattering sources for a number of experiments, and then calculates from these an approximation for the permittivity distribution in the medium. The second step uses this result as an initial guess for solving the shape reconstruction problem. A key point in this second step is the fusion of the 'level set technique' for representing the shapes of the reconstructed obstacles, and an 'adjoint field technique' for solving the nonlinear inverse problem. In each step, a forward and an adjoint Helmholtz problem are solved based on the permittivity distribution which corresponds to the latest best guess for the representing level set function. A correction for this level set function is then calculated *directly* by combining the results of these two runs. Numerical experiments are presented which show that the derived method is able to recover one or more objects with nontrivial shapes given noisy cross-borehole EM data.

*Center for Electromagnetics Research, Northeastern University, Boston, MA, E-mail: dorn@cer.neu.edu, elmiller@ece.neu.edu, rappaport@neu.edu

Contents

1	Introduction	3
2	The physical experiment	6
2.1	The Helmholtz Equation	6
2.2	Formulation of the inverse problem	7
3	The shape reconstruction problem	8
3.1	Shape reconstruction and inverse scattering	8
3.2	The domains $\Omega^{(n)}$	9
3.3	Level set representation of the domains $\Omega^{(n)}$	10
4	Step 2: Solving the shape reconstruction problem	11
4.1	Function spaces	12
4.2	Operators	12
4.3	Linearized operators	14
4.4	A nonlinear Kaczmarz-type approach	15
4.5	The minimization problem	16
4.6	The adjoint linearized operators	17
4.7	The operators $T'_{jk}T'^{*}_{jk}$	18
4.8	Regularization of $T'_{jk}T'^{*}_{jk}$	18
4.9	Updating the level set function	20
4.10	Implementation: The levelART algorithm	22
5	Step 1: A Source-Type Adjoint Field method	23
5.1	Source-type methods	24
5.2	Solving the inverse scattering problem	25
5.3	Looking for a scattering source	26
5.4	Recovery of the scattering potential	28
5.5	Combining the results from single experiments	29
5.6	The initial level set function	30
5.7	Implementation: The STAF algorithm	31
6	Numerical Experiments	31
6.1	Discretization of the computational domain.	31
6.2	A full-view example	33
6.3	A cross-borehole situation with multiple objects	34
6.4	A cross-borehole situation with a single inclusion and variable conductivity	36

7	Summary and future directions	37
8	Acknowledgments	39
9	Appendix	39
9.1	Proof of theorem 5.1	39
9.2	Proof of theorem 5.2	40
10	Figures	43

1 Introduction

In this paper, we investigate the retrieval of an unknown number of penetrable objects (inclusions) imbedded in an inhomogeneous background medium based on observations of electromagnetic (EM) fields. The electromagnetic characteristics (permittivity and conductivity) of the background medium as well as of the material forming the inclusions are assumed to be known, but the main topological information concerning the number, sizes, shapes, and locations of the inclusions is missing and has to be reconstructed from the EM data.

One possible technique for using EM fields in cross-borehole tomography is ElectroMagnetic Induction Tomography (EMIT) [3, 18, 44, 49, 50, 51] which typically operates at frequencies between 1 to 20 kHz. In this frequency band, electromagnetic fields tend to diffuse rather than propagate as waves through the Earth. Penetration depths of 100 m or more are possible at these low frequencies, but the diffusional behavior of the fields makes the inverse problem severely ill-posed.

However, if the typical distances in the area of interest are not much larger than 10-20 m, we can use EM fields in the higher frequency band of 5 to 30 MHz instead. The wavelengths of these fields are typically between 2-15 m in moist soil, where the relative dielectric constant is typically around 20 [48]. Therefore, we can make use of wave propagation phenomena in the inversion process. In the present paper, we address this situation. The main application we have in mind is the imaging and monitoring of pollutant plumes at environmental cleanup sites given cross-borehole EM data, where the distances of the boreholes are not much larger than 10-20 m.

We assume that the known conductivity distribution is positive but small everywhere, and that the permittivity distribution in the medium has to be recovered. Inside the pollutant plumes, the permittivity is assumed to be constant with a known value, and the background permittivity is arbitrary but also known. Therefore, the task is to find the *number, shapes, sizes and locations of the pollutant plumes* from cross-borehole data gathered for a small number (less than 10) of frequencies. No

topological constraints are made on the shapes of these plumes. For example, they are allowed to be multiply connected, and to enclose 'cavities' or 'holes' filled with background material.

The main difficulties which arise in this situation are 1.) We want to allow for an (arbitrary) inhomogeneous background permittivity distribution in the inversion; 2.) The inverse problem is usually strongly nonlinear because of the high contrast of the permittivity values inside the plumes to the background medium; 3.) The data in our application are typically noisy and have only limited view; and 4.) The number of the plumes is typically unknown, and their shapes can have a complicated geometry.

In this paper, we propose a new shape reconstruction method which works in a two-step fashion in order to overcome these difficulties.

The first step of this combined inversion scheme plays the role of an initializing procedure for the second step, and employs a 'source-type' inversion method (which is described in more details in section 5) to deal with the high nonlinearity in the problem due to the presence of strong scatterers.

Then, the second step directly starts with the outcome of this initializing procedure, and continues by using a combination of the 'adjoint field technique' and a level set representation of the shapes until the inversion task is completed. Using a level set representation in this second step enables us to easily describe and keep track of complicated geometries which arise during the inversion process.

Both steps use an 'adjoint field technique' for the inversion which has the very useful property that the inverse problem can be solved approximately by making two uses of the same forward modelling code. Using a somewhat oversimplified description of our technique, the updates to the level set function are obtained by first making one pass through the code using the permittivity distribution corresponding to the latest best guess of the level set function, and then another pass with the adjoint operator applied to the differences in computed and measured data. Then the results of these two calculations are combined to determine updates to the level set function. The resulting procedure is iterative, and can be applied successively to parts of the data, e.g., data associated with one transmitter location and one frequency can be used to update the model before other transmitter locations and other frequencies are considered. This general procedure has several of the same advantages as wave equation migration in reflection seismology [10] and is also related to recent methods in EM migration introduced in Zhdanov *et al.* [51]. A similar technique has been successfully applied recently as part of an iterative nonlinear inversion scheme in [17, 18, 33].

The *level set method* was originally developed by Osher and Sethian for describ-

ing the motion of curves and surfaces [35, 41]. Since then, it has found applications in a variety of quite different situations. Examples are image enhancement, computer vision, interface problems, crystal growth, or etching and deposition in the microchip fabrication. For an overview we refer to [42].

The idea of using a level set representation as part of a solution scheme for inverse problems involving obstacles was first suggested by Santosa in [40]. More recently, a similar method was applied to a nonlinear inverse scattering problem by Litman *et al.* in [28]. In that work, an inverse transmission problem in free space is solved by a controlled evolution of a level set function. This evolution is governed by a Hamilton-Jacobi type equation, whose velocity function has to be determined properly in order to minimize a given cost functional.

The approach developed here does not lead to a Hamilton-Jacobi type equation. We follow an optimization approach, and employ a very specific inversion routine (an adjoint field technique) for solving it. This has the advantage that we do not have to propagate the level set function explicitly by computing a numerical Hamiltonian. Instead, our inversion routine provides us in each step with an update that has to be applied *directly* to the most recent level set function. Doing so, we automatically 'propagate' the level set function until the method converges.

This gain in simplicity, however, has its price. In order to arrive at an efficient scheme which is practically useful as well as easy to implement, we will apply some suitable approximations when deriving the algorithm. We will point out and discuss these approximations in those sections of the paper where they are applied.

For interesting approaches to solving shape recovery problems in various applications we refer to [23, 24, 26] and to the references therein. For alternative approaches to the shape reconstruction method in geophysical applications see [30, 39, 43] and the references therein. The treatment of more general inverse scattering problems is for example addressed in [4, 8, 11, 12, 15, 22, 31, 33, 38].

The paper is organized as follows. In section 2 we will present the basic equations of 2D EMs in a form convenient for development of the shape reconstruction technique. In section 3 we formulate the shape reconstruction problem and introduce the level set formulation of this problem. In section 4, we derive the basic shape reconstruction algorithm using level sets and adjoint fields. Section 5 describes how to calculate a suitable initialization for the shape reconstruction algorithm. In Section 6 numerical experiments are presented which demonstrate the performance of the algorithm in different situations. The final section summarizes the results of this paper and indicates some directions for future research.

2 The physical experiment

2.1 The Helmholtz Equation

We consider the 2D Helmholtz Equation

$$\Delta u + k^2(x)u = q(x) \quad \text{in } \mathbf{R}^2, \quad (1)$$

with complex wavenumber

$$k^2(x) = \omega^2 \mu_0 \epsilon_0 \left[\epsilon(x) + i \frac{\sigma(x)}{\omega \epsilon_0} \right]. \quad (2)$$

Here, $i^2 = -1$, ω denotes the angular frequency $\omega = 2\pi f$, μ_0 is the magnetic permeability in free space $\mu_0 = 4\pi \times 10^{-7}$ Henrys per meter, ϵ_0 is the dielectric permittivity in free space $\epsilon_0 = 8.854 \times 10^{-12}$ Farads per meter, ϵ is the relative dielectric permittivity (dimensionless), and σ is the electric conductivity in Siemens per meter. The form of (2) corresponds to time-harmonic line sources $\tilde{q}(x, t)$ which have a time-dependence $\tilde{q}(x, t) = q(x)e^{-i\omega t}$. For these sources we require that there exists a radius $r_0 > 0$ such that $\text{supp}(q) \subset\subset B_{r_0}(0)$, where $B_r(x) = \{y \in \mathbf{R}^2, |x - y| < r\}$ denotes the open ball centered in x with radius $r > 0$. For simplicity we assume throughout the paper that we can find a ball $B_R(0)$ with $R > r$ such that the complex wavenumber $k^2(x)$ is constant with value k_0^2 in $\mathbf{R}^2 \setminus B_R(0)$, and that for this k_0 the field u generated by (1) satisfies the Sommerfeld radiation condition

$$\lim_{r \rightarrow \infty} \sqrt{r} \left(\frac{\partial u}{\partial r} - ik_0 u \right) = 0 \quad (3)$$

with $r = |x|$ where the limit is assumed to hold uniformly in all directions $x/|x|$. With this assumption, the problem (1)-(3) possesses a uniquely determined solution u in \mathbf{R}^2 [12].

Furthermore we will consider in this paper only the case that the conductivity is positive everywhere, $\sigma > 0$ in \mathbf{R}^2 , and that it is small in some sense which will be specified later. Typical values in our geophysical examples will be $\sigma \approx 10^{-3} - 10^{-4}$ Siemens per meter or less [48].

We want to introduce some notation here which will be useful in the following. We denote the wavenumber $k^2(x)$ in short form by

$$k^2(x) = \kappa(x) = a\epsilon(x) + ib\sigma(x), \quad a = \omega^2 \mu_0 \epsilon_0, \quad b = \omega \mu_0. \quad (4)$$

We only consider positive frequencies $\omega > 0$ such that $a, b > 0$.

2.2 Formulation of the inverse problem

We assume that we are given p different source distributions q_j , $j = 1, \dots, p$. For each of these sources, data are gathered at the detector positions x_d , $d = 1, \dots, D_j$, for various frequencies f_k , $k = 1, \dots, K$. The total number of receivers D_j , as well as their positions x_d , might vary with the source q_j . We assume, for simplicity in the notation, that these positions do not depend on the frequency f_k . This restriction is, however, not necessary for the derivation of the inversion method. We require that there exists a radius $r_1 > 0$ such that all receiver positions are inside the ball of radius r_1 , i.e. $x_{jd} \in B_{r_1}(0)$ for all $d = 1, \dots, D_j$, $j = 1, \dots, p$.

For a given source q_j and a given frequency f_k we collect a set of data \tilde{G}_{jk} which is described by

$$\tilde{G}_{jk} = \left(\tilde{u}_{jk}(x_{j1}), \dots, \tilde{u}_{jk}(x_{jd}), \dots, \tilde{u}_{jk}(x_{jD_j}) \right)^T \in Z_j \quad (5)$$

with $Z_j = \mathbb{C}^{D_j}$ being the data space corresponding to a single experiment using one source and one frequency only. In (5), the fields \tilde{u}_{jk} solve (1)-(3) with the correct permittivity distribution $\tilde{\epsilon}(x)$, i.e.

$$\Delta \tilde{u}_{jk} + [a_k \tilde{\epsilon}(x) + ib_k \sigma(x)] \tilde{u}_{jk} = q_j(x) \quad \text{in } \mathbb{R}^2 \quad (6)$$

with

$$a_k = \omega_k^2 \mu_0 \epsilon_0, \quad b_k = \omega_k \mu_0, \quad \omega_k = 2\pi f_k. \quad (7)$$

In a slightly more formal way, we define for a given source q_j the measurement operator M_j acting on solutions u of (1) by

$$M_j u = \left(\int_{\mathbb{R}^2} u(x) \delta(x - x_{jd}) dx \right)_{d=1, \dots, D_j}^T. \quad (8)$$

With this notation, (5) is written as

$$\tilde{G}_{jk} = M_j \tilde{u}_{jk}, \quad j = 1, \dots, p, \quad k = 1, \dots, K. \quad (9)$$

We gather these data sets $\tilde{G}_{j,k}$ for all sources q_j , $j = 1, \dots, p$, and all frequencies f_k , $k = 1, \dots, K$, and the aim is to recover from this collection of data sets

$$\tilde{G} = (\tilde{G}_{1,1}, \dots, \tilde{G}_{p,K})^T \quad (10)$$

the unknown parameter distribution $\tilde{\epsilon}(x)$ in the domain of interest.

In the application of *EM cross-borehole tomography*, the sources and receivers are typically situated in some boreholes, and the permittivity distribution ϵ (and/or the conductivity distribution σ) between these boreholes has to be recovered from

the gathered data. In the 2D geometry considered here, typical sources are time-harmonic line sources which can be modelled in (1) by

$$q_j(x) = J_j \delta(x - x_j), \quad j = 1, \dots, p, \quad (11)$$

where x_j denotes the 2D coordinates of the j -th line source, $j = 1, \dots, p$, and the complex number J_j is the strength of the source. We will use these sources in our numerical experiments in section 6.

3 The shape reconstruction problem

In this section we formulate the shape reconstruction problem which we want to solve, and cast it in a form which makes use of the level set representation of the domains.

3.1 Shape reconstruction and inverse scattering

To start with we introduce some terminology which we will use throughout the paper.

Definition 3.1 *Let us assume that we are given a constant $\hat{\epsilon} > 0$, an open ball $B_r(0) \subset \mathbb{R}^2$ with $r > \max(r_0, r_1) > 0$, and a bounded function $\epsilon_b : \mathbb{R}^2 \rightarrow \mathbb{R}$. We call a pair (Ω, ϵ) , which consists of a compact domain $\Omega \subset\subset B_r(0)$ and a bounded function $\epsilon : \mathbb{R}^2 \rightarrow \mathbb{R}$, admissible if we have*

$$\epsilon|_{\Omega} = \hat{\epsilon}, \quad \epsilon|_{\mathbb{R}^2 \setminus \Omega} = \epsilon_b|_{\mathbb{R}^2 \setminus \Omega}. \quad (12)$$

In other words, a pair (Ω, ϵ) is *admissible* if ϵ is equal to a preassigned constant value $\hat{\epsilon}$ inside of Ω , and equal to the preassigned *background permittivity* ϵ_b outside of Ω . The domain Ω is called the *scattering domain*.

Remark 3.1 *For an admissible pair (Ω, ϵ) , and for given $\hat{\epsilon}$, ϵ_b , the permittivity ϵ is uniquely determined by Ω .*

With this definition, we can now formulate the *shape reconstruction problem*.

Shape reconstruction problem. Let us assume that we are given a constant $\hat{\epsilon} > 0$, a bounded function $\epsilon_b : \mathbb{R}^2 \rightarrow \mathbb{R}$, and some data \tilde{G} as in (10). Find a domain $\tilde{\Omega}$ such that the admissible pair $(\tilde{\Omega}, \tilde{\epsilon})$ reproduces the data, i.e. (9) holds with \tilde{u}_{jk} given by (6) for $j = 1, \dots, p$, $k = 1, \dots, K$.

Using the same notation and assumptions as in definition 3.1, we want to formulate another inverse problem which we will call the *inverse scattering problem* and which will play an important part when solving the shape reconstruction problem.

Inverse Scattering Problem. Let us assume that we are given a bounded function $\epsilon_b : \mathbb{R}^2 \rightarrow \mathbb{R}$, and some data \tilde{G} as in (10). Find a bounded function $\tilde{\epsilon}_s : \mathbb{R}^2 \rightarrow \mathbb{R}$ with $\text{supp}(\epsilon_s) \subset\subset B_r(0)$ such that $\tilde{\epsilon} = \epsilon_b + \tilde{\epsilon}_s$ reproduces the data, i.e. (9) holds with \tilde{u}_{jk} given by (6) for $j = 1, \dots, p$, $k = 1, \dots, K$.

The inverse scattering problem gives rise to the following decomposition of ϵ in \mathbb{R}^2 .

Decomposition of $\epsilon(x)$:

$$(i) \quad \epsilon = \epsilon_b + \epsilon_s \quad \text{in } \mathbb{R}^2 \quad (13)$$

$$(ii) \quad \text{supp}(\epsilon_s) \subset\subset B_r(0). \quad (14)$$

In other words, the permittivity distribution ϵ is decomposed into the background distribution ϵ_b and the perturbation ϵ_s which is assumed to have compact support and which we will refer to as the *scattering potential* in the following.

Solving the shape reconstruction problem requires only to find the *shape* of the domain $\tilde{\Omega}$, since the function $\tilde{\epsilon}$ is then uniquely determined by (12). Solving the inverse scattering problem, on the other hand, amounts to finding the entire function $\tilde{\epsilon}_s$ from the given data, which is much harder to do. However, it will turn out that finding a good *approximate* solution of the inverse scattering problem is much easier to achieve and will provide us with an excellent initial guess for starting our shape reconstruction routine.

Definition 3.1 allows us to formulate a first version of the strategy which we want to use for solving the shape reconstruction problem.

Strategy for solving the shape reconstruction problem. Construct a series of admissible pairs $(\Omega^{(n)}, \epsilon^{(n)})$, $n = 0, 1, 2, \dots$, such that the misfit between the data (10) and the calculated data corresponding to $(\Omega^{(n)}, \epsilon^{(n)})$ decreases with increasing n , and ideally, i.e. in absence of noise, tends to zero in the limit $n \rightarrow \infty$. Use the *approximate* solution of the inverse scattering problem (i),(ii) to initialize this series by determining a good starting element $(\Omega^{(0)}, \epsilon^{(0)})$.

3.2 The domains $\Omega^{(n)}$

In our numerical examples, each of the domains $\Omega^{(n)}$ which we are looking for can be given as a collection of a finite number L_n of disjoint, compact subdomains $\Omega_l^{(n)}$,

$l = 1, \dots, L_n$, with

$$\Omega^{(n)} = \bigcup_{l=1}^{L_n} \Omega_l^{(n)}, \quad \Omega_l^{(n)} \cap \Omega_{l'}^{(n)} = \emptyset \quad \text{for } l \neq l'. \quad (15)$$

The shapes of these subdomains $\Omega_l^{(n)}$ can in principle be arbitrary. In particular, they are allowed to be multiply connected, and to enclose some 'cavities' or 'holes' filled with background material. Moreover, the number L_n of these subdomains might (and usually does) vary with the iteration number n . For the derivation of the inversion method, we assume that the boundaries $\partial\Omega_l^{(n)}$ of these domains are sufficiently smooth (e.g. C^1).

It is essential for the success and the efficiency of the reconstruction scheme to have a good and flexible way of keeping track of the shape evolution during the reconstruction process. The method we have chosen in our reconstruction algorithm is a *level set representation* of the shapes as it was suggested by Santosa [40]. This representation has the advantage that the level set functions, which are in principle only used for *representing* the shapes, can in a natural way be made part of the reconstruction scheme itself. Doing so, it is not necessary anymore to refer to the shapes of the domains until the reconstruction process is completed. The *final* shape is then recovered from the representing level set function easily. In the following we will discuss in a more formal way how this can be achieved.

3.3 Level set representation of the domains $\Omega^{(n)}$

Assume that we are given a domain $\Omega \subset\subset B_r(0)$. The characteristic function $\chi_\Omega : \mathbb{R}^2 \rightarrow \{0, 1\}$ is defined in the usual way as

$$\chi_\Omega(x) = \begin{cases} 1 & , \quad x \in \Omega \\ 0 & , \quad x \in \mathbb{R}^2 \setminus \Omega. \end{cases} \quad (16)$$

Definition 3.2 We call a function $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ a level set representation of Ω if

$$\chi_\Omega(x) = \Psi_\phi(x) \quad \text{on } \mathbb{R}^2 \quad (17)$$

where $\Psi_\phi : \mathbb{R}^2 \rightarrow \{0, 1\}$ is defined as

$$\Psi_\phi(x) = \begin{cases} 1 & , \quad \phi(x) \leq 0 \\ 0 & , \quad \phi(x) > 0. \end{cases} \quad (18)$$

For each function $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ there is a domain Ω associated with ϕ by (17),(18) which we call $\Omega[\phi]$. It is clear that different functions ϕ_1, ϕ_2 , $\phi_1 \neq \phi_2$, can be associated with the same domain $\Omega[\phi_1] = \Omega[\phi_2]$, but that different domains cannot have

the same level set representation. Therefore, we can use the level set representation for unambiguously specifying a domain Ω by any one of its associated level set functions.

The boundary $\Gamma = \partial\Omega[\phi]$ of a domain $\Omega[\phi]$, represented by the level set function ϕ , is defined as

$$\begin{aligned} \Gamma = \{x \in \mathbb{R}^2 : \quad & \text{for all } \epsilon > 0 \text{ we can find } x_1, x_2 \in B_\epsilon(x) \\ & \text{with } \phi(x_1) > 0 \text{ and } \phi(x_2) < 0\} \end{aligned} \quad (19)$$

Definition 3.3 We call a triple (Ω, ϵ, ϕ) , which consists of a domain $\Omega \subset\subset B_r(0)$ and bounded functions $\epsilon, \phi : \mathbb{R}^2 \rightarrow \mathbb{R}$, admissible if the pair (Ω, ϵ) is admissible in the sense of definition 3.1, and ϕ is a valid level set representation of Ω .

Remark 3.2 For an admissible triple (Ω, ϵ, ϕ) , and for given $\hat{\epsilon}, \epsilon_b$, the pair (Ω, ϵ) is uniquely determined by ϕ .

We use these definitions to reformulate our shape reconstruction problem.

Level set formulation of the shape reconstruction problem. Given a constant $\hat{\epsilon} > 0$, a background distribution ϵ_b , and some data \tilde{G} as in (10). Find a level set function $\tilde{\phi}$ such that the corresponding admissible triple $(\tilde{\Omega}, \tilde{\epsilon}, \tilde{\phi})$ reproduces the data, i.e. (9) holds with \tilde{u}_{jk} given by (6) for $j = 1, \dots, p$, $k = 1, \dots, K$.

The strategy for solving this shape reconstruction problem has to be reformulated, too. It reads now as follows.

Strategy for solving the reformulated shape reconstruction problem. Construct a series of admissible triples $(\Omega^{(n)}, \epsilon^{(n)}, \phi^{(n)})$, $n = 0, 1, 2, \dots$, such that the misfit between the data (10) and the calculated data corresponding to $(\Omega^{(n)}, \epsilon^{(n)}, \phi^{(n)})$ decreases with increasing n , and ideally, i.e. in absence of noise, tends to zero in the limit $n \rightarrow \infty$. For finding this series we only have to keep track of $\epsilon^{(n)}$ and $\phi^{(n)}$, but not of $\Omega^{(n)}$. The function $\epsilon^{(n)}$ is needed in each step for solving a forward problem (1), and a corresponding adjoint problem. The knowledge of $\phi^{(n)}$ is used in each step to determine $\epsilon^{(n)}$. The final level set function $\phi^{(N)}$, which satisfies some stopping criterion, is used to recover the final shape $\Omega^{(N)}$ via (17).

4 Step 2: Solving the shape reconstruction problem

In this section we derive the basic shape reconstruction method which uses adjoint fields and the level set representation introduced above. The initializing procedure ('Step 1') for this reconstruction routine will be discussed in section 5.

4.1 Function spaces

We want to specify now the function spaces which we will be working with. The main objective of this section is to introduce the inner products on these function spaces, which will become important when defining the adjoint linearized operators in sections 4.6 and 5.3.

The space of sources and scattering sources Y is defined as

$$Y = \left\{ q : \mathbb{R}^2 \rightarrow \mathbb{C}, \quad q = 0 \text{ on } \mathbb{R}^2 \setminus B_r(0), \quad \int_{B_r(0)} |q|^2 dx < \infty \right\}, \quad (20)$$

$$\langle q_1, q_2 \rangle_Y = \int_{B_r(0)} q_1(x) \overline{q_2(x)} dx, \quad (21)$$

where the bar means 'complex conjugate'. The space F of scattering potentials is defined as

$$F = \left\{ \epsilon_s : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad \epsilon_s = 0 \text{ on } \mathbb{R}^2 \setminus B_r(0), \quad \int_{B_r(0)} |\epsilon_s|^2 dx < \infty \right\}, \quad (22)$$

$$\langle \epsilon_{s,1}, \epsilon_{s,2} \rangle_F = \int_{B_r(0)} \gamma \epsilon_{s,1}(x) \epsilon_{s,2}(x) dx, \quad (23)$$

with some positive weighting factor $\gamma > 0$ which is introduced here for convenience. Analogously, the space of level set functions Φ is defined as

$$\Phi = \left\{ \phi : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad \phi = 0 \text{ on } \mathbb{R}^2 \setminus B_r(0), \quad \int_{B_r(0)} |\phi|^2 dx < \infty \right\}, \quad (24)$$

$$\langle \phi_1, \phi_2 \rangle_\Phi = \int_{B_r(0)} \gamma \phi_1(x) \phi_2(x) dx. \quad (25)$$

This space for the level set functions is mainly chosen in order to have an inner product available which is convenient for the derivation of the shape reconstruction algorithm. We mention that the regularity of an arbitrary function in Φ is, strictly speaking, not sufficient for our purposes, such that we will apply further regularity constraints on those level set functions $\phi \in \Phi$ which we choose for representing the boundaries $\partial\Omega[\phi]$.

The data space Z_j corresponding to source q_j , $j = 1, \dots, p$, was already introduced earlier, and is given by $Z_j = \mathbb{C}^{D_j}$, where D_j is the total number of receivers corresponding to source q_j .

4.2 Operators

In the following, we will introduce some operators which will enable us to formulate the shape reconstruction problem in a way suitable for deriving the inversion algorithm.

Given a constant $\hat{\epsilon}$ and a bounded function $\epsilon_b : \mathbb{R}^2 \rightarrow \mathbb{R}$. Then, with each level set function $\phi \in \Phi$ a uniquely determined scattering potential $\Lambda(\phi)$ is associated by putting

$$\Lambda(\phi)(x) = \begin{cases} \hat{\epsilon} - \epsilon_b(x) & , \quad \phi(x) \leq 0 \\ 0 & , \quad \phi(x) > 0. \end{cases} \quad (26)$$

With (18) we can write this also as

$$\Lambda(\phi)(x) = \Psi_\phi(x)(\hat{\epsilon} - \epsilon_b(x)) \quad , \quad x \in \mathbb{R}^2. \quad (27)$$

Notice that the operator Λ is chosen such that the triple (Ω, ϵ, ϕ) with $\epsilon = \epsilon_b + \Lambda(\phi)$ and domain $\Omega[\phi]$ forms an admissible triple (Ω, ϵ, ϕ) in the sense of definition 3.3. Moreover, for (Ω, ϵ, ϕ) an admissible triple, we see that $\Lambda(\phi)$ is just the scattering potential ϵ_s as defined in (13),(14)

$$\Lambda(\phi)(x) = \epsilon_s(x) = \chi_\Omega(x)(\hat{\epsilon} - \epsilon_b(x)) \quad , \quad x \in \mathbb{R}^2. \quad (28)$$

Let us assume now that we are given a background permittivity ϵ_b and that we have collected some data \tilde{G}_{jk} which correspond to the 'true' permittivity distribution

$$\tilde{\epsilon} = \epsilon_b + \tilde{\epsilon}_s, \quad (29)$$

where $\tilde{\epsilon}_s$ is the 'true' scattering potential. The *residual operators* R_{jk} map for a source position q_j and a frequency f_k a given scattering potential ϵ_s to the corresponding mismatch in the data

$$R_{jk} : F \longrightarrow Z_j \quad , \quad R_{jk}(\epsilon_s) = M_j u_{jk} - \tilde{G}_{jk} \quad (30)$$

where u_{jk} solves

$$\Delta u_{jk} + [a_k(\epsilon_b + \epsilon_s)(x) + ib\sigma(x)] u_{jk} = q_j \quad (31)$$

and M_j is the measurement operator defined in (8). From (9) we see that for the 'true' scattering potential the residuals vanish,

$$R_{jk}(\tilde{\epsilon}_s) = 0 \quad \text{for } j = 1, \dots, p, \quad k = 1, \dots, K, \quad (32)$$

if the data are noise-free.

The *forward operators* T_{jk} which map a given level set function $\phi \in \Phi$ into the corresponding mismatch in the data are defined by

$$T_{jk} : \Phi \longrightarrow Z_j \quad , \quad T_{jk}(\phi) = R_{jk}(\Lambda(\phi)) \quad (33)$$

for $j = 1, \dots, p, \quad k = 1, \dots, K$. The goal is to find a level set function $\tilde{\phi} \in \Phi$ such that

$$T_{jk}(\tilde{\phi}) = 0 \quad \text{for } j = 1, \dots, p, \quad k = 1, \dots, K. \quad (34)$$

We mention that all three operators Λ , R_{jk} and T_{jk} are nonlinear.

4.3 Linearized operators

For the derivation of the shape reconstruction algorithm, we will need expressions for the linearized operators corresponding to the nonlinear operators introduced above, and for their adjoints with respect to the given inner products. In this section, we define the linearized operators, and expressions for their adjoints are derived in section 4.6.

In Santosa [40] it is shown that, for a homogeneous background ϵ_b , the infinitesimal response $\delta\epsilon_s(x)$ in the scattering potential $\epsilon_s(x)$ to an infinitesimal change $\delta\phi(x)$ of the level set function $\phi(x)$ has the form

$$\delta\epsilon_s(x) = - [\hat{\epsilon} - \epsilon_b] \frac{\delta\phi(x)}{|\nabla\phi(x)|} \Big|_{x \in \partial\Omega[\phi]}. \quad (35)$$

The function $\delta\epsilon_s$ in (35) can be interpreted as a 'surface measure' on the boundary $\Gamma = \partial\Omega[\phi]$. Similar to (35), we would like to define the linearized operator $\tilde{\Lambda}'[\phi]$ by

$$\left(\tilde{\Lambda}'[\phi] \delta\phi \right) (x) = - [\hat{\epsilon} - \epsilon_b(x)] \frac{\delta\phi(x)}{|\nabla\phi(x)|} \hat{\delta}_\Gamma(x) \quad (36)$$

where $\hat{\delta}_\Gamma(x)$ denotes the Dirac delta distribution concentrated on $\Gamma = \partial\Omega[\phi]$. In this interpretation, (36) describes an infinitesimal 'surface load' of permittivity on Γ which has to be recovered from the mismatch in the data.

However, the expression on the right hand side of (36) is not an element of F which causes problems when we want to calculate the inner products defined in section 4.1. Therefore, we will introduce an approximation to the operator (36) which maps from Φ into F and which will be more convenient for the derivation of the reconstruction method.

For a given level set function $\phi \in \Phi$, let $\Gamma = \partial\Omega[\phi]$ and $B_\rho(\Gamma) = \cup_{y \in \Gamma} B_\rho(y)$ a small neighborhood of Γ with some given constant $0 < \rho \ll 1$. The (approximated) linearized operator $\Lambda'[\phi]$ is defined as

$$\Lambda'[\phi] : \Phi \longrightarrow F, \quad (\Lambda'[\phi] \delta\phi) (x) = - [\hat{\epsilon} - \epsilon_b(x)] \frac{\delta\phi(x)}{|\nabla\phi(x)|} C_\rho(\Gamma) \chi_{B_\rho(\Gamma)}(x). \quad (37)$$

Here, $C_\rho(\Gamma) = L(\Gamma)/\text{Vol}(B_\rho(\Gamma))$ where $L(\Gamma) = \int_{B_r(0)} \hat{\delta}_\Gamma(x) dx$ is the length of the boundary Γ , and $\text{Vol}(B_\rho(\Gamma)) = \int_{B_r(0)} \chi_{B_\rho(\Gamma)}(x) dx$ is the volume of $B_\rho(\Gamma)$. For a very small ρ we will get a very large weight $C_\rho(\Gamma)$, whereas for increasing ρ this weight $C_\rho(\Gamma)$ decreases accordingly. The operator defined in (37) maps now from Φ into F such that we can make use of the inner products defined on these spaces.

We mention that the term $|\nabla\phi(x)|$ in (35), (36), as well as the derivation of these expressions, implies some regularity constraint on ϕ . For example, $\phi \in C^1$ would be possible. Another possibility would be to use a 'signed distance function'

as a standard representation of the boundary [42]. We do not want to specify the regularity of ϕ at this point, but assume instead that it is 'sufficiently smooth' for our purposes.

The linearized residual operator $R'_{jk}[\epsilon_s]$ is defined by

$$R'_{jk}[\epsilon_s] : F \longrightarrow Z_j \quad , \quad R'_{jk}[\epsilon_s]\delta\epsilon_s = M_j v_{jk} \quad (38)$$

where v_{jk} solves the linearized equation

$$\Delta v_{jk} + [a_k(\epsilon_b + \epsilon_s)(x) + ib\sigma(x)] v_{jk} = -a_k \delta\epsilon_s(x) u_{jk}(x) \quad (39)$$

with u_{jk} a solution of (31). This representation can be derived by perturbing

$$\epsilon_s \rightarrow \epsilon_s + \delta\epsilon_s \quad , \quad u_{jk} \rightarrow u_{jk} + v_{jk}, \quad (40)$$

plugging this into (31) and neglecting terms which are of higher than linear order in the perturbations $\delta\epsilon_s, v_{jk}$.

Notice that the right hand side of (39) can be interpreted as a 'scattering source'. We will use this concept later for solving our inverse scattering problem approximately in order to find a starting guess for the shape reconstruction scheme. But we want to mention here already that the linearization assumption built into (38), (39), namely that v_{jk} is small compared to u_{jk} , will not be necessary when solving the inverse scattering problem. That will allow us to circumvent some difficulties which often arise in high contrast inverse problems due to the occurrence of strong nonlinearities.

As our third linearized operator, we introduce the linearized forward operator $T'_{jk}[\phi]$ by putting

$$T'_{jk}[\phi] : \Phi \longrightarrow Z_j \quad , \quad T'_{jk}[\phi]\delta\phi = R'_{jk}[\Lambda(\phi)] \Lambda'[\phi]\delta\phi. \quad (41)$$

All three operators $\Lambda'[\phi]$, $R'_{jk}[\epsilon_s]$, and $T'_{jk}[\phi]$ are linear.

4.4 A nonlinear Kaczmarz-type approach

The algorithm works in a 'single-step fashion' as follows. Instead of using the data (10) for all sources and all frequencies simultaneously, we only use the data for one source and frequency at a time while updating the linearized residual operator after each determination of the corresponding incremental correction $\delta\phi$. So, in each step we will look for a solution of the equation

$$T'_{jk}[\phi]\delta\phi_{jk} = -T_{jk}(\phi) \quad (42)$$

for a given source index $j = 1, \dots, p$ and a given frequency index $k = 1, \dots, K$. After correcting ϕ by

$$\phi \longrightarrow \phi + \delta\phi_{jk}, \quad (43)$$

we use the updated residual equation (42) to compute the next correction $\delta\phi_{j'k'}$. Doing this for one equation after the other, until each of the sources q_j and each of the frequencies f_k has been considered exactly once, will yield one complete sweep of the algorithm. This procedure is similar to the Kaczmarz method for solving linear systems, or the algebraic reconstruction technique (ART) in x-ray tomography [32] and the simultaneous iterative reconstruction technique (SIRT) as presented in [16]. Related approaches have also been employed in ultrasound tomography by Natterer and Wübbeling [33], in more general bilinear inverse problems by Natterer [34], in optical tomography by Dorn [17], and in 3D-electromagnetic induction tomography (EMIT) by Dorn *et alii* [18].

4.5 The minimization problem

Let us assume now that we are given a level set function $\phi^{(n)}(x)$ and a scattering potential $\epsilon_s^{(n)}(x)$ such that $(\Omega^{(n)}, \epsilon_b + \epsilon_s^{(n)}, \phi^{(n)})$ forms an admissible triple in the sense of definition 3.3. Using a data set \tilde{G}_{jk} corresponding to the fixed source position q_j and the frequency f_k , we want to find an update $\delta\phi^{(n)}$ to $\phi^{(n)}$ such that for the admissible triple

$$\begin{aligned} & \left(\Omega^{(n+1)}, \epsilon_b + \epsilon_s^{(n+1)}, \phi^{(n+1)} \right) := \\ & \left(\Omega[\phi^{(n)} + \delta\phi^{(n)}], \epsilon_b + \Lambda(\phi^{(n)} + \delta\phi^{(n)}), \phi^{(n)} + \delta\phi^{(n)} \right) \end{aligned} \quad (44)$$

the residuals in the data corresponding to this source and this frequency vanish

$$T_{jk}(\phi^{(n+1)}) = T_{jk}(\phi^{(n)} + \delta\phi^{(n)}) = 0. \quad (45)$$

Applying a Newton-type approach, we get from (45) a correction $\delta\phi^{(n)}$ for $\phi^{(n)}$ by solving

$$T'_{jk}[\phi^{(n)}]\delta\phi^{(n)} = -T_{jk}(\phi^{(n)}) = -\left(M_j u_{jk} - \tilde{G}_{jk}\right) \quad (46)$$

where u_{jk} satisfies (31) with $\epsilon_s = \Lambda(\phi^{(n)})$

$$\Delta u_{jk} + \left[a_k(\epsilon_b + \Lambda(\phi^{(n)}))(x) + ib\sigma(x) \right] u_{jk} = q_j(x) \quad (47)$$

and

$$\epsilon_b(x) + \Lambda(\phi^{(n)})(x) = \begin{cases} \hat{\epsilon} & , \quad x \in \Omega[\phi^{(n)}] \\ \epsilon_b(x) & , \quad x \in \mathbb{R}^2 \setminus \Omega[\phi^{(n)}]. \end{cases} \quad (48)$$

Since we have only few data given for one source and one frequency, equation (46) usually will have many solutions (in the absence of noise), such that we have to pick

one according to some criterion. We choose to take that solution which minimizes the energy norm of $\delta\phi^{(n)}$

$$\text{Min } \|\delta\phi^{(n)}\|_2 \quad \text{subject to} \quad T'_{jk}(\phi^{(n)})\delta\phi^{(n)} = -\left(M_j u_{jk} - \tilde{G}_{jk}\right). \quad (49)$$

This solution can be formulated explicitly. It is

$$\delta\phi_{\text{MN}}^{(n)} = -T'_{jk}[\phi^{(n)}]^* \left(T'_{jk}[\phi^{(n)}]T'_{jk}[\phi^{(n)}]^*\right)^{-1} \left(M_j u_{jk} - \tilde{G}_{jk}\right), \quad (50)$$

where $T'_{jk}[\phi^{(n)}]^*$ denotes the adjoint operator to $T'_{jk}[\phi^{(n)}]$.

4.6 The adjoint linearized operators

In order to calculate the minimal norm solution (50), we will need practically useful expressions for the adjoints of the linearized operators of section (4.3). We will present such expressions in this section. The calculation of the actions of these operators will typically require to solve an adjoint Helmholtz problem. This explains the name 'adjoint field method' of the inversion method employed here.

To start with, a simple calculation gives us the following theorem.

Theorem 4.1 *The adjoint operator $\Lambda'[\phi]^*$ which corresponds to the linearized operator $\Lambda'[\phi]$ is given by*

$$\Lambda'[\phi]^* : F \longrightarrow \Phi \quad , \quad (\Lambda'[\phi]^* \delta\epsilon_s)(x) = -[\hat{\epsilon} - \epsilon_b(x)] \frac{\delta\epsilon_s(x)}{|\nabla\phi(x)|} C_\rho(\Gamma) \chi_{B_\rho(\Gamma)}(x). \quad (51)$$

The next theorem describes the adjoint operator $R'_{jk}[\epsilon_s]^*$ which corresponds to $R'_{jk}[\epsilon_s]$. Its proof is analogous to the proof of Theorem 4.1 in the appendix, or to the proof given in a similar situation in Dorn *et al.* [18], and is therefore omitted here.

Theorem 4.2 *Let $\zeta = (\zeta_1, \dots, \zeta_{D_j})^T \in Z_j$ and let x_{jd} , $d = 1, \dots, D_j$ be the detector positions corresponding to q_j . Then the action of the adjoint operator $R'_{jk}[\epsilon_s]^*$ on ζ is given by*

$$R'_{jk}[\epsilon_s]^* \zeta = -\frac{1}{a_k} \mathcal{R}e(\overline{u_{jk} z_{jk}}) \chi_{B_r(0)} \quad (52)$$

where u_{jk} solves

$$\Delta u_{jk} + \kappa_k(x) u_{jk} = q_j(x), \quad (53)$$

and z_{jk} solves the 'adjoint equation'

$$\Delta z_{jk} + \kappa_k(x) z_{jk} = \sum_{d=1}^{D_j} \bar{\zeta}_d \delta(x - x_{jd}) \quad (54)$$

with

$$\kappa_k(x) = a_k[\epsilon_b(x) + \epsilon_s(x)] + ib_k \sigma(x) \quad (55)$$

and a_k, b_k defined as in (7).

Finally, by combining theorems 4.1 and 4.2, we get an expression for the adjoint operator $T'_{jk}[\phi]^*$ which corresponds to the linearized forward operator $T'_{jk}[\phi]$. It is described in the following theorem.

Theorem 4.3 *Let $\zeta = (\zeta_1, \dots, \zeta_{D_j})^T \in Z_j$ and let x_{jd} , $d = 1, \dots, D_j$ be the detector positions corresponding to q_j . Then the adjoint operator $T'_{jk}[\epsilon_s]^*$ acts on ζ in the following way*

$$\begin{aligned} T'_{jk}[\phi]^* \zeta &= \Lambda'[\phi]^* R'_{jk}[\Lambda(\phi)]^* \zeta \\ &= \frac{[\hat{\epsilon} - \epsilon_b(x)]}{a_k |\nabla \phi(x)|} \mathcal{R}e(\overline{u_{jk} z_{jk}}) C_\rho(\Gamma) \chi_{B_\rho(\Gamma)}(x), \end{aligned} \quad (56)$$

where u_{jk} solves (53) and z_{jk} solves (54) with ϵ_s replaced by $\Lambda(\phi)$.

4.7 The operators $T'_{jk} T'^*_{jk}$

Let us consider the operator

$$C_{jk}^{(n)} := T'_{jk}[\phi^{(n)}] T'^*_{jk}[\phi^{(n)}] \quad (57)$$

in (50) more closely. Using (41) it gets the form

$$C_{jk}^{(n)} = R'_{jk}[\Lambda(\phi^{(n)})] \Lambda'[\phi^{(n)}] \Lambda'[\phi^{(n)}]^* R'_{jk}[\Lambda(\phi^{(n)})]^*. \quad (58)$$

With (37), (51) we see that, due to the operator $\Lambda'[\phi^{(n)}] \Lambda'[\phi^{(n)}]^*$ in (58), $C_{jk}^{(n)}$ maps first from the data space to functions in F or Φ which are supported on $B_\rho(\Gamma)$, and then back to the data space. In a discretized setting, it might happen that for a coarse mesh (and a small ρ) the number of pixels representing $B_\rho(\Gamma)$ becomes close to (or even smaller than) the number of data points. This observation lets us expect that the inversion of $C_{jk}^{(n)}$ in (50) will be highly ill-conditioned and unstable. This is confirmed by our numerical experiments so far.

Therefore, we will regularize the inversion of $C_{jk}^{(n)}$ and the action of its inverse on the right hand side of (46).

4.8 Regularization of $T'_{jk} T'^*_{jk}$

A standard way of regularization is the Tychonov-Phillips regularization scheme which amounts to replacing the operator $T'_{jk}[\phi^{(n)}] T'^*_{jk}[\phi^{(n)}]$ in (50) by the operator

$$T'_{jk}[\phi^{(n)}] T'^*_{jk}[\phi^{(n)}] + \lambda I \quad (59)$$

with some suitably chosen regularization parameter $\lambda > 0$. A small λ means little regularization, whereas in the case of very noisy data we might wish to use a very

large λ such that (59) is dominated by the term λI and we can approximate it further by a simple multiplication with the regularization parameter λ .

However, motivated by the above mentioned observations, we choose a different form of regularization. First, we add a Tychonov-Phillips term to $\Lambda'[\phi^{(n)}] \Lambda'[\phi^{(n)}]^*$ such that the right hand side of (58) becomes

$$C_{jk}^{(n)} \approx R'_{jk}[\Lambda(\phi^{(n)})] \left(\Lambda'[\phi^{(n)}] \Lambda'[\phi^{(n)}]^* + \lambda I \right) R'_{jk}[\Lambda(\phi^{(n)})]^*. \quad (60)$$

Now, using a very large regularization parameter λ , we approximate (60) further by

$$C_{jk}^{(n)} \approx \lambda R'_{jk}[\Lambda(\phi^{(n)})] R'_{jk}[\Lambda(\phi^{(n)})]^*. \quad (61)$$

Since calculating the operator (61) in each step of the inversion routine is still very time-consuming, we approximate this operator further by replacing the argument $\Lambda(\phi^{(n)})$ by the background scattering potential which is zero. Therefore, we end up with the following approximation for $C_{jk}^{(n)}$

$$C_{jk}^{(n)} \approx \hat{C}_{jk} := R'_{jk}[0] R'_{jk}[0]^* \quad \text{for all } n = 1, 2, \dots \quad (62)$$

The multiplier λ is neglected in (62) since it becomes part of the relaxation parameter in our inversion scheme. We see that we have replaced in (62) the highly ill-conditioned and difficult to calculate operator $C_{jk}^{(n)}$ by a much better conditioned operator \hat{C}_{jk} which has to be computed only once and which can be precalculated before starting the inversion routine.

The next theorem tells us how to practically compute the operator \hat{C}_{jk} for a given background permittivity ϵ_b .

Theorem 4.4 *Let us assume that we are given a background permittivity distribution ϵ_b , a finite set of sources q_j , $j = 1, \dots, p$, and for each of these sources a finite set of receiver positions x_{jd} , $d = 1, \dots, D_j$. We apply each of the sources with K different frequencies f_k , $k = 1, \dots, K$. The operators \hat{C}_{jk} , $j = 1, \dots, p$, $k = 1, \dots, K$, are then described by $D_j \times D_j$ matrices of the form*

$$\hat{C}_{jk} = \left\{ \int_{B_r(0)} \overline{u_{jk}(x) \varphi_{dk}(x)} u_{jk}(x) \varphi_{lk}(x) \right\}_{\substack{l=1, \dots, D_j \\ d=1, \dots, D_j}}, \quad (63)$$

where $\varphi_{\nu k}$ solves

$$\Delta \varphi_{\nu k} + (a_k \epsilon_b + i b_k \sigma) \varphi_{\nu k} = \delta(x - x_{j\nu}) \quad (64)$$

and u_{jk} solves

$$\Delta u_{jk} + (a_k \epsilon_b + i b_k \sigma) u_{jk} = q_j. \quad (65)$$

The index ν in (64) can stand for a receiver index d or l . In (63), the receiver index l is the row index, and the receiver index d is the column index of \hat{C}_{jk} .

The proof of this theorem is similar to the proof of Theorem 5.2 given in the appendix such that we omit it here.

4.9 Updating the level set function

In order to calculate a correction $\delta\phi_{\text{MN}}^{(n)}$ by (50) we have to apply the operator $T'_{jk}[\phi^{(n)}]^*$ to the vector

$$\zeta := \hat{C}_{jk}^{-1} \left(M_j u_{jk} - \tilde{G}_{jk} \right). \quad (66)$$

An explicit formula for $T'_{jk}[\phi^{(n)}]$ was already given in (56)

$$T'_{jk}[\phi^{(n)}]^* \zeta = \frac{[\hat{\epsilon} - \epsilon_b(x)]}{a_k |\nabla \phi^{(n)}(x)|} \mathcal{R}e \left(\overline{u_{jk}(x) z_{jk}(x)} \right) C_\rho(\Gamma) \chi_{B_\rho(\Gamma)}(x), \quad (67)$$

where u_{jk} and z_{jk} solve (53)- (55) with ϵ_s replaced by $\Lambda(\phi^{(n)})$.

To stabilize the reconstruction scheme, we replace the term $|\nabla \phi(x)|$ in (67) by some constant c_1 . Doing so we avoid dividing by numerical derivatives which might cause instabilities due to numerical noise and roundoff errors. This is justified as long as $|\nabla \phi(x)|$ does not vary too much along the boundary. It turns out that the updates we apply in our numerical examples to the level set functions usually justify this assumption. In cases with limited view and very noisy data, however, we will apply an additional 'smoothing procedure' (which is described in section 4.10) to the level set functions near the boundary after each update in order to guarantee the necessary regularity for the succeeding steps.

With these modifications, (50) yields the following *update formula for the level set function*

$$\delta\hat{\phi}^{(n)}(x) = - \frac{\hat{\epsilon} - \epsilon_b(x)}{c_1 a_k} \mathcal{R}e \left(\overline{u_{jk}(x) z_{jk}(x)} \right) C_\rho(\Gamma) \chi_{B_\rho(\Gamma)}(x) \quad (68)$$

where u_{jk} and z_{jk} solve (53)- (55) with ζ given by (66) and ϵ_s replaced by $\Lambda(\phi^{(n)})$.

Notice that, although we did not explicitly impose any regularity constraints on the updates (68), they are in the range of $R'_{jk}[\Lambda(\phi^{(n)})]^*$ (up to the factor $\hat{\epsilon} - \epsilon_b(x)$) which implicitly gives us some information about the regularity we can expect. Our numerical experiments so far indicate that the degree of regularity which is achieved by applying (68) is typically sufficient 'for practical purposes' in those situations where the data are not too noisy and where we have suitably arranged receiver positions all around the obstacles. (This is the 'full view' situation.)

However, in cases of noisy and limited-view data, the resulting boundaries look rough and fuzzy, in particular when high-frequency data are used for the reconstruction. In these situations, we can improve the results by applying some additional regularization on ϕ . A possible way of doing so is to filter the level set function after

each update in order to smooth it locally. An example for such a procedure is 'curve shortening by diffusion', which is briefly described at the end of section 4.10.

We mention that an interesting (and from a mathematical point of view more satisfactory) alternative to this procedure would be to apply some additional regularity constraints already in the derivation of (68), such that we would not have to worry at all about the smoothness of the resulting level set functions. We will investigate possible ways of doing so in our future research.

4.10 Implementation: The levelART algorithm

In brief algorithmic form, the nonlinear Kaczmarz-type method for shape reconstruction using level sets (which we call for short 'levelART' because of its above mentioned similarity to the 'ART' algorithm in x-ray tomography) can be written in the following way.

Preparation step.

- Calculate \hat{C}_{jk} and

$$D_{jk} = \hat{C}_{jk}^{-1} \quad (69)$$

according to (63) for each source q_j , $j = 1, \dots, p$, and each frequency f_k , $k = 1, \dots, K$, and store in memory for later use.

- Build groups of frequencies $G_m = \{f_1, \dots, f_{K_m}\}$, $m = 1, \dots, M$.

Initialization.

$n = 0$;

$(\Omega^{(0)}, \epsilon^{(0)}, \phi^{(0)})$ given from STAF.

Reconstruction loop.

```

FOR  $m = 1 : M$       march over frequency groups  $G_m$ 
  FOR  $i = 1 : I_m$     perform  $I_m$  sweeps for frequency group  $G_m$ 
    FOR  $k = 1 : K_m$   march over frequencies in  $G_m$ 
      FOR  $j = 1 : p$    march over sources  $q_j$  for each frequency
         $\zeta_{jk} = D_{jk}(M_j u_{jk} - \tilde{G}_{jk})$ ;  $u_{jk}$  solves (53) with  $\epsilon^{(n)}$ 
         $\delta\phi_{jk} = -\frac{\hat{\epsilon} - \epsilon_b(x)}{a_k} \mathcal{R}e(\overline{u_{jk} z_{jk}}) \chi_{B_\rho(\Gamma)}$ ;  $z_{jk}$  solves (54) with  $\epsilon^{(n)}$  and  $\zeta_{jk}$ 
      END
       $\delta\phi^{(n)}(x) = \sum_{j=1}^p \delta\phi_{jk}(x)$ ;
       $\phi^{(n+1)} = C_{LS}^{(n)}(\phi^{(n)} + \eta \frac{C_\rho(\Gamma)}{c_1} \delta\phi^{(n)})$ ;      update level set function
      Optional step: 'curve shortening' by diffusion. See separate chart.
       $\epsilon^{(n+1)} = \epsilon_b + \Lambda(\phi^{(n+1)})$ ;  $n = n + 1$ ;      Reinitialization  $n \rightarrow n + 1$ 
    END
  END
  alternatively, some stopping criteria can be used here
END
END
 $(\Omega^{(N)}, \epsilon^{(N)}, \phi^{(N)}) = (\Omega[\phi^{(n)}], \epsilon_b + \Lambda(\phi^{(n)}), \phi^{(n)})$ ;      Final reconstruction.

```

Here, η is a relaxation parameter for the update of the level set function which is determined empirically. The constant $C_\rho(\Gamma)$ could be calculated explicitly for the actual curve $\Gamma^{(n)}$, or it could be approximated by some value corresponding to a simple geometrical object (to give an example, in case of a single circle it would be $C_\rho(\Gamma) = (2\rho)^{-1}$). In our numerical experiments so far, however, it is simply considered as part of η . The same holds true for c_1 . The constant ρ is in our numerical experiments chosen between 30-40 cm, which corresponds to 2-3 grid cells. The scaling factor $C_{\text{LS}}^{(n)}$ is determined after each update to keep the global minimum (or maximum) of the level set function at a constant value.

The following smoothing filter on the level set function is optional. We usually apply it when we use noisy high-frequency data for the reconstruction. Especially in the limited-view examples presented in sections 6.3 and 6.4, the application of this filter improves the reconstructions significantly. Smoothing the level set function with this filter has the effect of local curve shortening. Roughness and small scale oscillations in the reconstructed boundaries are smoothed out such that the 'energy' of the reconstructed boundaries is reduced. The filtering step can be described as follows.

Optional step: 'Curve shortening' by diffusion.

Introduce artificial time $\tau \in \mathbb{R}$. $\frac{\partial}{\partial \tau}$ = time derivative, Δ = Laplace operator.

Solve initial value problem (with absorbing boundary conditions) on $B_r(0)$:

$$\begin{aligned}\tilde{\phi}(x, 0) &= \phi^{(n+1)}(x); \\ \frac{\partial}{\partial \tau} \tilde{\phi}(x, \tau) &= \Delta \tilde{\phi}(x, \tau), \quad \tau \in [0, T_\lambda], \quad x \in B_r(0). \\ \phi^{(n+1)}(x) &= \tilde{\phi}(x, T_\lambda),\end{aligned}$$

with regularization parameter T_λ .

5 Step 1: A Source-Type Adjoint Field method

For starting our shape reconstruction method using level sets we will need an initial guess $(\Omega^{(0)}, \epsilon^{(0)}, \phi^{(0)})$.

Although it is possible just to create an arbitrary initial guess without using any data at all, we believe that it is important for the efficiency and the robustness of such a method to start it with a *good* initial guess. There are several reasons for this. First, when deriving the shape reconstruction method (which we will call for short 'levelART' in the following) we used a perturbation approach which is strictly

justified only when we already have a good first guess available. Moreover, we see in our numerical experiments that starting with a good initial guess stabilizes the shape reconstruction routine, in particular in cases where the data are incomplete and noisy. In addition, finding a good first guess reduces the amount of work which has to be done by the levelART routine itself, such that in the end the combined code will converge much more rapidly than levelART alone would do without a good initialization.

In this section, we will present the second key point of our combined inversion scheme, namely a *fast*, *inexpensive* and *stable* method for finding a very good first guess $(\Omega^{(0)}, \epsilon^{(0)}, \phi^{(0)})$ for levelART. This method is designed to share basic features with the levelART algorithm, such that it can be implemented in addition to levelART with almost no extra cost.

5.1 Source-type methods

In the framework of inverse scattering problems, the method we propose here can be considered as a 'source-type inversion method'. Roughly speaking, the general idea of source-type reconstruction methods in inverse scattering is to split a given nonlinear inverse scattering problem into two subproblems. The first one is linear, and tries to recover a virtual 'equivalent source' in the medium which would be able to fit the data if applied with the known background distribution. This equivalent source is related to the unknown scattering potential by a nonlinear 'constitutive' relation. Therefore, in the second part of the algorithm, a nonlinear inverse problem has to be solved to derive the scattering potential from the recovered equivalent source distribution.

This idea is not at all new. It has been applied for example in the Source-Type Integral Equation (STIE) method of Habashy *et al* [21], or in the method presented by Chew *et al.* in [9]. More recently, similar ideas have been applied by Abdullah *et al.* [1], Caorsi *et al.* [6], and van den Berg *et al.* [46, 47].

All of these approaches have in common that they use the source-type method as a stand-alone inversion scheme. Such a method has the advantage that it is not as sensitive to strong nonlinearities in the inverse problem as for example perturbation methods or the Born or Rytov approximation are [15, 20, 27].

On the other hand, interpreting the inverse scattering problem as an inverse source problem is not without drawbacks. For example, the existence of so-called 'non-radiating sources' or 'invisible sources' gives rise to a nonuniqueness in the inverse source problem, which is difficult to deal with when solving the nonlinear part [1, 14, 21]. Moreover, it is not clear at all how to combine properly the information corresponding to different experiments, since each experiment creates its own 'equiv-

alent sources' and its own 'invisible sources'. For more information about possible applications, advantages and drawbacks of the source-type scheme as a stand-alone inversion tool we refer to [1, 5, 6, 9, 14, 21, 46, 47].

Our approach is different from those mentioned above. We only want to find a good *approximation* to the scattering potential, and a corresponding *initial level set function* suitable to start the shape reconstruction routine. Moreover, we can make use of our prior information about the permittivity distribution. This will allow us to circumvent most of the problems of source-type schemes which have been mentioned above.

We will now describe this method, which we will call the Source-Type Adjoint Field (STAF) method, in more details.

5.2 Solving the inverse scattering problem

Consider the inverse scattering problem formulated in section 3.1. The decomposition (13), (14) reads

$$\tilde{\epsilon}(x) = \epsilon_b(x) + \tilde{\epsilon}_s(x) \quad (70)$$

with some (known) background distribution ϵ_b and the (unknown) scattering potential $\tilde{\epsilon}_s$ having compact support, $\text{supp}(\tilde{\epsilon}_s) \subset \subset B_r(0)$. The goal is to find $\tilde{\epsilon}_s(x)$ from the data (9).

We already mentioned above that we actually will *not* recover the entire function $\tilde{\epsilon}_s(x)$ from the data \tilde{G}_{jk} in this preprocessing step. All we will find is 1.) A very good first guess for the scattering potential $\epsilon_s^{(0)}$ which will be part of the initializing triple $(\Omega^{(0)}, \epsilon_b + \epsilon_s^{(0)}, \phi^{(0)})$, and 2.) A corresponding level set function $\phi^{(0)}$. We will freely make use of the prior information resulting from the knowledge of $\hat{\epsilon}$ inside the estimated scatterer $\Omega^{(0)}$. However, our numerical results so far indicate that the method proposed here -if suitably adapted- can actually be used to find, in addition to $\Omega^{(0)}$ and $\phi^{(0)}$, also a good first estimate $\hat{\epsilon}^{(0)}$ for the contrast $\hat{\epsilon}$. This will be used in our future work to start a reconstruction method which tries to recover $\tilde{\Omega}$ and $\hat{\epsilon}$ *simultaneously* from the given data.

For a fixed frequency f_k and a source q_j , let \tilde{u}_{jk} be the solution of

$$\Delta \tilde{u}_{jk} + (a_k(\epsilon_b + \tilde{\epsilon}_s)(x) + ib\sigma(x)) \tilde{u}_{jk} = q_j(x), \quad (71)$$

and let u_{jk} be the solution of the 'unperturbed' equation

$$\Delta u_{jk} + (a_k \epsilon_b(x) + ib\sigma(x)) u_{jk} = q_j(x). \quad (72)$$

Define

$$\tilde{v}_{jk} := u_{jk} - \tilde{u}_{jk}. \quad (73)$$

Subtraction of (72) from (71) shows that \tilde{v}_{jk} solves

$$\Delta \tilde{v}_{jk} + (a_k \epsilon_b(x) + ib\sigma(x)) \tilde{v}_{jk} = \tilde{Q}_{jk}^s(x), \quad (74)$$

where the 'scattering source' $\tilde{Q}_{jk}^s(x)$ is defined as

$$\tilde{Q}_{jk}^s(x) = a_k \tilde{\epsilon}_s(x) \tilde{u}_{jk}(x). \quad (75)$$

We introduce a 'source type' forward operator A_{jk} by putting

$$A_{jk} : Y \longrightarrow Z_j \quad , \quad A_{jk} Q_{jk}^s = M_j v_{jk} \quad (76)$$

where M_j is the measurement operator defined in (8), and v_{jk} solves

$$\Delta v_{jk} + (a_k \epsilon_b(x) + ib\sigma(x)) v_{jk} = Q_{jk}^s(x). \quad (77)$$

The operator A_{jk} is linear.

Let us assume now that we apply the 'correct' scattering source $\tilde{Q}_{jk}^s(x)$ defined by (75) as argument of A_{jk} . Then we know from (9), (29), (30) that

$$A_{jk} \tilde{Q}_{jk} = M_j \tilde{v}_{jk} = M_j (u_{jk} - \tilde{u}_{jk}) = M_j u_{jk} - \tilde{G}_{jk} = R_{jk}(0). \quad (78)$$

The vectors $R_{jk}(0)$ are easily computed by solving a forward problem on the background distribution (72). Therefore, all we have to do to get back the scattering source \tilde{Q}_{jk}^s from the data \tilde{G}_{jk} is to solve (78) for \tilde{Q}_{jk}^s . Doing so amounts to solving an ill-posed but *linear* inverse problem.

Once we have recovered $\tilde{Q}_{jk}^s(x)$, we want to get back $\tilde{\epsilon}_s(x)$ out of it by using the constitutive relation (75). This second part of the inversion scheme can be interpreted as solving a *nonlinear* inverse problem since $\tilde{u}_{jk}(x)$ depends on $\tilde{\epsilon}_s(x)$.

Notice that $\tilde{Q}_{jk}^s(x)$ varies with different sources and frequencies, but that $\tilde{\epsilon}_s(x)$ is the same for all sources and all frequencies (if we neglect dispersion). We will make use of this observation when we try to solve the nonlinear part (75). In the following, we describe the method which we will use to recover the scattering source $\tilde{Q}_{jk}^s(x)$ from a given data set \tilde{G}_{jk} for a fixed source q_j and a fixed frequency f_k .

5.3 Looking for a scattering source

Since for a fixed (primary) source position and a fixed frequency we have only few data given to recover \tilde{Q}_{jk}^s , and since we have to take into account that also 'non-radiating' and 'invisible' sources have been generated in the experiment, we assume that there will be many solutions (in absence of noise) of (78). To pick one we are looking for the solution with minimal norm

$$\text{Min } \|Q_{jk}^s\|_Y \quad \text{subject to} \quad A_{jk} Q_{jk}^s = R_{jk}(0). \quad (79)$$

It is given by

$$Q_{jk,MN}^s = A_{jk}^* \left(A_{jk} A_{jk}^* \right)^{-1} R_{jk}(0), \quad (80)$$

where A_{jk}^* denotes the adjoint operator to A_{jk} .

The following theorem, which is proven in the appendix, tells us how to calculate the action of A_{jk}^* on a vector $\zeta \in Z_j$ in an efficient way.

Theorem 5.1 *Let $\zeta = (\zeta_1, \dots, \zeta_{D_j})^T \in Z_j$ and let x_{jd} , $d = 1, \dots, D_j$ be the detector positions corresponding to the source q_j . Then, $A_{jk}^* \zeta$ is given by*

$$A_{jk}^* \zeta = \overline{z_{jk}} \chi_{B_r(0)}, \quad (81)$$

where z_{jk} solves

$$\Delta z_{jk} + (a_k \epsilon_b + i b_k \sigma) z_{jk} = \sum_{d=1}^{D_j} \overline{\zeta_d} \delta(x - x_{jd}). \quad (82)$$

Corollary 5.1 *Let $\varphi_{\nu k}$ solve*

$$\Delta \varphi_{\nu k} + (a_k \epsilon_b + i b_k \sigma) \varphi_{\nu k} = \delta(x - x_{j\nu}). \quad (83)$$

Then, we can write (81) in the alternative form

$$\left(A_{jk}^* \zeta \right) (x) = \overline{\sum_{d=1}^{D_j} \zeta_d \varphi_{dk}(x) \chi_{B_r(0)}(x)} = \sum_{d=1}^{D_j} \zeta_d \overline{\varphi_{dk}(x) \chi_{B_r(0)}(x)}. \quad (84)$$

The next theorem, which is proven in the appendix, gives an explicit expression for the operators $A_{jk} A_{jk}^*$.

Theorem 5.2 *Let us assume that we are given a background permittivity distribution ϵ_b and a finite set of receiver positions x_{jd} , $d = 1, \dots, D_j$. The operators $A_{jk} A_{jk}^*$, $j = 1, \dots, p$, $k = 1, \dots, K$, are then described by $D_j \times D_j$ matrices of the form*

$$A_{jk} A_{jk}^* = \left\{ \int_{B_r(0)} \overline{\varphi_{dk}(x)} \varphi_{lk}(x) \right\}_{\substack{l=1, \dots, D_j \\ d=1, \dots, D_j}}, \quad (85)$$

where $\varphi_{\nu k}$ solves (83) and the index ν can stand for a receiver index d or l . In (85), the receiver index l is the row index, and the receiver index d is the column index of \hat{C}_{jk} .

Remark. Notice that (85) does not depend on the sources q_j , but only on the arrangement of the detectors and on the background permittivity distribution ϵ_b . The operators $A_{jk} A_{jk}^*$ can be precomputed before starting the inversion routine,

and this has to be done only once for each frequency as long as we use the same arrangement of detectors for all sources q_j , $j = 1, \dots, p$.

In the case of noisy data we will invert $A_{jk}A_{jk}^* + \lambda I$ instead of $A_{jk}A_{jk}^*$ in (80) with a suitably chosen regularization parameter $\lambda > 0$. This amounts to applying Tychonov-Phillips regularization.

5.4 Recovery of the scattering potential

After we have found a scattering source Q_{jk}^s which satisfies (79), we want to use the constitutive relation

$$Q_{jk}^s(x) = a_k \tilde{\epsilon}_s(x) \tilde{u}_{jk}(x), \quad (86)$$

which holds for the 'correct' scattering source \tilde{Q}_{jk}^s according to (75), to find an approximation for $\tilde{\epsilon}_s(x)$.

Let \tilde{u}_{jk} be a solution of (71) and u_{jk} a solution of (72). We decompose Q_{jk}^s , \tilde{u}_{jk} , u_{jk} and $\tilde{\epsilon}_s$ into amplitude and phase

$$Q_{jk}^s(x) = |Q_{jk}^s(x)| e^{ir(x)}, \quad \tilde{u}_{jk} = |\tilde{u}_{jk}| e^{i\tilde{s}(x)}, \quad (87)$$

$$u_{jk} = |u_{jk}| e^{is(x)}, \quad \epsilon_s(x) = |\epsilon_s(x)| e^{it(x)}, \quad (88)$$

where we have omitted the subscripts jk in the argument functions r , \tilde{s} , s , and t for simplicity in the notation. Making use of the fact that $\tilde{\epsilon}_s(x) \in \mathbb{R}$ we see that

$$t(x) \in \{0, \pi\} \quad \text{for all } x \in \mathbb{R}^2. \quad (89)$$

With (87),(88) equation (79) decomposes into two equations, one for the amplitude and one for the phase. They are

$$|Q_{jk}^s(x)| = a_k |\tilde{\epsilon}_s(x)| |\tilde{u}_{jk}|, \quad (90)$$

$$r(x) = \tilde{s}(x) + t(x). \quad (91)$$

The observation in our numerical experiments is that, although $s(x)$ and $\tilde{s}(x)$ might be quite different from each other for large perturbations $\tilde{\epsilon}_s(x)$, the amplitudes $|u_{jk}(x)|$ and $|\tilde{u}_{jk}(x)|$ most often do not differ too much from each other in the scattering region. Therefore, in our applications it is a reasonable approximation to assume that

$$|\tilde{u}_{jk}(x)| \approx |u_{jk}(x)| \quad \text{in } B_r(0). \quad (92)$$

With this approximation, (90) yields the following estimate for $|\tilde{\epsilon}_s(x)|$

$$|\tilde{\epsilon}_s^{(jk)}(x)| \approx \frac{|Q_{jk}^s(x)|}{a_k |u_{jk}(x)|} \quad \text{in } B_r(0). \quad (93)$$

We have added the indices j and k on the left hand side of (93) to indicate that we have used only the data \tilde{G}_{jk} corresponding to source q_j and frequency f_k for its determination.

Notice that the step (92), (93) is *nonlinear* since taking the amplitude of a complex number is a nonlinear operation. Therefore, the approach presented here is quite different from the usual Born approximation which approximates \tilde{u}_{jk} by u_{jk} .

For the purposes of the present paper, the determination of $|\tilde{\epsilon}_s(x)|$ is already sufficient in order to get a good first guess for the scattering potential $\epsilon_s^{(0)}(x)$ and for the level set function $\phi^{(0)}(x)$, since we can now make use of our prior information about the correct value of $\hat{\epsilon}$ in (12).

We mention, however, that the recovery of the phase $t(x)$ is also possible from (86). (This will be necessary for example when we try to recover $\tilde{\Omega}$ and $\hat{\epsilon}$ simultaneously from the given set of data \tilde{G}_{jk} .) We can do this by using equation (91). We already mentioned that the assumption $\tilde{s}(x) \approx s(x)$ might be quite wrong for situations with large scattering potentials $\tilde{\epsilon}_s(x)$. However, in our situation we only have to decide whether $t(x)$ is zero or π , which means that we have to determine whether in (91) $r(x) = \tilde{s}(x)$ or $r(x) = \tilde{s}(x) \pm \pi$ is satisfied. Therefore, a reasonable estimate for $t(x)$ is to put

$$t^{(jk)}(x) = \begin{cases} 0 & , \quad |s(x) - r(x)| < \pi/2, \\ \pi & , \quad \text{elsewhere.} \end{cases} \quad (94)$$

Our numerical experiments so far show that a suitable combination of these estimates resulting from many source positions gives a very good reconstruction of the phase $t(x)$ in $B_r(0)$ even in the situation of limited view and noisy data. We will not need this estimate in the present paper.

5.5 Combining the results from single experiments

We can combine now the estimates $|\tilde{\epsilon}_s^{(jk)}(x)|$ from many source positions q_j , $j = 1, \dots, p$, by putting

$$|\tilde{\epsilon}_s^{(k)}(x)| = \frac{1}{p} \sum_{j=1}^p |\tilde{\epsilon}_s^{(jk)}(x)|. \quad (95)$$

If we want to take into account also the information corresponding to many frequencies we can do so by putting

$$|\tilde{\epsilon}_s(x)| \approx \frac{1}{pK} \sum_{k=1}^K \sum_{j=1}^p |\tilde{\epsilon}_s^{(jk)}(x)| \quad (96)$$

where the sum is over all frequencies f_k , $k = 1, \dots, K$, and all source positions q_j , $j = 1, \dots, p$. In (96) we have neglected dispersion.

A similar strategy can be employed to improve the estimates for the phase $t^{(jk)}(x)$.

5.6 The initial level set function

We are now ready to define the initial triple $(\Omega^{(0)}, \epsilon^{(0)}, \phi^{(0)})$.

We assume that we are working in a high contrast situation, such that exactly one of the following conditions is satisfied

$$\hat{\epsilon} - \epsilon_b(x) \gg 0 \quad \text{for all } x \in \tilde{\Omega} \quad (97)$$

$$\hat{\epsilon} - \epsilon_b(x) \ll 0 \quad \text{for all } x \in \tilde{\Omega}. \quad (98)$$

Since we know $\hat{\epsilon}$ and $\epsilon_b(x)$, we know the constant

$$\text{sign}(\tilde{\Omega}) := \begin{cases} 1 & , \quad \text{if (97) holds} \\ -1 & , \quad \text{if (98) holds.} \end{cases} \quad (99)$$

Let us assume that we want to use $|\tilde{\epsilon}_s^{(k)}(x)|$ as defined in (95) for a fixed frequency f_k to specify the level set function $\phi^{(0)}$. Choose a threshold value $0 < \gamma_{\text{LS}} < 1$ (in our numerical examples presented in section 6 we use $\gamma_{\text{LS}} = 0.7$) and define

$$\epsilon_{\text{LS}} := \gamma_{\text{LS}} \max_{x \in B_r(0)} |\tilde{\epsilon}_s^{(k)}(x)|. \quad (100)$$

For the level set zero $L_0^{(0)}$ of $\phi^{(0)}$ we require that

$$L_0^{(0)} = \left\{ x \in B_r(0) : |\tilde{\epsilon}_s^{(k)}(x)| = \epsilon_{\text{LS}} \right\}. \quad (101)$$

This means that we want all points of $B_r(0)$ where the reconstruction $|\tilde{\epsilon}_s^{(k)}(x)|$ has exactly the value ϵ_{LS} to be mapped to zero by the level set function $\phi^{(0)}$

$$\phi^{(0)}(x) = 0 \quad \text{for all } x \in L_0^{(0)}. \quad (102)$$

The level set function is now defined as

$$\phi^{(0)}(x) = C_{\text{LS}}^{(0)} \text{sign}(\tilde{\Omega}) \left(\epsilon_{\text{LS}} - |\tilde{\epsilon}_s^{(k)}(x)| \right), \quad (103)$$

where $C_{\text{LS}}^{(0)}$ is some suitably chosen scaling factor. Notice that (102) and (103) are consistent.

The initial scattering domain $\Omega^{(0)}$ and the permittivity $\epsilon^{(0)}$ are defined as

$$\Omega^{(0)} = \Omega[\phi^{(0)}], \quad \epsilon^{(0)} = \epsilon_b + \Lambda(\phi^{(0)}). \quad (104)$$

Together with $\phi^{(0)}$ they form an admissible triple $(\Omega^{(0)}, \epsilon^{(0)}, \phi^{(0)})$.

5.7 Implementation: The STAF algorithm

In brief algorithmic form, the Source Type Adjoint Field (STAF) scheme can be written in the following way.

Preparation step.

- Select a group of frequencies $G_s = \{f_1, \dots, f_{K_s}\}$ which are used for the STAF reconstruction. Typically, this is just one frequency.
- Calculate $A_{jk}A_{jk}^*$ according to (85) for each frequency $f \in G_s$. The operator $A_{jk}A_{jk}^*$ does not depend on the index j if we use the same detector positions for all sources.
- Calculate

$$B_{jk} = \left(A_{jk}A_{jk}^*\right)^{-1} \quad \text{or} \quad B_{jk} = \left(A_{jk}A_{jk}^* + \lambda I\right)^{-1} \quad (105)$$

for all $f \in G_s$ and store in memory for later use.

Reconstruction step.

FOR $k = 1 : K_s$

FOR $j = 1 : p$

$$R_{jk}(0) = M_j u_{jk} - \tilde{G}_{jk}, \quad u_{jk} \text{ solves (72)}$$

$$\zeta_{jk} = B_{jk} R_{jk}(0), \quad B_{jk} \text{ from (105)}$$

$$Q_{jk}^s = A_{jk}^* \zeta_{jk} = \overline{z_{jk}} \chi_{B_r(0)}, \quad z_{jk} \text{ solves (82)}$$

$$|\tilde{\epsilon}_s^{(jk)}(x)| = \frac{|Q_{jk}^s(x)|}{a_k |u_{jk}(x)|}$$

END

END

$$|\tilde{\epsilon}_s(x)| = \frac{1}{pK_s} \sum_{k=1}^{K_s} \sum_{j=1}^p |\tilde{\epsilon}_s^{(jk)}(x)|, \quad x \in B_r(0)$$

$$\phi^{(0)}(x) = C_{\text{LS}}^{(0)} \text{sign}(\tilde{\Omega})(\epsilon_{\text{LS}} - |\tilde{\epsilon}_s(x)|) \quad \text{as in (103)}$$

$$(\Omega^{(0)}, \epsilon^{(0)}, \phi^{(0)}) = (\Omega[\phi^{(0)}], \epsilon_b + \Lambda(\phi^{(0)}), \phi^{(0)}).$$

6 Numerical Experiments

6.1 Discretization of the computational domain.

In our numerical experiments, we use a Finite-Differences Frequency Domain (FDFD) code written in MATLAB for solving (1)-(3). The code uses appropriately designed

perfectly matched layers (PML) to avoid reflections at the artificial computational boundaries [36, 37].

The system which results after discretization is solved by a simple Gauss elimination scheme, which is reflected in the implementation shown in sections 4.10 and 5.7. The LR -factorization corresponding to the most recent best guess is used there to calculate the fields for all transmitters and all receivers simultaneously. Therefore, the computational cost for solving all necessary forward and adjoint problems is just one LR -factorization for STAF, and one LR -factorization per update for levelART. If a different solver is used (e.g. GMRES or QMR), then we might find more efficient strategies than those presented in sections 4.10 and 5.7. We mention also that an iterative solver has been developed recently in [25] which is optimized to work on several source distributions simultaneously.

The physical domain is partitioned into 100×100 elementary cells (pixels) in the first numerical example, and into 180×110 elementary cells in the second and third example. Each of these grid cells has a physical size of about $0.14 \times 0.14 \text{ m}^2$, such that the total computational domain in the first example covers an area of $14 \times 14 \text{ m}^2$, and in the other two examples of $15 \times 25 \text{ m}^2$. The eight layers which are closest to the boundaries of the computational domain are used as a PML.

We will refer to the first numerical example as the 'full-view' situation, and to the other two numerical examples as the 'limited-view', 'cross-borehole' or 'geophysical' situations. This terminology is motivated by the source and receiver geometries used, which are as follows.

In the full-view example, we have 64 sources and receivers given which surround the domain of interest. Each source position is at the same time a receiver position and vice versa. The distance of two adjacent sources or receivers from each other is four pixels or about 55 cm. The area enclosed by these sources and receivers has a size of $10 \times 10 \text{ m}^2$.

In the two limited-view examples, 74 sources and receivers are positioned equally spaced in two boreholes. The distance of two adjacent sources or receivers from each other is again 4 pixels or 55 cm, and the distance of the two boreholes from each other is about 10 m.

We mention that, in all of our numerical examples, the regions beyond the source and receiver positions are part of the inversion problem, too. This means, the area which has to be recovered from the data is the whole area situated between the PML boundaries. In some of our numerical experiments, artifacts can be observed developping in the outer areas during the early stages of the reconstruction process.

We apply time-harmonic dipole sources of the form (11) with frequencies of $f = 5, 10, 15, 20, 25$, or 30 MHz. In our examples, this corresponds to wavelengths

in the background medium between 2 meters for $f = 30$ MHz and 13 meters for $f = 5$ MHz. The size of an individual grid cell is chosen such that each of these wavelengths is sampled by at least 16 pixels in order to avoid numerical artifacts due to undersampling.

The data in our numerical examples are generated by running the FDFD forward modelling code on the correct permittivity and conductivity distributions. Using the same forward code for creating the data and for doing the reconstruction is usually called 'inverse crime'. Therefore, to make sure that the situations we model in our experiments are as realistic as possible, we have tested the forward modelling code thoroughly, and add Gaussian noise with signal-to-noise ratios between 10 and 5 dB to the real and imaginary parts of the generated data.

6.2 A full-view example

Our first numerical example tests whether the derived algorithm is able to reconstruct a relatively complicated shape in the ideal situation where sources and receivers completely surround the area of interest. The geometry of this example is shown in Figure 1. The positions of the sources and receivers are indicated by dots in the Figure. The background medium in this example consists of a homogeneous conductivity distribution $\sigma_b = 3.0 \times 10^{-4}$ Siemens/m, and a homogeneous permittivity distribution $\epsilon_b = 20$. Inside the object, the permittivity is $\hat{\epsilon} = 15$, having a moderate contrast to the background distribution.

Notice that an interesting feature of this geometry is the 'hole' in the body of the object which is difficult to reconstruct. We will see that, during the evolution of the permittivity in levelART, the boundaries of the reconstructed domain will split and merge in the attempt to recover this geometry correctly.

First, we test the STAF algorithm in Figure 2 by reconstructing the permittivity $\epsilon^{(0)}$ using only the data $G_s = \{30 \text{ MHz}\}$. These data are noisy with a signal-to-noise ratio (SNR) of 10 dB in the real and the imaginary parts. Compare the result with the upper left image of Figure 3 where we used (noise-free) data with frequency $G_s = \{5 \text{ MHz}\}$.

A comparison of reconstructed permittivities for different frequencies between 5 and 30 MHz shows that -in the ideal situation of sources and receivers completely surrounding the area of interest- the STAF algorithm usually yields already a decent approximation to the shape of the inclusion when applied to the data with the highest frequencies, whereas it yields a reconstruction with decreasing resolution when applied to data corresponding to lower frequencies.

Therefore, it seems to be most efficient to apply the STAF algorithm directly to the highest frequency data, such that we do not need at all any low-frequency

information for the reconstruction. We will see in the following two geophysical examples that this is certainly not true in applications where we have only data with limited view available. In these cases, the use of lower frequency data stabilizes the reconstruction process, and is necessary for preparing the final reconstruction step using the higher frequency data.

We also want to demonstrate the performance of the levelART reconstruction scheme when applied to this geometry. We start the algorithm by using as initial permittivity $\epsilon^{(0)}$ the low resolution STAF reconstruction which is shown in the upper left image of Figure 3, and the corresponding initial level set function $\phi^{(0)}$. No noise is added to the data. Figure 3 shows different stages in the reconstruction process. We first apply levelART with a frequency of 10 MHz to this initial guess and run it for 30 sweeps. Then, we run levelART with 20 MHz for 30 more sweeps, and finally for another 30 sweeps with 30 MHz. The final reconstruction $\epsilon^{(N)}$ is shown in the lower right image of Figure 3. The final level set function $\phi^{(N)}$ corresponding to this reconstruction is displayed in Figure 14.

We see from this example that the shape reconstruction algorithm using level sets is able to split and merge boundaries easily in order to build up relatively complicated geometries. In the present situation, splitting and merging of boundaries was necessary for building the 'hole' in the reconstructed domain.

6.3 A cross-borehole situation with multiple objects

In our second numerical example, we consider a situation which is typical for geophysical applications. Comparable situations occur for example when we wish to monitor pollutant plumes at environmental cleanup sites from cross-borehole EM data.

We assume that we have 74 sources and receivers equidistantly distributed over two boreholes. The distance of the boreholes from each other is 10 meters, and the distance of two adjacent sources or receivers is 55 cm. The area between the two boreholes has to be monitored given the gathered data. The geometry is shown in Figure 4.

The background permittivity distribution in this example consists of four tilted layers with values of $\epsilon_b = 21$ in the top layer, and then continuing downwards with 20, 19, and again 21 for the deepest layer. The conductivity distribution σ_b is homogeneous with a value of $\sigma_b = 3.0 \times 10^{-4}$ S/m everywhere.

Embedded in this background are three compact inclusions as shown in Figure 4. The permittivity inside these inclusions is $\hat{\epsilon} = 5$, having a high contrast to the background values. The three inclusions are oriented such that there are two 'channels' of background material between them, one of them in the vertical and

one in the horizontal direction. The difficulty in this example is to separate the three inclusions from each other from the limited-view data. In particular, the reconstruction of the vertical channel is critical since we expect that the resolution in the horizontal direction will suffer from the missing data.

Again, we first apply the STAF reconstruction scheme to the data to get a first guess of the permittivity distribution $\epsilon^{(0)}$. Figure 5 shows the result for $G_s = \{30 \text{ MHz}\}$. Comparing this result with the reconstruction for the same frequency in our first numerical example, we conclude that the performance of STAF for high-frequency data is in the limited-view case not as good as in the case where we can use data with full view. We observe that the vertical resolution of the reconstruction is still good, whereas in the horizontal direction severe artifacts build up which reduce the quality of the high-frequency STAF reconstructions as an initial guess for levelART.

Figure 6 shows the corresponding STAF reconstruction using $G_s = \{5 \text{ MHz}\}$. We do not observe any artifacts in this reconstruction which might be caused by the limited view in the data. Therefore, we see that the decreased resolution of STAF using low-frequency data is in this situation compensated by a much higher robustness with respect to missing data. Keeping this in mind, we will typically start our reconstructions in the limited-view geometry by using the STAF result which correspond to (one or more of) the lowest available frequencies.

Starting out from the permittivity $\epsilon^{(0)}$ as shown in Figure 6, and the corresponding level set function $\phi^{(0)}$, we want to use the levelART algorithm in order to calculate a series of shape deformations which transforms the initial shape into the correct permittivity distribution. Figure 6 shows a reconstruction which uses data where the real part and the imaginary part have been contaminated by additive Gaussian noise with a signal-to-noise ratio (SNR) of 10 dB before starting the reconstruction process. Figure 7 shows the same reconstruction scheme, but with an even lower SNR of now 5 dB in the data.

Different strategies are possible for levelART. Which one works best, depends on the specific situation, for example the number and arrangement of sources and detectors, and on the noise level of the data. The strategy we use here (for both SNR values) is the following: First, we apply 20 steps of levelART with a frequency of 15 MHz to the initial guess. The result is shown in the lower left images of Figures 6 and 7. At this stage, the task of splitting the initial object into three subsets is almost completed.

Then, we apply levelART with a combination of three frequencies, namely 20, 25, and 30 MHz. This means that in a given sweep each of these three frequencies is used exactly once in the prescribed order, before starting again with the lowest frequency

(20 MHz) for the new sweep. This final step completes the reconstruction in just a few (about 10) sweeps. The succeeding sweeps do not improve the reconstruction significantly. Moreover, the norms of the residuals approach a constant value as the Figures 8 to 11 show.

Figures 8 to 11 show the evolution of the norms of the residuals during the reconstruction process for different signal-to-noise ratios in the data. The graphs with the symbol 'o' correspond to noiseless data, the graphs for a SNR of 10 dB are indicated by '*', and those for a SNR of 5 dB by '+'.

Figure 8 shows the norms of the residuals for $f = 15$ MHz during the 20 steps using the data with this frequency. The norms of the residuals decrease in all cases continually during the reconstruction process. Figures 9 to 11 show the evolution of the norms during the final 30 sweeps with the frequencies 20, 25 and 30 MHz. We see that after 10 sweeps the residuals approach some constant value which depends on the noise level of the data. In the lower right image of Figure 6, the reconstruction for a signal-to-noise ratio of 10 dB after completion of these 10 sweeps is shown. The corresponding level set function is shown in Figure 15.

Our experience is that marching over the higher frequencies in the described fashion stabilizes the inversion procedure especially in the limited-view situation considered here. However, so far we do not have any theoretical analysis which supports this observation.

Notice the artifacts which appear in the case of an extremely low SNR of 5 dB in Figure 7. These artifacts remain more or less stable when applying levelART with a constant frequency of 15 MHz, and disappear when marching to the higher frequencies in the succeeding reconstruction step. Notice that we also apply a diffusion ('curve shortening') filter for these higher frequencies, see section 4.10.

We observe again that levelART has no problems in propagating and tracking these multiple artifacts, even when they finally shrink and disappear. Notice also that, similar to the first numerical example, the algorithm splits the original boundary in order to arrive at the three separated inclusions forming the final reconstruction.

6.4 A cross-borehole situation with a single inclusion and variable conductivity

In our third numerical example, we want to test a situation where the conductivity inside the inclusions is different from the given background values. Since we did not take these conductivity changes into account when deriving the reconstruction algorithm, the question arises how much the performance of the reconstruction method will be effected by such changes in the conductivity distribution.

We assume therefore that we know the two values $\hat{\epsilon}$ and $\hat{\sigma}$ of the permittivity and the conductivity inside the inclusions, but both of them are different from the background values. We run the STAF routine as usual, which amounts to treating the conductivity changes simply as an additional form of noise. When applying levelART, we calculate the updates $\delta\phi^{(n)}$ for the level set function $\phi^{(n)}$ in the same way as derived above, but when determining the corresponding updated permittivity distribution $\epsilon^{(n)}$, we update at the same time the conductivity distribution $\sigma^{(n)}$ such that $\sigma(x) = \hat{\sigma}$ where the level set function $\phi^{(n)}$ has negative values. Strictly speaking, we still invert only for the permittivity ϵ , but we make use of the fact that inside the obstacles the conductivity and the permittivity are closely related to each other.

Figure 12 shows the geometry of this example, and Figure 13 shows the results for two different conductivity values $\hat{\sigma} = 1.0 \times 10^{-6}$ S/m and $\hat{\sigma} = 1.0 \times 10^{-2}$ S/m. Notice that these two values differ from each other and from the background value $\sigma_b = 3.0 \times 10^{-4}$ S/m by orders of magnitude! In both cases, Gaussian noise has been added to the real and imaginary parts of the data with a signal-to-noise ratio of 10 dB before starting the reconstruction routine.

As before, we start the reconstruction with the STAF guess corresponding to a frequency of 5 MHz. After only six sweeps of levelART, using the frequencies 15, 20, 25, and 30 MHz one after the other in each sweep, we arrive in both cases at a very good reconstruction of the permittivity distribution and of the conductivity distribution. We conclude that the performance of the reconstruction method (STAF and levelART) is not significantly effected by the changes in the conductivity distribution.

This robustness with respect to changes in the conductivity certainly has its limits. However, the example presented here makes us confident that in practical situations, when the conductivity value $\hat{\sigma}$ inside the obstacles is approximately known and not too large ($< 1.0 \times 10^{-2}$ S/m), the shape reconstruction method will perform stably and reliably and will give us a good reconstruction of the actual permittivity distribution.

7 Summary and future directions

We have presented a stable and efficient two-step shape reconstruction algorithm for EM cross-borehole tomography which uses adjoint fields and level sets. We have shown that this method is able to recover one or more objects with nontrivial shapes given noisy cross-borehole EM data.

The first step of this combined inversion scheme plays the role of an initializing procedure for the second step, and employs a 'source-type' inversion scheme to deal

with the high nonlinearity in the problem due to the presence of strong scatterers.

Although we believe that the preprocessing routine we propose here will work well in most situations, it can be replaced by any other preprocessing tool which shares the main features of the derived algorithm.

The second step of the inversion routine starts directly with the outcome of this initializing procedure, and continues by using a combination of an 'adjoint field technique' and a level set representation of the shapes until the inversion task is completed. We have shown that using a level set representation in this second step enables us to easily describe and keep track of complicated geometries which arise during the inversion process.

We mention that the FDFD routine, which has been employed in both steps to solve the forward and the adjoint Helmholtz problems, can be replaced by any other more efficient Helmholtz solver which has been tested to work reliably in the given situation.

We have shown in our numerical experiments that the proposed reconstruction scheme performs stably with respect to changes in the conductivity distribution, although these conductivity changes have not been taken into account for the derivation of the scheme. It would be desirable, however, to extend the reconstruction scheme to work simultaneously on the permittivity and the conductivity distribution. This seems to be possible, and we will address this problem in our future work.

We also assume that we know the permittivity values inside the obstacles, and that we only have to recover their shapes and their locations. In our future research, we will investigate the situation where both, the shapes and the permittivity values inside the obstacles, have to be recovered from the given data.

Throughout the paper, we have not clearly specified the degree of regularity which we require for the level set functions ϕ representing the domains Ω . A possible choice would be $\phi \in C_0^1(B_r(0))$ (i.e. continuously differentiable on $B_r(0)$ and zero on $\partial B_r(0)$), which would require some additional regularization in our numerical experiments. We also have introduced in (72) an approximated linearized operator $\Lambda'[\phi]$ motivated by our wish to use convenient inner products. Are there any function spaces which are more useful for our purposes? Will a practically useful reconstruction scheme result if we use different inner products instead of introducing $\Lambda'[\phi]$? To answer both questions, a thorough theoretical analysis has to be done to investigate the implications of using different function spaces for the level set representation.

The main ideas of the reconstruction method presented here are not restricted to a 2D geometry. Therefore, we believe that it is possible to extend the method to a more realistic 3D situation. All what is needed for this is an efficient forward

solver for the 3D system of Maxwell's equations. A forward solver which has been tested for such situations has been presented in [7, 18]. Moreover, applications to situations in medical imaging [2, 11], or in the nondestructive testing of materials [45], seem interesting and possible.

8 Acknowledgments

This work has been supported by a grant from the U. S. Dept. of Energy DE-FG07-97ID3566, by the Center for Subsurface Sensing and Imaging Systems at Northeastern University, by A CAREER Grant from the National Science Foundation MIP-9623721, and by The Army Research Office, Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013.

9 Appendix

9.1 Proof of theorem 5.1

Green's formula for an infinite domain (without boundary terms since $\sigma > 0$) reads for general v, z

$$\begin{aligned} \int_{\mathbb{R}^2} [\Delta v + \kappa_k v] z \, dx + \int_{\mathbb{R}^2} v(x) \left(\sum_{d=1}^{D_j} \overline{\zeta_d} \delta(x - x_{jd}) \right) dx \\ = \int_{\mathbb{R}^2} v [\Delta z + \kappa_k z] \, dx + \sum_{d=1}^{D_j} \left(\int_{\mathbb{R}^2} v(x) \delta(x - x_{jd}) dx \right) \overline{\zeta_d} \end{aligned} \quad (106)$$

where we have used the notation $\kappa_k = a_k \epsilon_b + i b_k \sigma$. Let now v_{jk} be a solution of (77), and z_{jk} a solution of (82). Then the first term on the left hand side of (106) reads

$$\int_{\mathbb{R}^2} Q_{jk}^s(x) z_{jk}(x) \, dx = \left\langle Q_{jk}^s, \overline{z_{jk}} \chi_{B_r(0)} \right\rangle_Y, \quad (107)$$

whereas the second term on the right hand side is

$$\langle M_j v_{jk}, \zeta \rangle_{Z_j} = \left\langle A_{jk} Q_{jk}^s, \zeta \right\rangle_{Z_j}. \quad (108)$$

The remaining two terms cancel each other because of (82). Therefore, (106) gets the form

$$\left\langle Q_{jk}^s, \overline{z_{jk}} \chi_{B_r(0)} \right\rangle_Y = \left\langle A_{jk} Q_{jk}^s, \zeta \right\rangle_{Z_j}, \quad (109)$$

which proves the theorem.

9.2 Proof of theorem 5.2

Since $A_{jk}A_{jk}^*$ is a linear operator acting on the finite-dimensional data space, it is sufficient to find the action of $A_{jk}A_{jk}^*$ on each of the basis vectors $e_d = (0, \dots, 0, 1, 0, \dots, 0)^T$, $d = 1, \dots, D_j$, where the '1' is at the d -th position. From (83) we see that

$$\left(A_{jk}^* e_d\right)(x) = \overline{\varphi_{dk}(x)} \chi_{B_r(0)}(x). \quad (110)$$

Application of A_{jk} yields for the l -th component ($l = 1, \dots, D_j$)

$$\left(A_{jk}A_{jk}^* e_d\right)_l = v_{jk}(x_{jl}), \quad (111)$$

where v_{jk} is given by

$$v_{jk}(y) = \int_{B_r(0)} G_k(y, x) \overline{\varphi_{dk}(x)} dx \quad (112)$$

and $G_k(y, x)$ is Green's function satisfying

$$\Delta G_k(y, x) + (a_k \epsilon_b + i b_k \sigma) G_k(y, x) = \delta(y - x). \quad (113)$$

Reciprocity yields $G_k(x_{jl}, x) = \varphi_{lk}(x)$. Therefore, we get from (111), (112)

$$\left(A_{jk}A_{jk}^* e_d\right)_l = \int_{B_r(0)} \varphi_{lk}(x) \overline{\varphi_{dk}(x)} dx, \quad (114)$$

which proves the theorem.

References

- [1] Abdullah H and Louis A K 1999 The approximate inverse for solving an inverse scattering problem for acoustic waves in an inhomogeneous medium *Inverse Problems* **15** 1213-1229
- [2] Albanese R A, Medina R L and Penn J W 1994 Mathematics, medicine and microwaves *Inverse Problems* **10** 995-1007
- [3] Alumbaugh D L and Morrison H F 1995 Theoretical and practical considerations for crosswell electromagnetic tomography assuming a cylindrical geometry *Geophysics* **60** (3) 846-870
- [4] Angell T S, Kleinman R E and Roach F G 1987 An inverse transmission problem for the Helmholtz equation *Inverse Problems* **3** 149-80
- [5] Angell T S, Jiang X and Kleinman R E 1997 A distributed source method for inverse acoustic scattering *Inverse Problems* **13** 531-46
- [6] Caorsi S and Gragnani G L 1999 Inverse-scattering method for dielectric objects based on the reconstruction of the nonmeasurable equivalent current density *Radio Science* **34** (1) 1-8

- [7] Champagne N J, II, Berryman J G, Buettner H M, Grant J B and Sharpe R M 1999 A finite-difference frequency-domain code for electromagnetic induction tomography, in the *Proceedings of SAGEEP*, Oakland, CA, March 14–18, 931–940
- [8] Chen Y (1997) Inverse scattering via Heisenberg’s uncertainty principle *Inverse Problems* **13** 253-282
- [9] Chew W C, Wang Y M, Otto G, Lesselier D and Bolomey J Ch 1994 On the inverse source method of solving inverse scattering problems *Inverse Problems* **10** 547-553
- [10] Claerbout J F 1976 *Fundamentals of Geophysical Data Processing: With Applications to Petroleum Prospecting* McGraw-Hill, New York
- [11] Colton D and Monk P 1994 The detection and monitoring of leukemia using electromagnetic waves: mathematical theory *Inverse Problems* **10** 1235-1251
- [12] Colton D L and Kress R 1998 *Inverse Acoustic and Electromagnetic Scattering Theory* 2nd edn (Berlin: Springer)
- [13] Davis J L and Annan A P 1989 Ground penetrating radar for high resolution mapping of soil and rock stratigraphy *Geophys. Prospect.* **37** 531-51
- [14] Devaney A J and Sherman G C 1984 Nonuniqueness in inverse source and scattering problems *IEEE Trans. Antennas Propag.* **30** 1034-1037
- [15] Devaney A J 1986 Reconstructive tomography with diffracting wavefields *Inverse Problems* **2** 161-83
- [16] Dines K A and Lytle R J 1979 Computerized geophysical tomography *Proc. IEEE* **67** 1065–1073
- [17] Dorn O 1998 A transport-backtransport method for optical tomography *Inverse Problems* **14** 1107–1130
- [18] Dorn O, Bertete-Aguirre H, Berryman J G and Papanicolaou G C 1999 A nonlinear inversion method for 3D electromagnetic imaging using adjoint fields *Inverse Problems* **15** 1523-1558
- [19] Fisher E, McMechan G A and Annan A P 1992 Acquisition and processing of wide-aperture ground penetrating radar data *Geophysics* **57** 495-504
- [20] Habashy T M, Groom R W, Spies B R 1993 Beyond the Born and Rytov Approximations: A Nonlinear Approach to Electromagnetic Scattering *Journal of Geophysical Research* **98** (B2) 1759-1775
- [21] Habashy T M, Oristaglio M L and de Hoop A T 1994 Simultaneous nonlinear reconstruction of two-dimensional permittivity and conductivity *Radio Science* **29** (4) 1101-1118
- [22] Hettlich F 1995 Frechet derivatives in inverse obstacle scattering *Inverse Problems* **11** 371-382

- [23] Hettlich F and Rundell W 1997 Recovery of the support of a source term in an elliptic differential equation *Inverse Problems* **13** 959-976
- [24] Kaup P G, Santosa F and Vogelius M 1996 Method of imaging corrosion damage in thin plates from electrostatic data *Inverse Problems* **12** 279-93
- [25] Kilmer M, Miller E L and Rappaport C 1999 QMR-Based Projection Techniques for the Solution of Non-Hermitian Systems with Multiple Right Hand Sides, submitted to *SIAM Journal on Scientific and Statistical Computing*
- [26] Kolehmainen V, Arridge S R, Lionheart W R B, Vauhkonen M and Kaipio J P 1999 Recovery of region boundaries of piecewise constant coefficients of an elliptic PDE from boundary data *Inverse Problems* **15** 1375-1391
- [27] Ladas K T and Devaney A J 1992 Iterative methods in geophysical diffraction tomography *Inverse Problems* **8** 119-132
- [28] Litman A, Lesselier D and Santosa F 1998 Reconstruction of a two-dimensional binary obstacle by controlled evolution of a level-set *Inverse Problems* **14** 685-706
- [29] MATLAB: High-performance numeric computation and visualization software - Reference Guide 1992 *MathWorks* Natick MA
- [30] Miller E, Kilmer M and Rappaport C 1999 A New Shape-Based Method for Object Localization and Characterization from Scattered Field Data, submitted to *IEEE Trans. Geoscience and Remote Sensing*
- [31] Nachman A 1996 Global uniqueness for a two-dimensional inverse boundary value problem *Ann. Math.* **143** 71-96
- [32] Natterer F 1986 *The Mathematics of Computerized Tomography* (Stuttgart: Teubner)
- [33] Natterer F and Wübbeling F 1995 A propagation-backpropagation method for ultrasound tomography *Inverse Problems* **11** 1225-1232
- [34] Natterer F 1996 Numerical Solution of Bilinear Inverse Problems *Preprints* "Angewandte Mathematik und Informatik" 19/96-N Münster
- [35] Osher S and Sethian J 1988 Fronts propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations *Journal of Computational Physics* **56** 12-49
- [36] Rappaport C M 1996 Interpreting and Improving the PML Absorbing Boundary Condition Using Anisotropic Lossy Mapping of Space *IEEE Transactions on Magnetism* **32** (3) 968-974
- [37] Rappaport C M, Kilmer M and Miller E 1999 Accuracy Considerations in Using the PML ABC with FDFD Helmholtz Equation Computation *Journal of Numerical Modelling* in press
- [38] Ramm A G 1986 *Scattering by obstacles* (Dordrecht: Reidel)

- [39] Rozier C, Lesselier D, Angell D and Kleinman R E 1996 Shape retrieval of an obstacle immersed in shallow water from single frequency fields using a complete family method *Inverse Problems* **13** 487-508
- [40] Santosa F 1996 A Level-Set Approach for Inverse Problems Involving Obstacles *ESAIM: Control, Optimization and Calculus of Variations* **1** 17-33
- [41] Sethian J A 1990 Numerical algorithms for propagating interfaces: Hamilton-Jacobi equations and conservation laws *J. Diff. Geom.* **31** 131-61
- [42] Sethian J A 1999 Level Set Methods and Fast Marching Methods (2nd ed) Cambridge University Press
- [43] Souriau L, B Duchene, Lesselier D and Kleinman R E 1996 Modified gradient approach to inverse scattering for binary objects in stratified media *Inverse Problems* **12** 463-481
- [44] Spies B P and Habashy T M 1995 Sensitivity analysis of crosswell electromagnetics *Geophysics* **60** 834-845
- [45] Special section on the conference on 'inverse problems, control and shape optimization' held 1998 in Carthage (Tunisia) 1999 *Inverse Problems* **15** (1) 1-134
- [46] Van den Berg P M and Kleinman R E 1997 A contrast source inversion method *Inverse Problems* **13** 1607-20
- [47] Van den Berg P M, van Broekhoven A L and Abubakar A 1999 Extended contrast source inversion *Inverse Problems* **15** 1325-1344
- [48] von Hippel A 1953 *Dielectric Materials and Applications* (Wiley: New York) 3-4
- [49] Ward S H and Hohmann G W 1987 Electromagnetic theory for geophysical applications, in *Electromagnetic Methods in Applied Geophysics – Theory*, Vol 1, Nabighian M N (ed.), SEG, Tulsa, Oklahoma 131-311
- [50] Wilt M J, Morrison H F, Becker A, Tseng H W, Lee K H, Torres-Verdin C, and Alumbaugh D 1995 Crosshole electromagnetic tomography: A new technology for oil field characterization, *The Leading Edge* **14** 173-177
- [51] Zhdanov M S, Traynin P and Booker J R 1996 Underground imaging by frequency-domain electromagnetic migration *Geophysics* **61**, 666-682

10 Figures

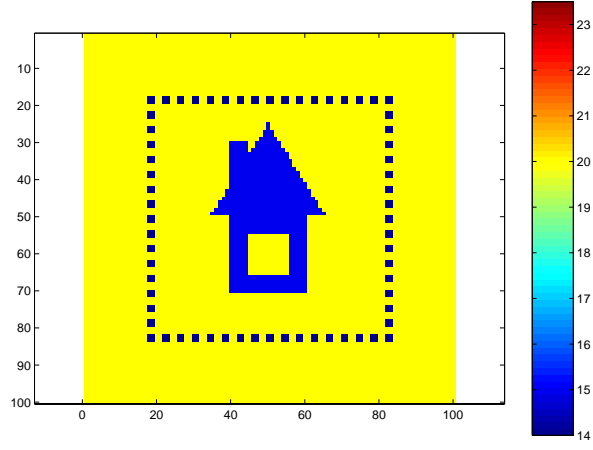


Figure 1: Original object for the example with full view. The dots in the figure indicate the source and receiver positions. The permittivity in the background is $\epsilon_b = 20$, and in the object $\hat{\epsilon} = 15$.

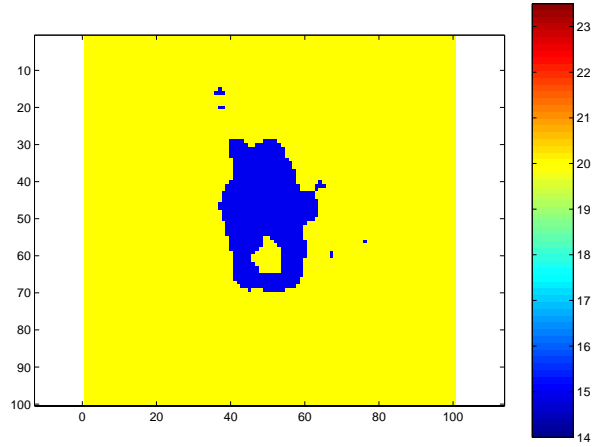


Figure 2: STAF reconstruction of permittivity distribution for the example with full view using $f = 30$ MHz and noisy data with 10 dB SNR.

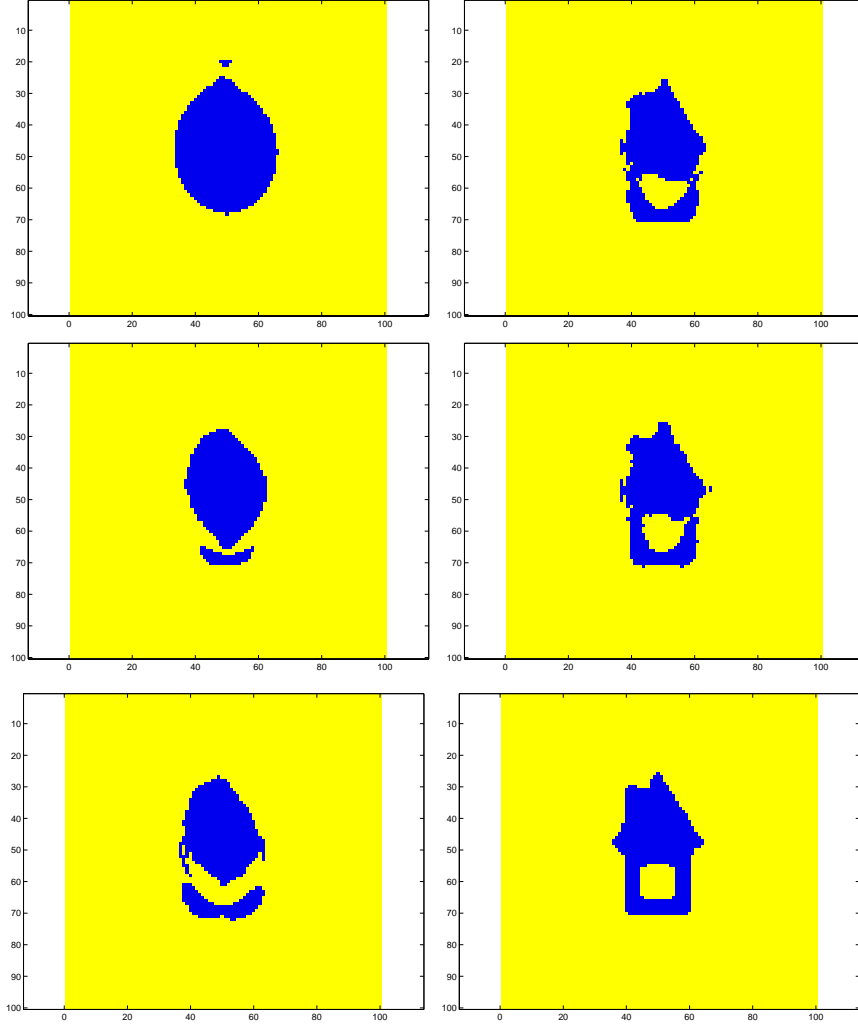


Figure 3: Evolution of permittivity $\epsilon^{(n)}$. Left column from top to bottom: STAF reconstructions of $\epsilon^{(0)}$ for 5 MHz (top left); This is the starting guess for the following reconstruction using levelART. After 10 steps of levelART with 10 MHz; After 30 steps with 10 MHz; Right column from top to bottom: After 10 steps with 20 MHz; After 30 steps with 20 MHz; Final reconstruction after 30 steps of levelART with 30 MHz (bottom right). The algorithm used noise-free data.

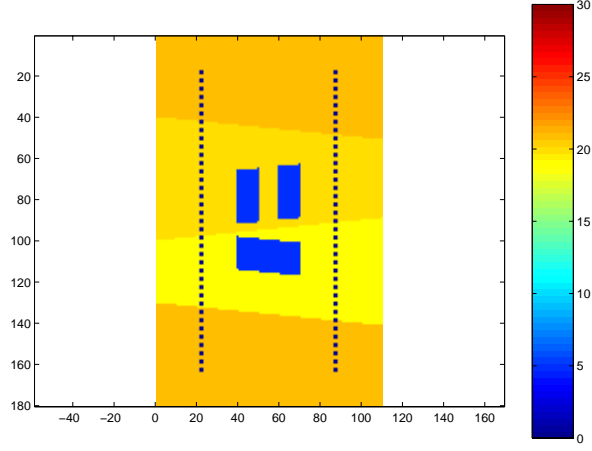


Figure 4: True permittivity distribution in the cross-borehole example. The dots in the figure indicate the source and receiver positions. The permittivity in the background layers is (from top to bottom) $\epsilon_b = 21, 20, 19$, and 21 . Inside the object it is $\hat{\epsilon} = 5$.

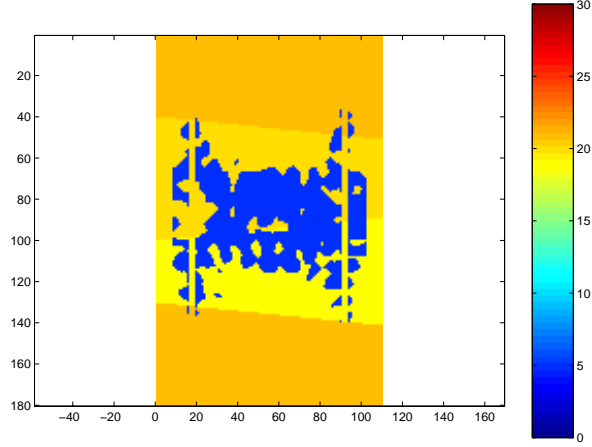


Figure 5: STAF reconstruction of permittivity distribution for cross-borehole example using noisy data with 30 MHz and 10 dB SNR.

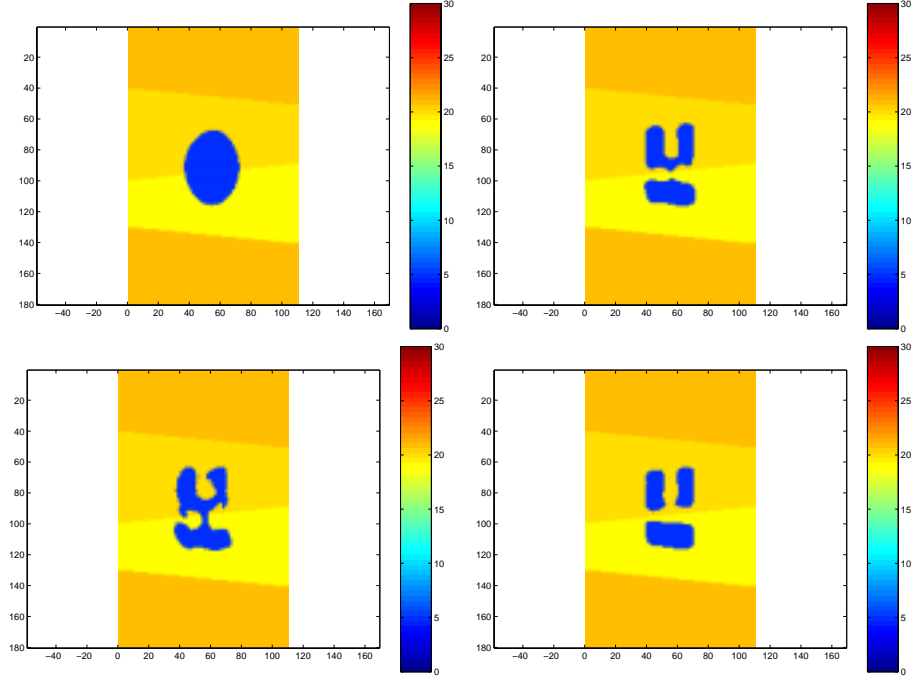


Figure 6: Evolution of the permittivity distribution $\epsilon^{(n)}$ in the cross-borehole example using noisy data with 10 dB SNR and limited view. Top left: STAF reconstruction of $\epsilon^{(0)}$ for 5 MHz. This is the starting guess for the following reconstruction using the levelART algorithm. Bottom left: After 20 steps of levelART with 15 MHz; Top right: After 2 sweeps with 20, 25, and 30 MHz; Bottom right: After 10 sweeps of levelART with 20, 25, and 30 MHz.

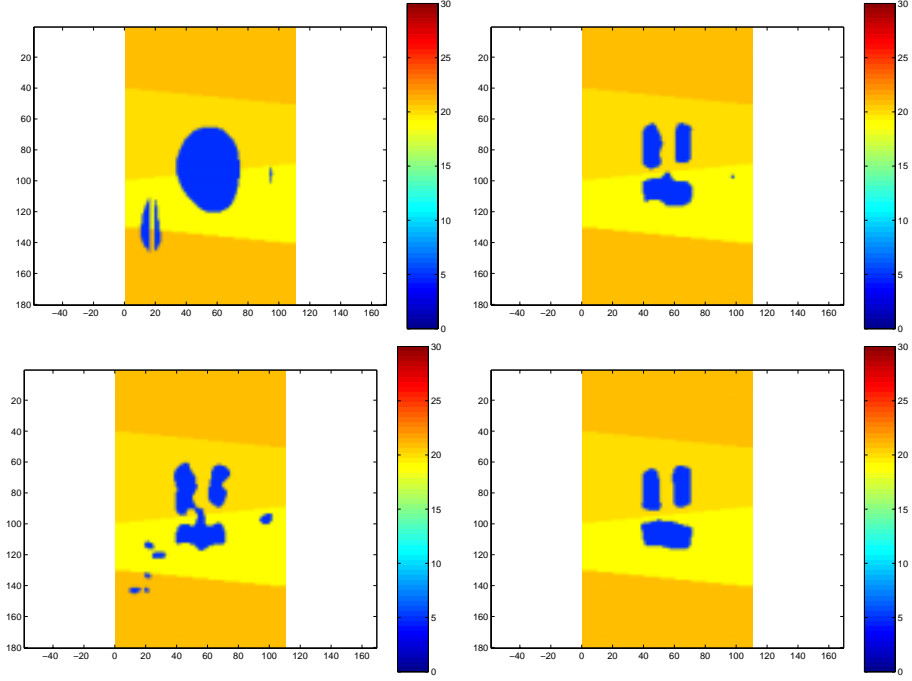


Figure 7: Evolution of the permittivity distribution $\epsilon^{(n)}$ in the cross-borehole example using noisy data with 5 dB SNR and limited view. Top left: STAF reconstruction of $\epsilon^{(0)}$ for 5 MHz. This is the starting guess for the following reconstruction using the levelART algorithm. Bottom left: After 20 steps of levelART with 15 MHz; Top right: After 2 sweeps with 20, 25, and 30 MHz; Bottom right: After 10 sweeps of levelART with 20, 25, and 30 MHz.

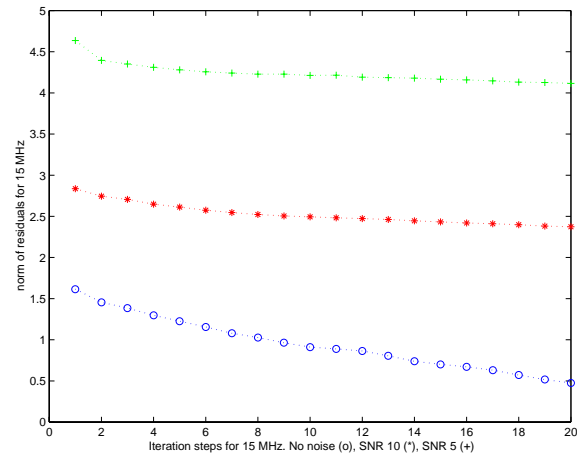


Figure 8: Norm of residuals for 15 MHz in cross-borehole example.

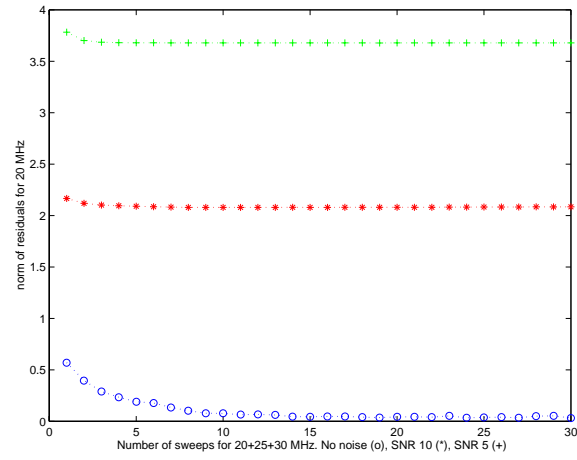


Figure 9: Norm of residuals for 20 MHz in cross-borehole example.

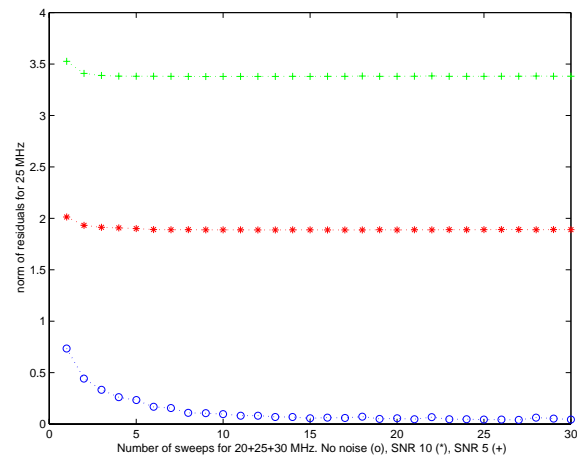


Figure 10: Norm of residuals for 25 MHz in cross-borehole example.

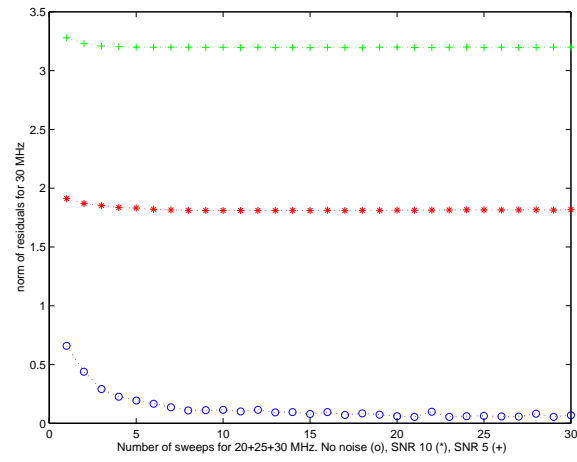


Figure 11: Norm of residuals for 30 MHz in cross-borehole example.

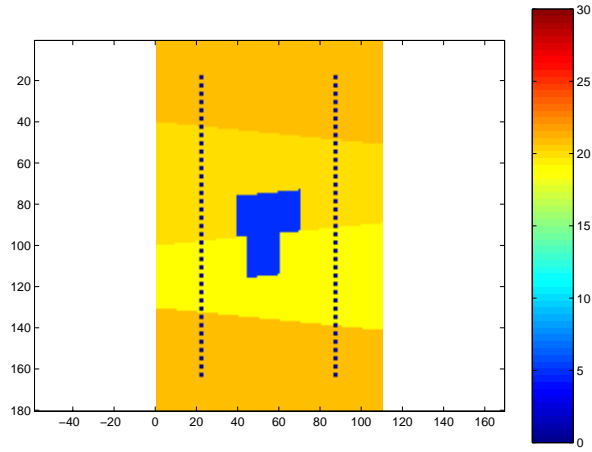


Figure 12: Original permittivity distribution $\epsilon^{(n)}$ in the cross-borehole example using noisy data with 10 dB SNR and limited view. The permittivity ϵ_b in the background is the same as in figure 4, and in the inclusion it is $\hat{\epsilon} = 5$. The conductivity in the background is $\sigma_b = 3.0 \times 10^{-4}$ S/m, and in the inclusion it is either $\hat{\sigma} = 1.0 \times 10^{-6}$ S/m (first example), or $\hat{\sigma} = 1.0 \times 10^{-2}$ S/m (second example). The dots in the figure indicate the source and receiver positions.

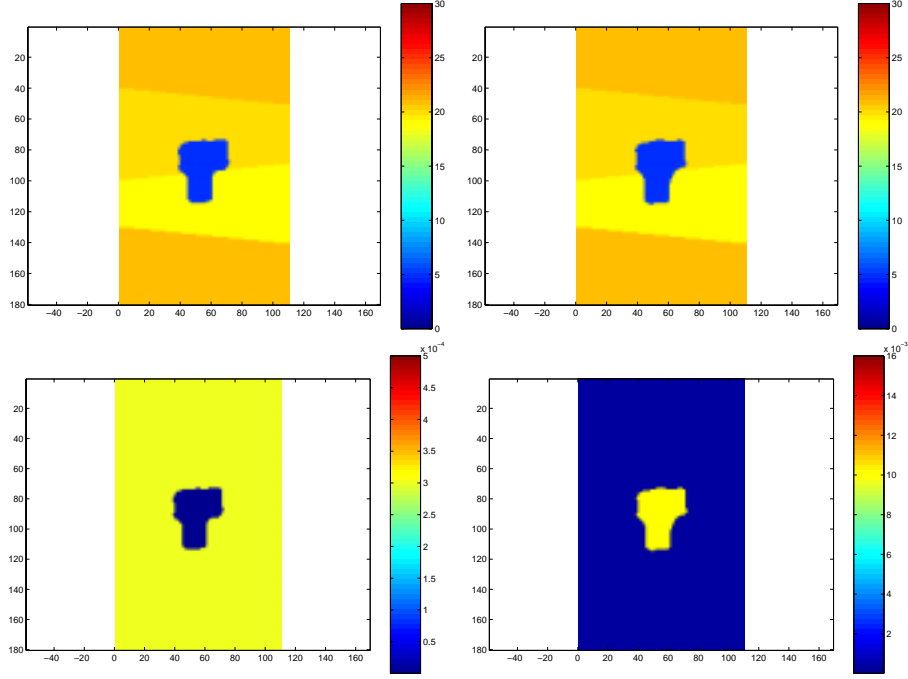


Figure 13: Reconstruction of permittivity and conductivity distributions after applying STAF with 5 MHz, and 6 sweeps of levelART with 15, 20, 25, and 30 MHz. Left column: Example with $\hat{\sigma} = 1.0 \times 10^{-6}$ S/m inside the obstacle. Shown is the reconstructed permittivity (top) and reconstructed conductivity (bottom). Right column: Example with $\hat{\sigma} = 1.0 \times 10^{-2}$ S/m inside the obstacle. Reconstructed permittivity (top) and reconstructed conductivity (bottom). The conductivity was treated as noise in STAF, and was considered linked to the current reconstruction of the permittivity distribution in levelART. All data were contaminated with white Gaussian noise of 10 dB SNR before starting the reconstruction process.

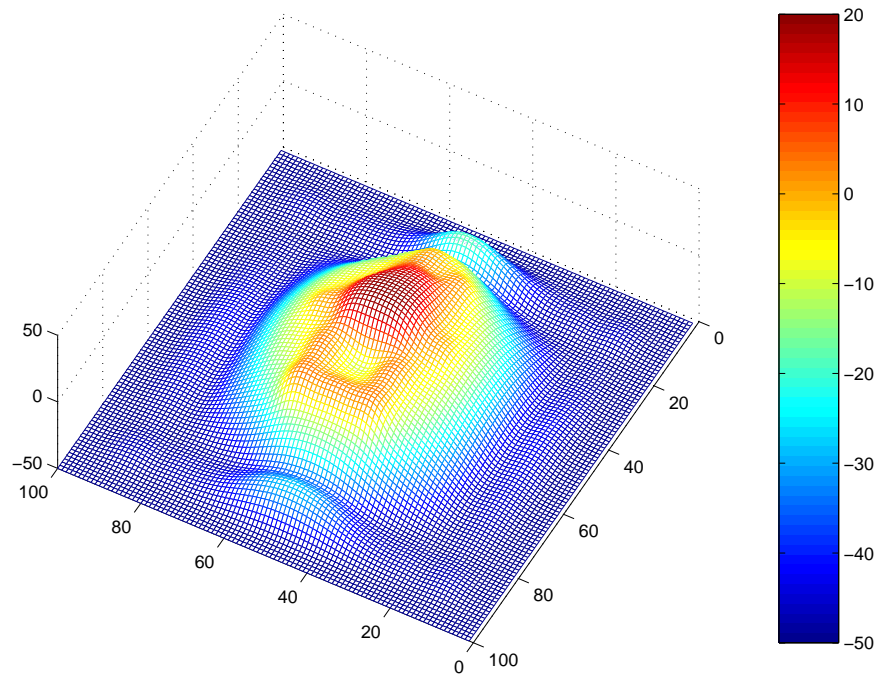


Figure 14: Final level set function $-\phi^{(N)}$ for the full-view example. Figure 3 shows the corresponding permittivity distribution $\epsilon^{(N)}$.

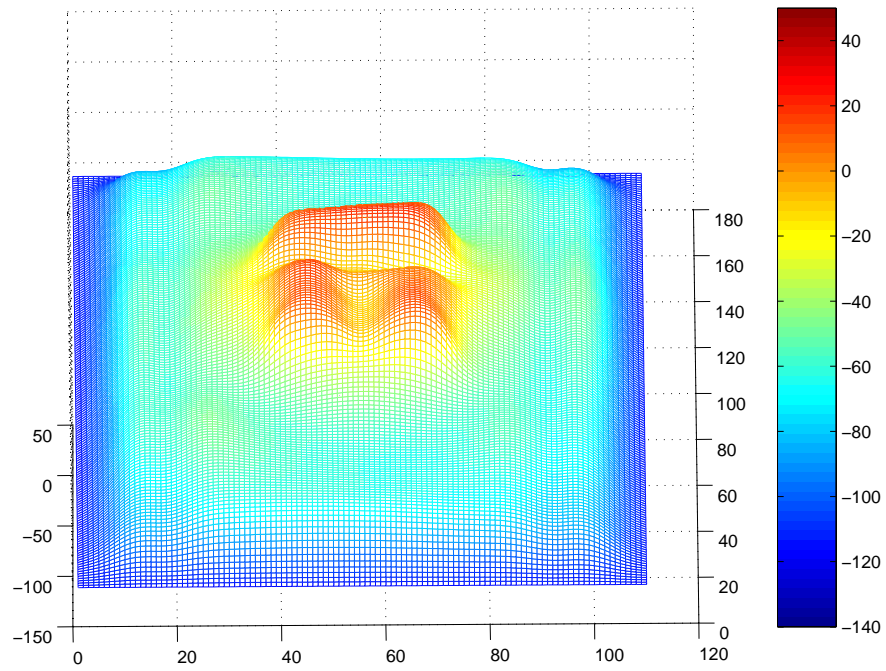


Figure 15: Final level set function $-\phi^{(N)}$ for the cross-borehole example using data with 10 dB SNR. Figure 6 shows the corresponding permittivity distribution $\epsilon^{(N)}$.

Template Matching Based Object Recognition with Unknown Geometric Parameters

Roger M. Dufour¹, Eric L. Miller¹, and Nikolas P. Galatsanos²

¹Communications and Digital Signal Processing Center

Department of Electrical and Computer Engineering

409 Dana Bldg.

Northeastern University

Boston, MA 02115

Tel.: (617) 373-8386

Fax: (617) 373-8627

email: rdufour@ece.neu.edu

email: elmiller@ece.neu.edu

²Electrical and Computer Engineering Department

Illinois Institute of Technology

Chicago, IL 60616

email: npg@ece.iit.edu

December 7, 2000

Abstract

In this paper we examine the problem of locating an object in an image when size and rotation are unknown. Previous work has shown that with known geometric parameters, an image restoration method can be useful by estimating a delta function at the object location. When the geometric parameters are unknown, this method becomes impractical because the likelihood surface to be minimized across size and rotation has numerous local minima and areas of zero gradient. In this paper, we propose a new approach where a smooth approximation of the template is used to minimize a well-behaved likelihood surface. A coarse-to-fine approximation of the original template using a diffusion-like equation is used to create a library of templates. Using this library, we can successively perform minimizations which are locally well-behaved. As detail is added to the template, the likelihood surface gains local minima, but previous estimates place us within a well-behaved “bowl” around the global minimum, leading to an accurate estimate. Numerical experiments are shown which verify the value of this approach.

EDICS

1 Introduction

A common image processing problem is to determine the location of an object using a template when the size and rotation of the true target are unknowns [1–3]. An algorithm for finding the solution should be robust to noise, accurate across a wide range of object configurations, and computationally efficient. With known geometric parameters (i.e. size and rotation) and additive Gaussian noise, the classic solution for object localization is a whitening filter followed by a matched filter. Other estimators have been proposed, such as the phase only matched filter (POMF) and the symmetric phase only matched filter (SPOMF) which give better location discrimination than the standard matched filter [3]. However these filters do not use additional information regarding the background noise which may be available or estimable. Alternatively, one may formulate the localization problem in the framework of image reconstruction, where the image to be recovered is a delta function at the location of the object and the blurring kernel is the template to be matched [1]. This approach makes available image reconstruction methods such as the linear least squares estimator (LLSE) and maximum likelihood estimator which can incorporate background statistics. The final step is simply to select the point of highest response in the recovered image.

Since they rely upon an accurate template the methods mentioned above are not sufficient when geometric parameters such as size and rotation are unknown. Here, one must compose an estimator which is invariant to these parameters or simultaneously estimate these parameters. With the invariant approach [4–7], a discriminant function is composed in the spatial, frequency or another domain which is invariant to an affine transformation of the template. The discriminant function is usually computed from local or global features of the image, boundaries within the image or regions of the image. The invariant function score can then be used for detection or classification. The second approach and the one followed in this paper is to estimate the parameters. By estimating

the parameters, an accurate template match can be achieved as in the methods mentioned earlier. This approach is also taken with the Fourier Mellin Matched Filter (FMMF) [2]. In the FMMF, the amplitude of the Fourier transform is mapped to a log-polar coordinate system. The translational dependence is eliminated because it appears only in the phase of the Fourier transform. The transformation to the log-polar coordinate system converts the rotation and scaling parameters into translational parameters. Scale and rotation are then solved with a matched filter or phase only matched filter. While computationally fast and a simple method, it is not very robust to noise or clutter. Additionally if the template does not accurately match the object in the image, this method again will often fail. Another approach is to minimize a likelihood surface defined from the difference between the matched template and the image. However, this surface is not amenable to minimization due to numerous local minima and areas of small or zero gradient. Randomized techniques can be used to minimize the surface such as simulated annealing [8, 9] or by using a jump diffusion technique [10, 11]. Here one updates the parameter set with a step increment, the increment becoming smaller as the algorithm runs until it settles into, hopefully, the global minimum.

We propose a deterministic method for searching this surface using a progression of templates. The early templates are smooth approximations of the exact template. This results in a search of a well-behaved approximation of the true likelihood surface and therefore we can get close to the global minimum of the cost surface. A standard optimization routine such as the Newton algorithm can be used to find the best fit solution. Using this estimate as a starting point, we add more detail to our template and search again. As we add detail, the surface becomes more ill-behaved, but the previous estimates have placed us within a “basin of attraction” of the global minimum. This method is similar to the Graduated Non-Convexity (GNC) approach [12–14] in that it allows us to locally search a series of approximations to the likelihood surface and in a small number of

iterations will carry us to an accurate solution. To generate the templates for the search, we use a diffusion like equation which allows fast Fourier based computations of the templates, coupled with the Fourier based image restoration method [1]. This leads to a Fourier domain algorithm which is not computationally burdensome.

We demonstrate the performance of the algorithm using synthetic images, infrared images and optical images. The experimental evidence shows that this approach is able to accurately estimate the size, rotation and location of an object across a wide range of signal to noise ratios and clutter.

In Section 2 we will introduce the problem setup and the solution method via an impulse estimation routine. In Section 3, we present a method for developing the template library. Section 4 shows the Newton algorithm used for parameter estimation. Section 5 discusses the Cramer-Rao bounds of the parameter estimates. In Section 6, we present our numerical experiments. Finally, in Section 7 we present some conclusions and directions for future work.

2 Background

Abu-Naser et. al. [1] formulated the problem of template matching with an image restoration framework, taking a delta function at the template location as the signal to be restored and the target template as the blurring kernel, which leads to the convolution equation,

$$g(\mathbf{r}) = f(\mathbf{r}; \boldsymbol{\theta}^0) * \delta(\mathbf{r} - \bar{\mathbf{r}}^0) + n(\mathbf{r}). \quad (1)$$

where the position vector $\mathbf{r}^T = [r_x \ r_y]$, and $*$ represents two dimensional convolution. Here the object is represented by the template, f , at some location $\bar{\mathbf{r}}^0$. We have described the template using a parameter set, $\boldsymbol{\theta}$ (where $\boldsymbol{\theta}^0$ is the true parameter set), which specifies the geometric parameters

of the template (i.e. the size s and rotation ϕ), as

$$\boldsymbol{\theta}^T = [s \ \phi]. \quad (2)$$

The template is a rotated and resized version of the standard template, $f(\mathbf{r})$, given as

$$f(\mathbf{r}; \boldsymbol{\theta}) = f\left(\frac{1}{s}\mathbf{M}(\phi)\mathbf{r}\right), \text{ where } \mathbf{M}(\phi) = \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix}. \quad (3)$$

As in [1] for estimation purposes we assume that the noise and the δ are zero-mean stochastic functions with spectra given as

$$S_\delta(\mathbf{k}) = \sigma_\delta^2 \text{ and, } S_n(\mathbf{k}) = \sigma_n^2, \quad (4)$$

where $\mathbf{k}^T = [k_x \ k_y]$ are the spatial frequencies used in the two dimensional Fourier Transform

$$\tilde{f}(\mathbf{k}) = \mathcal{F}\{f(\mathbf{r})\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\mathbf{r}) \exp(-j2\pi\mathbf{k} \cdot \mathbf{r}) \, d\mathbf{r}. \quad (5)$$

With this framework, we now use a maximum likelihood estimate of the parameter set $\hat{\boldsymbol{\theta}}$, while using the delta restoration method [1] for the location estimation $\hat{\mathbf{r}}$. Since the additive noise is assumed Gaussian, the negative log-likelihood for our parameters is

$$l(\boldsymbol{\theta}, \bar{\mathbf{r}}|g) = \frac{1}{\sigma_n^2} \|f(\mathbf{r}; \boldsymbol{\theta}) * \delta(\mathbf{r} - \bar{\mathbf{r}}) - g(\mathbf{r})\|_2^2. \quad (6)$$

The maximum likelihood solution is then achieved by minimizing (6) as

$$\{\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}\} = \arg \min_{\{\bar{\mathbf{r}}, \boldsymbol{\theta}\}} l(\boldsymbol{\theta}, \bar{\mathbf{r}}|g). \quad (7)$$

With a given set of geometric parameters $\boldsymbol{\theta}$, we can construct an estimate of position following Abu-Naser [1] by first making a Linear Least Squares Estimate (LLSE) of the delta function, [1]

$$\hat{\delta}(\mathbf{r}) = \mathcal{F}^{-1} \left\{ \frac{\tilde{f}_0^*(\mathbf{k}; \boldsymbol{\theta})}{|\tilde{f}_0(\mathbf{k}; \boldsymbol{\theta})|^2 + \left(\frac{\sigma_n}{\sigma_\delta}\right)^2} \tilde{g}(\mathbf{k}) \right\}, \quad (8)$$

and selecting the point of maximum response as our position estimate, as

$$\hat{\mathbf{r}}(\boldsymbol{\theta}) = \arg \max_{\mathbf{r}} \hat{\delta}(\mathbf{r}). \quad (9)$$

Substituting (9) into (7), we have the maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$ as

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \frac{1}{\sigma_n^2} \|f(\mathbf{r}, \boldsymbol{\theta}) * \delta(\mathbf{r} - \hat{\mathbf{r}}(\boldsymbol{\theta})) - g(\mathbf{r})\|_2^2 \equiv \arg \min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}; f). \quad (10)$$

The minimization of $J(\boldsymbol{\theta}; f)$ could be accomplished by calculating $J(\boldsymbol{\theta}; f)$ on a dense grid in the $\phi - s$ space to find the minimum point. However, the calculations required to perform this are generally prohibitive. The computation of the surface shown in Figure 1 needed 67 billion flops. Instead, we seek to use standard optimization tools to find the minimum, , the gradient descent algorithm we propose used only 260 million flops to find the minimum of the surface in Figure 1. The problem which we encounter when we attempt to do this is that the likelihood surface is not amenable to a gradient descent methods because of local minima and other areas of zero gradient. We find that these problem areas are a direct result of the shape of the template, f ; specifically,

non-convexities in the template lead to non-convexities in the likelihood surface. By non-convexities we mean a template which when viewed as a surface (height as a function of position in the image plane) has multiple extreme points, instead of a surface which has a single minimum or maximum. This can be seen in Figure 1 where we have a simple non-convex target template and the $J(\theta; f)$ surface directly below it. The example given here is for estimating the same two peaked object with size $s = 1$ and rotation $\phi = 0$ with noiseless data. We can see that the minimum of this surface is a sharp point at the correct parameters, but the irregularity of the surface makes descent-type minimizations impractical. Conversely, if we use a smooth rotationally invariant template, shown on the right, to perform the match we obtain the second surface which is far better behaved. The minimum of this surface while close to 1.0 is much broader in the direction of s , meaning that the estimate of s is not as accurate as the estimate which can be produced by the exact template. As for rotation, the second template gives no information whatsoever as it is rotationally invariant. Thus the primary objective of this work is to capture the accuracy of the exact template estimation with the more easily minimizable surface of the approximate template.

We formulate a method which uses smooth approximate templates which allow us to minimize on a well-behaved surface. Since the results of the minimizations which we obtain from the approximations are not as accurate as those which could be obtained from the exact template, we will add detail back to our templates and minimize again with the previous estimate starting our current estimate within a well-behaved region around the global minimum.

3 Template Progression

In this section we will detail a method of generating target templates which are approximations of the true template but which are sufficiently smooth so that the likelihood surface is locally

well-behaved. This will allow us to make successive estimates beginning at smooth approximation which will generally locate us near the global optimum and refine that estimate with more detailed approximations of the template until a sufficiently accurate estimate of the parameters is reached.

We will generate a spectrum of templates, indexed by t as f_t , ranging from a smooth template f_0 to the exact template $f = \lim_{t \rightarrow \infty} f_t$. We desire that the most smooth template be convex to induce the behavior in the likelihood surface which we discussed in the previous section. A relatively simple choice of this is a Gaussian blob matched (in the two-norm sense) in size and amplitude to the original template given as

$$f_0(\mathbf{r}) = A \exp\left(-\frac{\|\mathbf{r}\|_2^2}{2w^2}\right). \quad (11)$$

While successive estimation will hopefully locate us inside a “well of attraction” of the global minimum where we will avoid local minima, discontinuities in any template can yield discontinuities on our surface. We therefore choose a method for generating templates which in the continuous domain produces continuous templates. A diffusion like process performs this adequately. We specify the Fourier transform of a template f_t as

$$\tilde{f}_t(\mathbf{k}) = \left(\tilde{f}_E(\mathbf{k}) - \tilde{f}_0(\mathbf{k})\right) \exp\left(\frac{-\|\mathbf{k}\|^2}{t}\right) + \tilde{f}_0(\mathbf{k}), \quad (12)$$

where \tilde{f}_E and \tilde{f}_0 are the Fourier transforms of the exact and most smooth templates, respectively.

A progression for the two peak template is shown in Figure 2. We see that at small values of t , the template is a smooth approximation of the true template, as t increases, the true template emerges. Associated with this are the likelihood surfaces related to each template. Here we see that at small values of t , the surface is very smooth has no rotational localization and a very broad scale localization. As t increases, the ill-behavior returns, but previous estimates place us within

the area of the global minimum, and our estimate becomes more accurate. The final solution for the parameters will of course be a local minimum, but may not necessarily be the global minimum. As will be shown later, the rate at which the templates evolve influences the final outcome of the optimization. Slower t -schedules as expected lead to more accurate estimates of the global minimum.

We can now describe a complete algorithm using the template progression as

1. Begin at $t = 0$.
2. Construct f_t with (12).
3. Minimize the likelihood surface constructed via the equations

$$\hat{\mathbf{r}}_t(\boldsymbol{\theta}) = \arg \max_{\mathbf{r}} \mathcal{F}^{-1} \left\{ \frac{\tilde{f}_t^*(\mathbf{k}; \boldsymbol{\theta})}{|\tilde{f}_t(\mathbf{k}; \boldsymbol{\theta})|^2 + \left(\frac{\sigma_n}{\sigma_\delta}\right)^2} \tilde{g}(\mathbf{k}) \right\}. \quad (13)$$

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \frac{1}{\sigma_n^2} \|f_t(\mathbf{r}; \boldsymbol{\theta}) * \delta(\mathbf{r} - \hat{\mathbf{r}}_t(\boldsymbol{\theta})) - g(\mathbf{r})\|_2^2. \quad (14)$$

The minimization in 14 is performed via a Newton algorithm given in section 4.

4. Increase t and proceed to step 2. The t -schedule should be chosen to take small steps at low values of t where the algorithm is more sensitive to changes in the likelihood surface. This agrees with the observations in [13] for the method of graduated non-convexity and also demonstrated in the analysis of the bounds on estimation accuracy in Section 5. In our implementation, a small value is chosen for the initial value of t and it is doubled for each subsequent t . If a longer schedule is desired, a smaller multiplier is used.

4 Newton Algorithm

The Newton algorithm [15] is used for the minimizations in our paper. The Newton algorithm implemented here seeks to minimize the squared error in the estimated image against the data according to

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}; f_t) = \arg \min_{\boldsymbol{\theta}} \int e_t^2(\mathbf{r}; \boldsymbol{\theta}) d\mathbf{r} \quad (15)$$

where e_t is the error image given as

$$e_t(\mathbf{r}; \boldsymbol{\theta}) = f_t(\mathbf{r}; \boldsymbol{\theta}) * \delta(\mathbf{r} - \hat{\mathbf{r}}_t(\boldsymbol{\theta})) - g(\mathbf{r}). \quad (16)$$

The Newton iterative procedure produces updates of the parameter vector as [15]

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \mathbf{p}^{(k)}, \quad (17)$$

$$\text{with } (\mathbf{U}_t(\boldsymbol{\theta}^{(k)}) + \mathbf{S}_t(\boldsymbol{\theta}^{(k)}))\mathbf{p}^{(k)} = - \int \mathbf{J}_t(\mathbf{r}; \boldsymbol{\theta}^{(k)}) e_t(\mathbf{r}; \boldsymbol{\theta}^{(k)}) d\mathbf{r}, \quad (18)$$

where $\mathbf{p}^{(k)}$ is the update vector for the parameter set, \mathbf{J}_t is the Jacobian vector of the error function, and \mathbf{U}_t and \mathbf{S}_t are functions of the Jacobian and Hessian that are described below.

The Jacobian vector is the vector of first derivative functions of the of the error as

$$\begin{aligned} \mathbf{J}_t(\mathbf{r}; \boldsymbol{\theta}) &= \nabla_{\boldsymbol{\theta}} e_t(\mathbf{r}; \boldsymbol{\theta}) = \left[\frac{\partial e_t(\mathbf{r}; \boldsymbol{\theta})}{\partial s} \quad \frac{\partial e_t(\mathbf{r}; \boldsymbol{\theta})}{\partial \phi} \right]^T \\ &= \nabla_{\boldsymbol{\theta}} f_t(\mathbf{r}; \boldsymbol{\theta}) * \delta(\mathbf{r} - \hat{\mathbf{r}}_t(\boldsymbol{\theta})) + f_t(\mathbf{r}; \boldsymbol{\theta}) * \delta(\mathbf{r} - \hat{\mathbf{r}}_t(\boldsymbol{\theta})) * \nabla_{\boldsymbol{\theta}} \delta(\mathbf{r} - \hat{\mathbf{r}}_t(\boldsymbol{\theta})) \end{aligned} \quad (19)$$

$$= \nabla_{\boldsymbol{\theta}} f_t(\mathbf{r} - \hat{\mathbf{r}}_t(\boldsymbol{\theta}); \boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}} \hat{\mathbf{r}}_t(\boldsymbol{\theta}) \nabla_{\mathbf{r}} f_t(\mathbf{r} - \hat{\mathbf{r}}_t(\boldsymbol{\theta}); \boldsymbol{\theta}). \quad (20)$$

The first term of (20) is directly computable from the template function, however the second term is

troublesome. The gradient $\nabla_{\boldsymbol{\theta}} \hat{\mathbf{r}}_t(\boldsymbol{\theta})$ is not guaranteed to exist everywhere, and even if it does exist, it cannot be calculated in closed form since $\hat{\mathbf{r}}_t(\boldsymbol{\theta})$ involves a maximization. Numerical computation of $\nabla_{\boldsymbol{\theta}} \hat{\mathbf{r}}_t(\boldsymbol{\theta})$ is also complicated by the large granularity of image pixels with respect to the usual size of $\nabla_{\boldsymbol{\theta}} \hat{\mathbf{r}}_t(\boldsymbol{\theta})$. By this, we mean that if we attempt to approximate an element of $\nabla_{\boldsymbol{\theta}} \hat{\mathbf{r}}_t(\boldsymbol{\theta})$ (for example $\frac{\partial \hat{r}_x(\boldsymbol{\theta})}{\partial s}$) by the relation

$$\frac{\partial \hat{r}_x(s, \phi)}{\partial s} \approx \frac{\hat{r}_x(s + \Delta s, \phi) - \hat{r}_x(s, \phi)}{\Delta s}, \quad (21)$$

we find that the numerator is either zero (if the two estimates of position are usually the same pixel), or arbitrarily large (if they are different pixels) because Δs is made small. Therefore we used an alternative method to obtain a more accurate approximation of the elements of $\nabla_{\boldsymbol{\theta}} \hat{\mathbf{r}}_t(\boldsymbol{\theta})$. First we increase s until $\hat{r}_x(s, \phi)$ moved at least by a pixel and denoted this point by s_1 . Similarly, we then find another point by decreasing s and this point is denoted by s_2 . The derivative is then approximated by

$$\frac{\partial \hat{r}_x(s, \phi)}{\partial s} \approx \frac{\hat{r}_x(s_2, \phi) - \hat{r}_x(s_1, \phi)}{s_2 - s_1}. \quad (22)$$

A similar approach was used to compute the remaining three elements of $\nabla_{\boldsymbol{\theta}} \hat{\mathbf{r}}_t(\boldsymbol{\theta})$, that is $\frac{\partial \hat{r}_y(\boldsymbol{\theta})}{\partial s}$, $\frac{\partial \hat{r}_x(\boldsymbol{\theta})}{\partial \phi}$, and $\frac{\partial \hat{r}_y(\boldsymbol{\theta})}{\partial \phi}$.

The disadvantage of this method is that it is computationally intensive since we must calculate $\hat{\mathbf{r}}_t(\boldsymbol{\theta})$ at many points. Using this method, we find that the elements of the matrix $\nabla_{\boldsymbol{\theta}} \hat{\mathbf{r}}_t(\boldsymbol{\theta})$ are typically two to three orders of magnitude smaller than $\nabla_{\boldsymbol{\theta}} f_t(\mathbf{r} - \hat{\mathbf{r}}_t(\boldsymbol{\theta}); \boldsymbol{\theta})$ in (20). This is demonstrated by the figures in Table 1, which shows a comparison between the average magnitude of $\nabla_{\boldsymbol{\theta}} f_t(\mathbf{r}; \boldsymbol{\theta})$ and the elements of $\nabla_{\boldsymbol{\theta}} \hat{\mathbf{r}}_t(\boldsymbol{\theta})$ for 20 runs (1095 iterations) of the algorithm for the two-peak example shown earlier. It is reasonable therefore to simplify the calculation by eliminating this term, and

using the approximation for the Jacobian given by

$$\mathbf{J}_t(\mathbf{r}; \boldsymbol{\theta}) \approx \nabla_{\boldsymbol{\theta}} f_t(\mathbf{r} - \hat{\mathbf{r}}_t(\boldsymbol{\theta}); \boldsymbol{\theta}). \quad (23)$$

The Hessian of the error, or the matrix of second derivatives, is

$$\begin{aligned} \mathbf{H}_t(\mathbf{r}; \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}^2 e_t(\mathbf{r}; \boldsymbol{\theta}) = & [\nabla_{\boldsymbol{\theta}}^2 f_t(\mathbf{r}; \boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}}^2 \hat{\mathbf{r}}_t(\boldsymbol{\theta}) \nabla_{\mathbf{r}} f_t(\mathbf{r}; \boldsymbol{\theta}) \\ & - 2 \nabla_{\boldsymbol{\theta}} \hat{\mathbf{r}}_t(\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}} \nabla_{\mathbf{r}} f_t(\mathbf{r}; \boldsymbol{\theta}) + (\nabla_{\boldsymbol{\theta}} \hat{\mathbf{r}}_t(\boldsymbol{\theta}))^2 \nabla_{\mathbf{r}}^2 f_t(\mathbf{r}; \boldsymbol{\theta})] * \delta(\mathbf{r} - \hat{\mathbf{r}}_t(\boldsymbol{\theta})) \end{aligned} \quad (24)$$

Similarly to the Jacobian, the terms of the Hessian which contain the gradients of $\hat{\mathbf{r}}_t(\boldsymbol{\theta})$ are usually not significant and can cause computational problems, so we disregard those terms. The approximate Hessian which we used is given by

$$\mathbf{H}_t(\mathbf{r}; \boldsymbol{\theta}) \approx \nabla_{\boldsymbol{\theta}}^2 f_t(\mathbf{r} - \hat{\mathbf{r}}_t(\boldsymbol{\theta}); \boldsymbol{\theta}). \quad (25)$$

The two matrices on the left hand side of (18) are computed from the Jacobian and Hessian. The first is the inner product of the Jacobian with itself, and the second is given as the integral of the Hessian components with the error. Therefore,

$$\mathbf{U}_t(\boldsymbol{\theta}) = \int \mathbf{J}_t(\mathbf{r}; \boldsymbol{\theta}) \mathbf{J}_t^T(\mathbf{r}; \boldsymbol{\theta}) d\mathbf{r}. \quad (26)$$

Specifying the elements of $\mathbf{S}_t(\boldsymbol{\theta})$, by $s_{i,j}(\boldsymbol{\theta})$ and the elements of $\mathbf{H}_t(\mathbf{r}; \boldsymbol{\theta})$ by $h_{i,j}(\mathbf{r}; \boldsymbol{\theta})$ we have

$$s_{i,j}(\boldsymbol{\theta}) = \int h_{i,j}(\mathbf{r}; \boldsymbol{\theta}) e_t(\mathbf{r}; \boldsymbol{\theta}) d\mathbf{r}. \quad (27)$$

The Newton algorithm is then an iteration of (18) and (17) until the likelihood as evaluated by

(10) ceases to change significantly. That is while

$$J(\boldsymbol{\theta}^{(k)}; f_t) - J(\boldsymbol{\theta}^{(k-1)}; f_t) > \tau \quad (28)$$

for some small value of τ .

5 Performance and Bounds

The variance of the parameter estimates is constrained by the Cramer-Rao (CR) bounds. We find that the CR lower bounds upon the estimates are a function of the noise power and the mismatch between the approximate template and the true target template. As a function of the approximate template, the bounds show the accuracy of the estimates at various values of t . Furthermore since they are proportional to the curvature of the local basin of attraction of the likelihood surface, they can also describe the manner in which the surface is evolving.

We will arrive at the bounds as the diagonal elements of the inverse of the Fisher information matrix [16, 17]. Since we can express the likelihood as a function of the error as

$$l_t(\boldsymbol{\theta}, \bar{\mathbf{r}}; g) = \frac{1}{\sigma_n^2} \int e_t^2(\mathbf{r}; \boldsymbol{\theta}, \bar{\mathbf{r}}) d\mathbf{r}, \quad (29)$$

we can express the derivatives with respect to the template parameters as

$$\frac{\partial l_t(\boldsymbol{\theta}, \bar{\mathbf{r}}; g)}{\partial \theta_i} = \frac{2}{\sigma_n^2} \int e_t(\mathbf{r}; \boldsymbol{\theta}, \bar{\mathbf{r}}) \frac{\partial e_t(\mathbf{r}; \boldsymbol{\theta}, \bar{\mathbf{r}})}{\partial \theta_i} d\mathbf{r}, \quad (30)$$

and with respect to location as

$$\frac{\partial l_t(\boldsymbol{\theta}, \bar{\mathbf{r}}; g)}{\partial r_x} = \frac{2}{\sigma_n^2} \int e_t(\mathbf{r}; \boldsymbol{\theta}, \bar{\mathbf{r}}) \frac{\partial e_t(\mathbf{r}; \boldsymbol{\theta}, \bar{\mathbf{r}})}{\partial r_x} d\mathbf{r}, \quad \text{and} \quad \frac{\partial l_t(\boldsymbol{\theta}, \bar{\mathbf{r}}; g)}{\partial r_y} = \frac{2}{\sigma_n^2} \int e_t(\mathbf{r}; \boldsymbol{\theta}, \bar{\mathbf{r}}) \frac{\partial e_t(\mathbf{r}; \boldsymbol{\theta}, \bar{\mathbf{r}})}{\partial r_y} d\mathbf{r}, \quad (31)$$

Expanding the error term, and taking the expectations we arrive at the result

$$\begin{aligned} E \frac{\partial l_t(\boldsymbol{\theta}, \bar{\mathbf{r}}; g)}{\partial \theta_i} \frac{\partial l_t(\boldsymbol{\theta}, \bar{\mathbf{r}}; g)}{\partial \theta_j} &= \frac{4}{\sigma_n^4} \int \frac{\partial f_t(\mathbf{r} - \bar{\mathbf{r}})}{\partial \theta_i} (f_t(\mathbf{r} - \bar{\mathbf{r}}) - f_0(\mathbf{r})) d\mathbf{r} \int \frac{\partial f_t(\mathbf{r} - \bar{\mathbf{r}})}{\partial \theta_j} (f_t(\mathbf{r} - \bar{\mathbf{r}}) - f_0(\mathbf{r})) d\mathbf{r} \\ &\quad + \frac{4}{\sigma_n^2} \int \frac{\partial f_t(\mathbf{r} - \bar{\mathbf{r}})}{\partial \theta_i} \frac{\partial f_t(\mathbf{r} - \bar{\mathbf{r}})}{\partial \theta_j} d\mathbf{r}, \end{aligned} \quad (32)$$

where we take $\bar{\mathbf{r}}^0 = \mathbf{0}$ without loss of generality. The terms with respect to location are similar.

We have a term in the information generated by the model mismatch and another generated by the noise term. We can construct the Fisher information matrix as

$$\mathbf{I} = \begin{bmatrix} E \left(\frac{\partial l}{\partial s} \right)^2 & E \frac{\partial l}{\partial s} \frac{\partial l}{\partial \phi} & E \frac{\partial l}{\partial s} \frac{\partial l}{\partial r_x} & E \frac{\partial l}{\partial s} \frac{\partial l}{\partial r_y} \\ E \frac{\partial l}{\partial s} \frac{\partial l}{\partial \phi} & E \left(\frac{\partial l}{\partial \phi} \right)^2 & E \frac{\partial l}{\partial \phi} \frac{\partial l}{\partial r_x} & E \frac{\partial l}{\partial \phi} \frac{\partial l}{\partial r_y} \\ E \frac{\partial l}{\partial s} \frac{\partial l}{\partial r_x} & E \frac{\partial l}{\partial \phi} \frac{\partial l}{\partial r_x} & E \left(\frac{\partial l}{\partial r_x} \right)^2 & E \frac{\partial l}{\partial r_x} \frac{\partial l}{\partial r_y} \\ E \frac{\partial l}{\partial s} \frac{\partial l}{\partial r_y} & E \frac{\partial l}{\partial \phi} \frac{\partial l}{\partial r_y} & E \frac{\partial l}{\partial r_x} \frac{\partial l}{\partial r_y} & E \left(\frac{\partial l}{\partial r_y} \right)^2 \end{bmatrix}. \quad (33)$$

Then from the elements of the inverse of the Fisher information matrix we have the lower bounds upon the variance of the parameter estimations.

In Figure 3 we see the standard deviation bounds for the estimator versus the progression index t for the two-peak template examined in Section 3. It is seen that the size estimate is least accurate at $t = 0$, and asymptotically approaches the CRB for the exact template as t approaches infinity. The rotational estimate begins unbounded since f_0 has no rotational component. Thus at low

values of t , we can only make a rough estimate of the size and no estimate of rotation. We can see that as the value of t increases, the bounds become lower also, indicating a more accurate estimate may be achievable. The bounds for the location estimate perform similarly to that of size.

In Figure 4 we see the standard deviation bounds versus noise. These behave as expected with exact estimation possible in the no noise case and estimation performance degrading as noise increases.

These bounds also influence the t -schedule, or the rate at which the template evolves. At $t = 0$, the template is smoothest and the estimates are the least accurate; however, this is also when the surface is the most well behaved and has the broadest well around the global minimum. The template must evolve in such a way that the estimate remains within the well. Using the CRB as a gauge of the curvature, we see that initially small changes in t are necessary so as not to too drastically change the surface. However as t increases larger steps can be taken since the surface evolves slower with respect to t . This type of schedule is similar to that used in the Graduated Non-Convexity approach in [12–14].

6 Numerical Experiments

In this section we present numerical experiments from the proposed algorithm and performance comparisons with the Fourier Mellin Matched Filter estimate [2]. We first present a synthetic target estimation example using a Monte Carlo simulation analysis. Then we show results of the proposed algorithm when applied to infrared (IR) and optical images.

Figure 5 (a) shows the two-peak target shown earlier buried in noise with 0 dB SNR. Figure 5 (b) and (c) show the estimates of s and ϕ at each iteration of the Newton algorithm. Here we used a t -schedule with four values, $\{10, 5, 2.5, 0\}$. We see here that the algorithm converged to close to

the true parameters in 25 iterations. The true parameters were a size of 1.0 and a rotation angle of 0.25 the estimations converged to values of 0.9914 for size and 0.2354 for rotation.

In Figure 6, we show scatter plots of estimates from 400 simulations of the two-peak example at varying signal to noise ratios. The true parameters are a size of 0.5 and a rotation of 0.25. The ellipses are used to show estimates which are “accurate”, that is are close to the true parameters. The center of the ellipse is at the true parameters and the ellipses are drawn at 3 times the CRB computed from the equations in the previous section. The interiors of these ellipses can be viewed as regions of sufficient confidence in which “accurate” estimates of the parameters fall. We observe that in (a), (b) and (c), the number of estimates which are outside the ellipse, and are therefore considered misses, decreases. The number of misses increases for the highest SNR because of a bias in the size estimates which when combined with the tightness of the CRB causes us to register more misses at high SNR levels than may be warranted by the fact that these estimates are close to the true parameters.

Table 2 summarizes the Monte Carlo runs of the example just presented. Each line shows the result of 400 simulations of the algorithm for a specific level of noise and t -schedule. The error in the estimates is composed of two components, one caused by the variance of the local minimum around the true parameters, and the second caused by the the algorithm becoming trapped in a local minimum which is far from the true parameters. The first type of error is bounded below by the CR bounds, and in practice is usually so small as to effect the target location estimation by less than one pixel. The second type of error constitutes a catastrophic miss by the algorithm. This error is demonstrated in Figure 7 with a pair of Monte Carlo runs at two different t -schedules. From the first scatter plot, we see that the majority of estimates are grouped around the true parameters, but 35 of the estimates missed dramatically. These are instances where the algorithm became caught in a local minimum which was not the global minimum, and thus resulted in an

enormous error. The second scatter plot shows the same Monte Carlo runs with a t -schedule which has twice as many stops and thus has twice the computational burden. Here, the number of misses was reduced from 35 to 18. By progressing through the t -schedule at a slower rate, we can reduce the chances of become trapped in a local minimum at the expense of more computations.

In Figure 8 the performance for these Monte Carlo runs is compared with that of the Fourier Mellin Matched Filter [2]. The plot shows the standard deviation of the error in the location estimate versus SNR. We see that at high SNR, both methods perform well, but as the SNR decreases then error in the FMMF increases at a faster rate.

Figure 9 shows an example of estimating the size, rotation and location of a vehicle in an IR image. Shown first is the target template which we are using followed by the data image. Below this is are the estimation values for size and rotation for each iteration. The iterations are divided into several sections by vertical dotted lines denoting the respective value of t for each stage of the algorithm. We can see that the algorithm converges in 60 iterations to values which upon matching the template to the data appear to be appropriate. Also shown is the position estimate of the target which also appears to be correct.

Figure 10 shows an example with a real optical image. The first image shows a cup in a simple background which is close to white noise. The second image shows the estimated position of the template with proper size and rotation. We see here that the algorithm settled into this estimate after 42 iterations and that it is an accurate estimate. In Figure 11, we show the output of the Fourier-Mellin matched filter which fails to isolate the true size and rotation for this image.

7 Conclusions and Future Work

In this work we have examined the problem of finding a target in a noisy image. Following in the work of Abu-Naser [1], we represent the problem as an image restoration problem where the object to be reconstructed is a delta function encoding the target location and the blur is a target template. Previous work showed that the restoration can be performed using a LLSE and choosing the location of maximum response [1]. However, this did not consider the problem of unknown geometric parameters, i.e. the size and rotation of the target within the image. To estimate size and rotation of the object, we choose to follow the previous approach for location estimation and minimize the negative of the likelihood function across size and rotation to find the true parameters.

This approach was not amenable to minimization, however, due to local minima and regions of zero derivative on the likelihood surface. To overcome these difficulties, we present a new method of generating a library of target templates which range from smooth convex approximations to the exact target. Using this library of templates, it is possible to sequentially estimate the parameter set using standard optimization tools.

Furthermore, since this method relies upon a functional description of the target template, we believe it to be extensible to more complicated target detection problems. Three dimensional perspective mappings of targets could be functionally described, allowing us to locate a target under a three dimensional rotation. Also, since all targets degenerate to the same template at the extremely smooth scale and the algorithm returns a likelihood score as it converges towards the true parameters, we could perform target classification by allowing our library to branch towards several targets or target classes and pruning those which score poorly. Overall, we believe this algorithm could be extended for multiple target, multiple template recognition with reasonable computational costs.

References

- [1] A. Abu-Naser, N.P. Galatsanos, M.N. Wernick, and D. Schonfeld. Object recognition based on impulse restoration using the expectation maximization algorithm. *Journal of the Optical Society of America A: Optics and Image Science*, 15:2327, 1998.
- [2] Q. Chen, M. Defrise, and F. Deconinck. Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and image recognition. *IEEE Trans. on PAMI*, 16(12), December 1994.
- [3] H. Lester and S.R. Arridge. Survey of heirarchical non-linear medical image registration. *Pattern Recognition*, 32(1):129–149, 1999.
- [4] Zhiqian Wang and Jezekiel Ben-Arie. Model based segmentation and detection of affine transformed shapes in cluttered images. In *IEEE International Conference on Image Processing*. IEEE, October 1998.
- [5] Isaac Weiss. Geometric invariants and object recognition. *International Journal of Computer Vision*, 10(3):207–231, 1993.
- [6] David Cyganski and Richard Vaz. A linear signal decomposition approach to affine invariant contour identification. *Pattern Recognition*, 28(12):1845–1853, 1995.
- [7] J. Ben-Arie and Z. Wang. Pictorial recognition of objects employing affine invariance in the frequency domain. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(6):604–618, June 1998.
- [8] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

- [9] E. Aarts and J. Korst. *Simulated Annealing and Boltzman Machines*. John Wiley, 1990.
- [10] Matthew Cooper, Ulf Grenander, Michael Miller, and Anuj Srivastava. Accommodating geometric and thermodynamic variability for forward-looking infrared sensors. In *Presented at Aerosense*. SPIE, April 1997.
- [11] Anuj Srivastava. A bayesian approach to geometric subspace estimation. In *Presented at IEEE SSAP workshop*. IEEE, September 1998.
- [12] Andrew Blake and Andrew Zisserman. *Visual Reconstruction*. The MIT Press, 1987.
- [13] Mila Nikolova, Jerome Idier, and Ali Mohammad-Djafari. Inversion of large-support ill-posed linear operators using a piecewise gaussian mrf. *IEEE Transactions on Image Processing*, 7(4):571–585, April 1998.
- [14] Mila Nikolova. Markovian reconstruction using a gnc approach. *IEEE Transactions on Image Processing*, 8(9):1204–1220, September 1999.
- [15] L.E. Scales. *Introduction to Non-Linear Optimization*. Springer-Verlag, New York, NY, 1985.
- [16] S.D. Silvey. *Statistical Inference*. Chapman and Hall, 1975.
- [17] Louis L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley, Reading, MA, 1991.

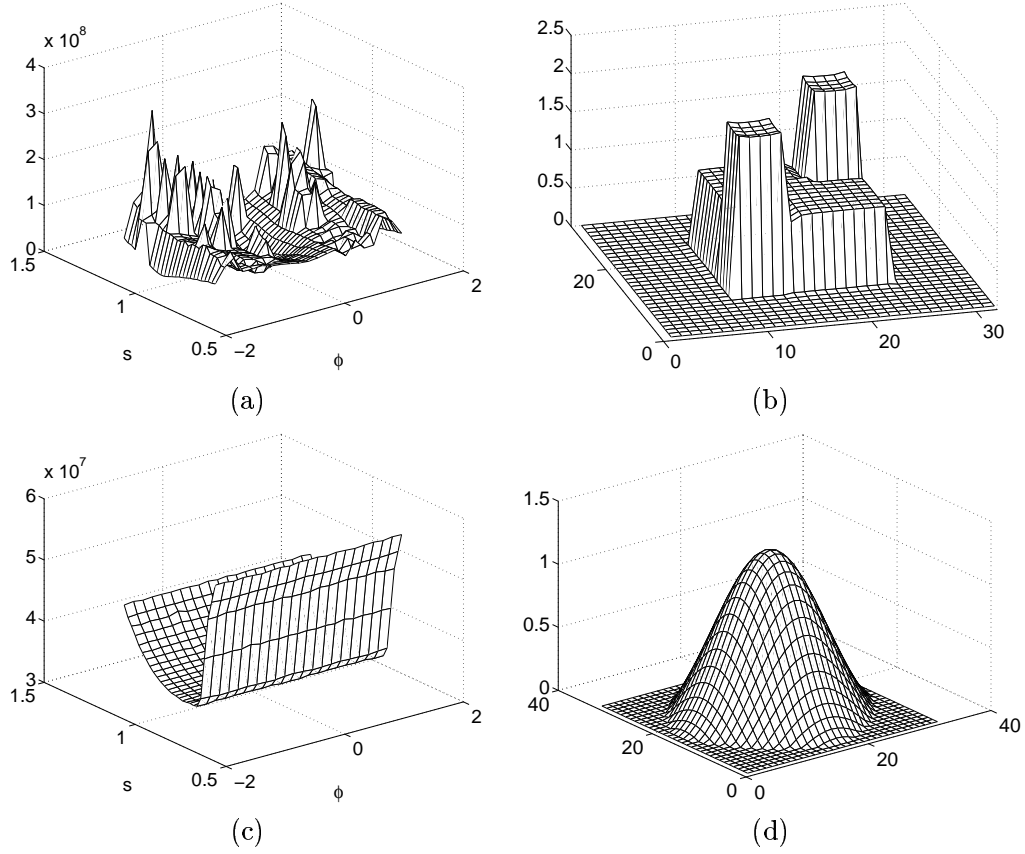


Figure 1: Likelihood surfaces and associated templates. (a) Exact surface (b) exact template (c) smooth approximate surface (d) smooth approximate template.

$ \nabla_{\theta} f(\mathbf{r}; \theta) $	$\frac{\partial \hat{r}_x}{\partial s}$		$\frac{\partial \hat{r}_y}{\partial s}$		$\frac{\partial \hat{r}_x}{\partial \phi}$		$\frac{\partial \hat{r}_y}{\partial \phi}$	
mean	mean	std	mean	std	mean	std	mean	std
1.3534	0.0090	0.0334	0.0090	0.0334	0.0087	0.0406	0.0087	0.0406

Table 1: Values of elements of $\nabla_{\theta} \hat{\mathbf{r}}(\theta)$ for 1095 iterations.

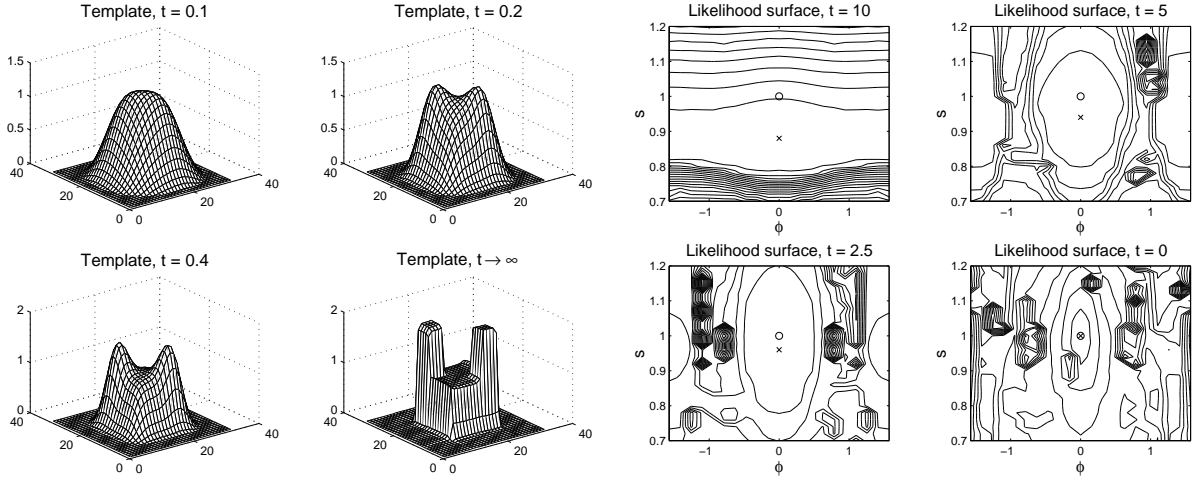


Figure 2: Four templates and contour plots of the associated likelihood surfaces. The exact parameters are marked by the "o", the local minimum found by gradient descent is marked by the "x".

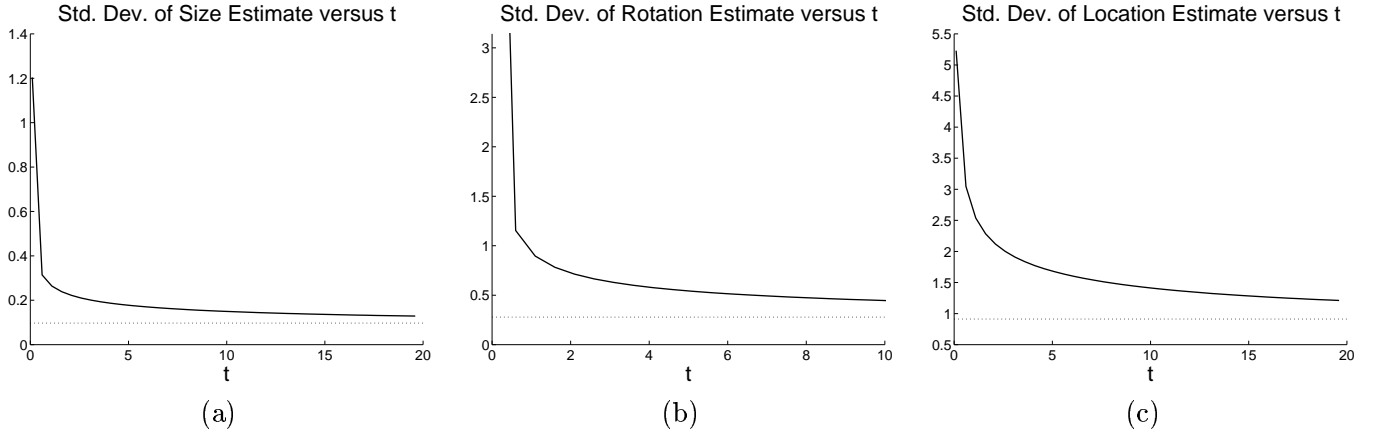


Figure 3: The Cramer-Rao bounds upon the parameter estimates with respect to t for (a) size estimates (b) rotation estimates and (c) location estimates.

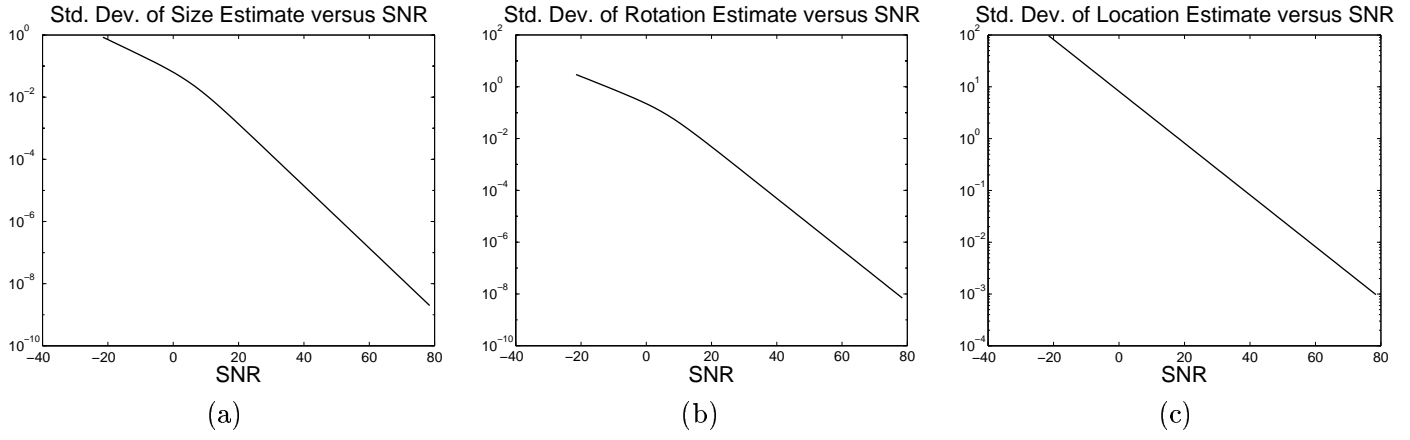


Figure 4: The Cramer-Rao bounds upon the parameter estimates with respect to SNR for (a) size estimates (b) rotation estimates and (c) location estimates.

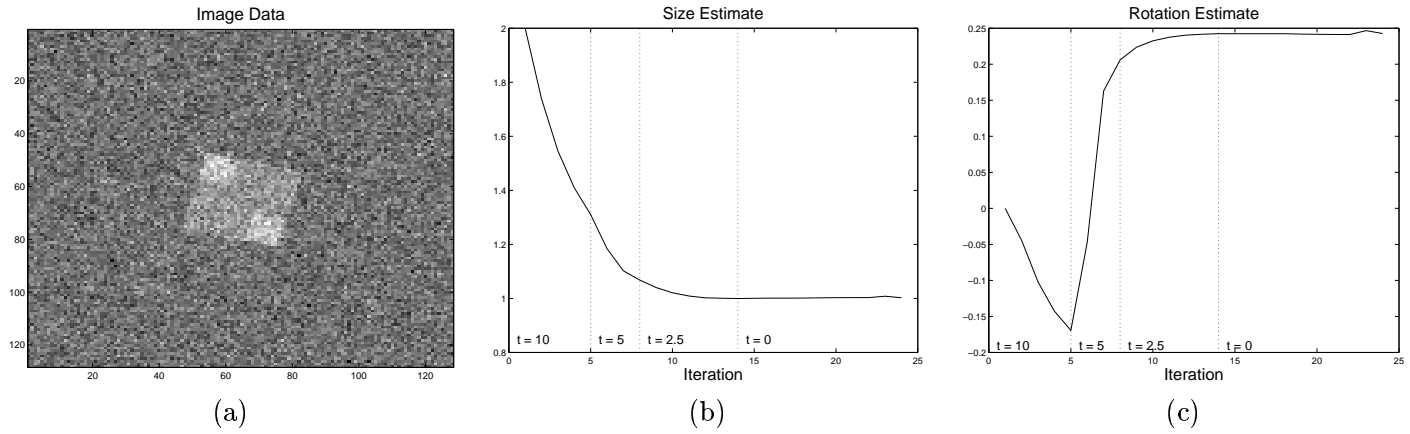


Figure 5: Synthetic example. The two-peak target in 0 dB SNR. Accurate parameter estimation was achieved in 24 iterations with four values of t .

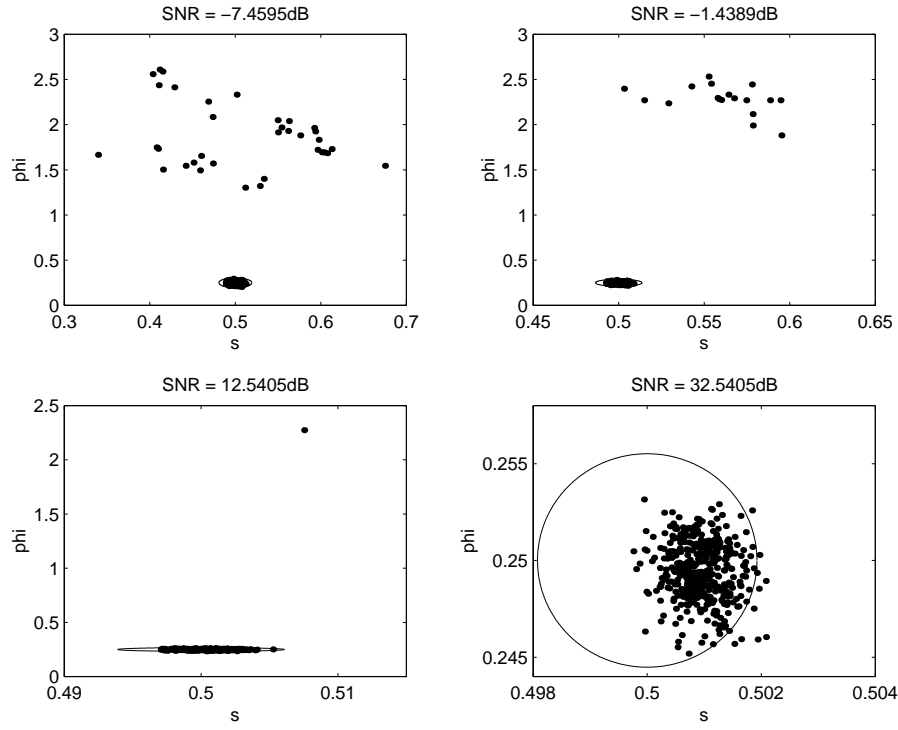


Figure 6: Scatter plots of estimations from 400 simulations of the two-peak example. The ellipse is drawn at 3 times the CRB around the exact parameters, and is used to define “good” estimations.

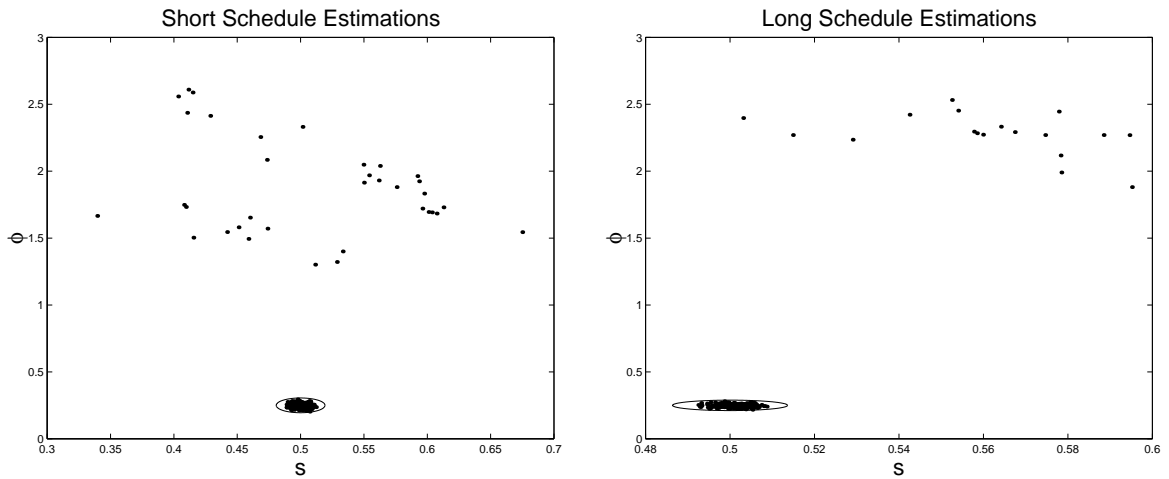


Figure 7: Scatter plots of estimations from 400 simulations of the two-peak example for two schedules, longer schedules result in less misses.

SNR (dB)	Schedule	Size		Rotation		\hat{r}_x		\hat{r}_y	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
-7.4595	short	0.50538	0.020603	0.42739	0.57641	64.8575	3.4557	65.0825	2.5947
-1.4389	short	0.50357	0.013774	0.33951	0.42302	64.9575	2.2921	65.0325	1.5756
12.5405	short	0.50105	0.0025995	0.25451	0.10889	65.025	0.5	64.9825	0.35
18.5611	short	0.50097	0.00095061	0.24918	0.0034543	65	0	65	0
32.5405	short	0.50104	0.00013374	0.24942	0.00046999	65	0	65	0
38.5611	short	0.50105	0.0002985	0.24943	0.0010505	65	0	65	0
-7.4595	long	0.5013	0.024512	0.38909	0.4705	64.785	2.5238	64.905	1.6379
-1.4389	long	0.50221	0.015672	0.32243	0.34788	64.895	1.9425	65.0325	1.1245
12.5405	long	0.50084	0.0014079	0.25412	0.10135	65.03	0.6	64.985	0.3
18.5611	long	0.50085	0.00096719	0.24918	0.0035128	65	0	65	0
32.5405	long	0.50093	0.00043025	0.24944	0.0015144	65	0	65	0
38.5611	long	0.50092	0.00030425	0.24944	0.0010707	65	0	65	0

Table 2: Estimation errors for size, rotation and location.

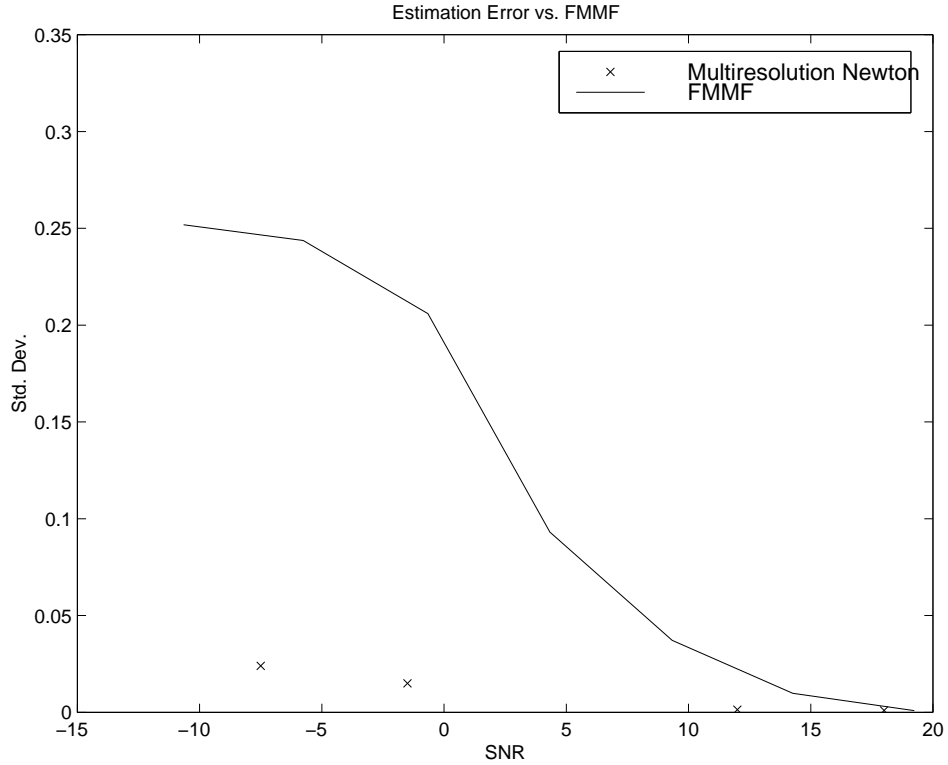


Figure 8: Estimation error for location for 400 simulations using the Fourier Mellin Matched Filter and the current algorithm. The current algorithm performs significantly better at low SNR values.

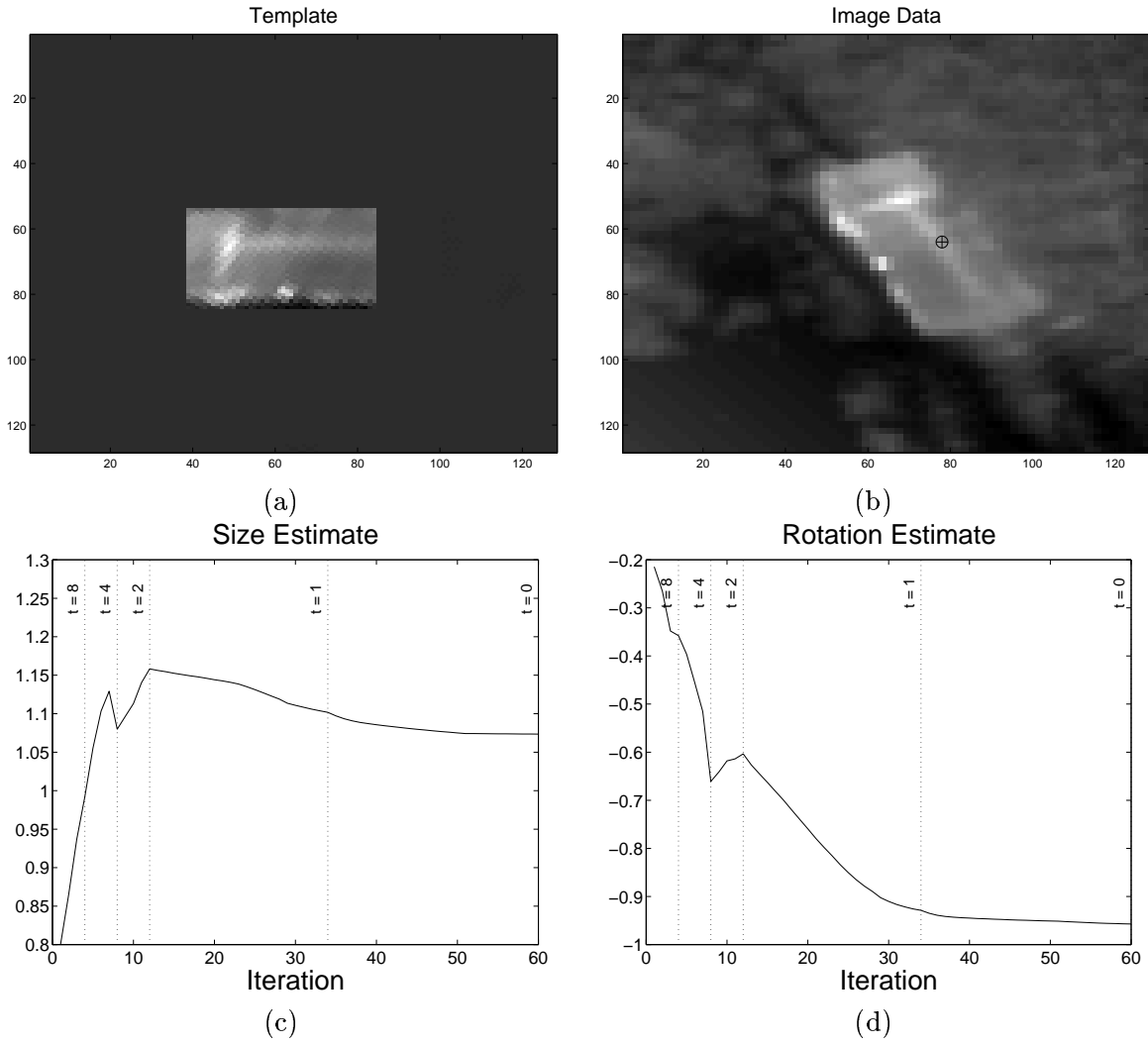


Figure 9: Infrared target example. (a) The template (b) the data with estimated center location (c) the estimated size at each iteration (d) the estimated rotation at each iteration.

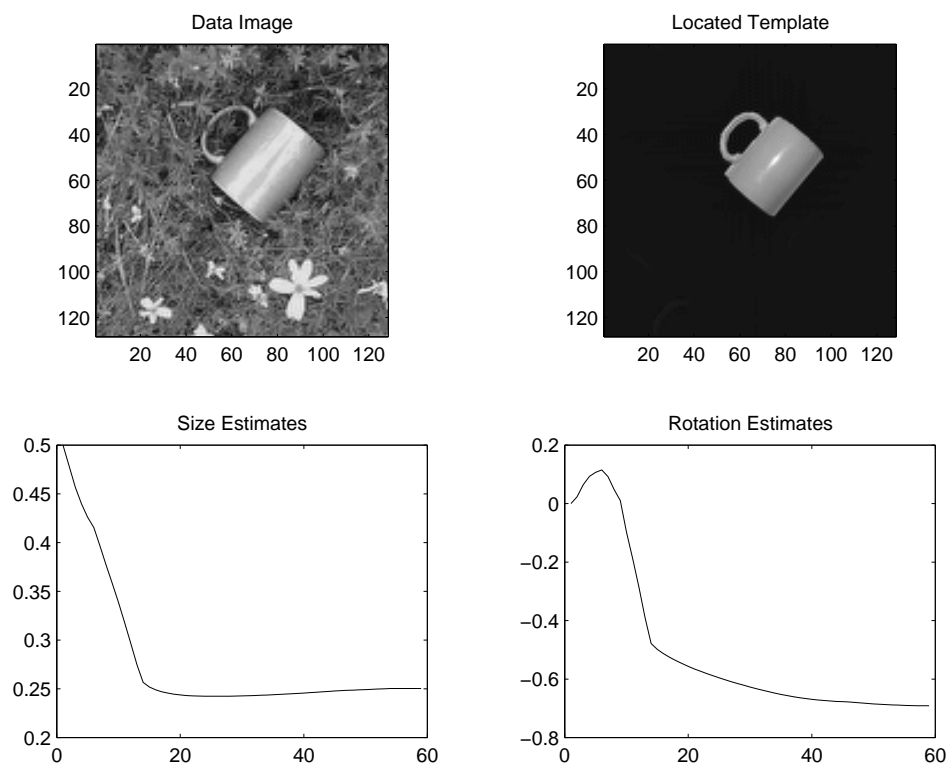


Figure 10: Optical example of a cup in a cluttered scene. (a) The data (b) the template at estimated size, rotation and location (c) the estimated size at each iteration (d) the estimated rotation at each iteration.

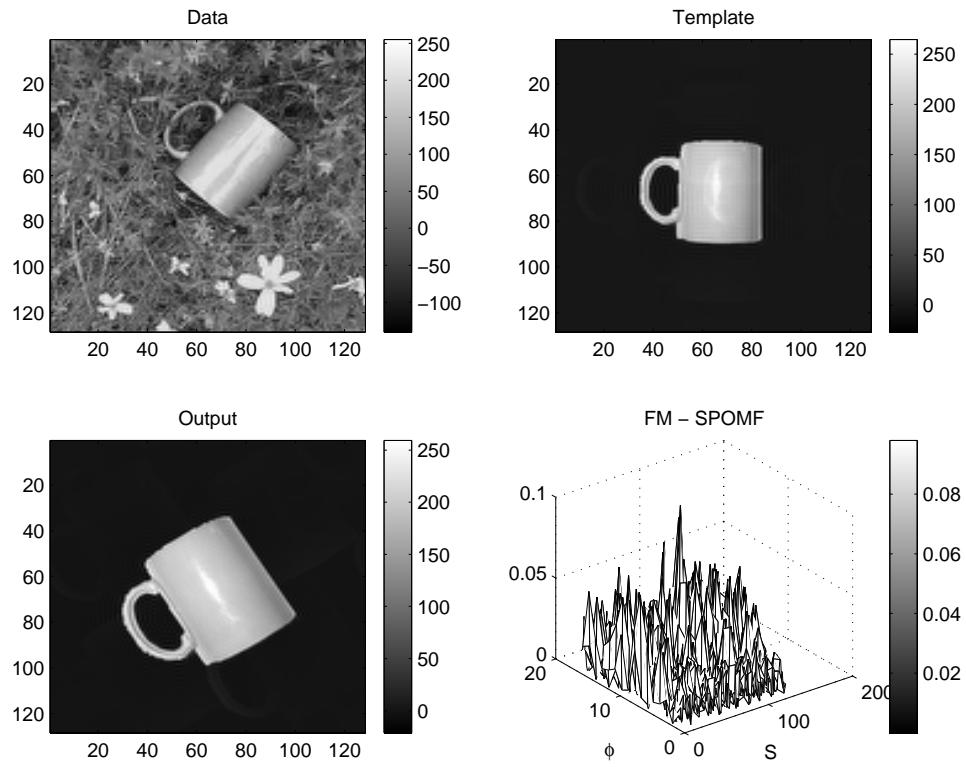


Figure 11: The FMMF of the cup in a cluttered scene. The FMMF fails to accurately estimate the parameters.

Statistical Method to Detect Subsurface Objects Using Array Ground Penetrating Radar Data

Xiaoyin Xu, Eric L. Miller, Carey M. Rappaport and Gary D. Sower

Abstract

We introduce a combination of high-dimensional analysis of variance (HANOVA) and sequential probability ratio test (SPRT) to detect buried objects from an array ground penetrating radar (GPR) surveying a region of interest in a progressive manner. Using HANOVA, we exploit the transient characteristic of GPR signals in the time domain to extract information about buried objects at fixed positions of the array. Based on the output of the HANOVA, the SPRT is employed to make detection decisions recursively as the array moves downtrack. The method is on-line implementable and of low computational complexity. Our approach is validated using field-data from a landmine detection application.

Index Terms

Analysis of variance (ANOVA), GPR mine detection, array signal processing, sequential detection, transient signal analysis.

I. INTRODUCTION

Ground penetrating radar (GPR) is widely used in detecting subsurface objects such as buried landmines, unexploded ordnance, and utility lines [1]. Compared with other subsurface sensing technologies, GPR has a few advantages. First, it is sensitive to changes in all three electromagnetic characteristics of a media, electric permittivity, electric conductivity, and magnetic permeability. Thus GPR is capable of detecting both metallic and non-metallic

X. Xu, E. L. Miller, and C. M. Rappaport are with the Center for Subsurface Sensing and Imaging Systems, Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA. G. D. Sower is with the EG&G MSI Inc., 2450 Alamo Ave. S.E., Albuquerque, NM 87106. This work was supported by an ARO MURI on Demining under Grant DAAG55-97-1-0013

objects. Second, unlike sensors that can only survey an area directly beneath them, GPR can survey an area in front of it [2, 3]. Therefore a GPR system can be used to detect dangerous objects before the system moves over and past them. This can be important for operations such as buried landmine detection and unexploded ordnance remediation.

A typical GPR transmitter/receiver configuration is shown in Fig. 1(a). The system consists of one transmitter and one receiver. The transmitter emits a short pulse of electromagnetic energy and the receiver collects the echo for a certain time period. The exact type of the transmitter and receiver, shape of the electromagnetic pulse, and system setup depend on the specific application of the GPR [1, 4–6]. To improve performance and efficiency, a GPR array is usually employed to sweep a large area in a relatively short time. Fig. 1(b) shows a typical GPR array moving in the x -direction. At every stop of the array, the GPR array operates in the following sequence: 1) the first transmitter radiates a pulse into the ground and then turns off, 2) the first receiver turns on to collect reflected signal, 3) the first receiver turns off after a short time, usually 10 to 20 ns. The above process repeats from every pair of transmitters and receivers and then the GPR array moves to next position. Based on the echoes, the processing objective is to determine if an object is present in the GPR's field of view.

The inherent near-field nature of the GPR detection problem coupled with the fact that the objects of interest are embedded in an inhomogeneous halfspace with a typically rough interface present some significant challenges in the area of GPR signal processing. Indeed assuming one has detailed knowledge of the air-earth interface as well as the electrical properties of the subsurface, just modeling the received signal using, for example, a three dimensional finite difference time domain code, is a daunting task [4, 7]. The use of such a forward model in any form of on-line processing routine where one might need to account

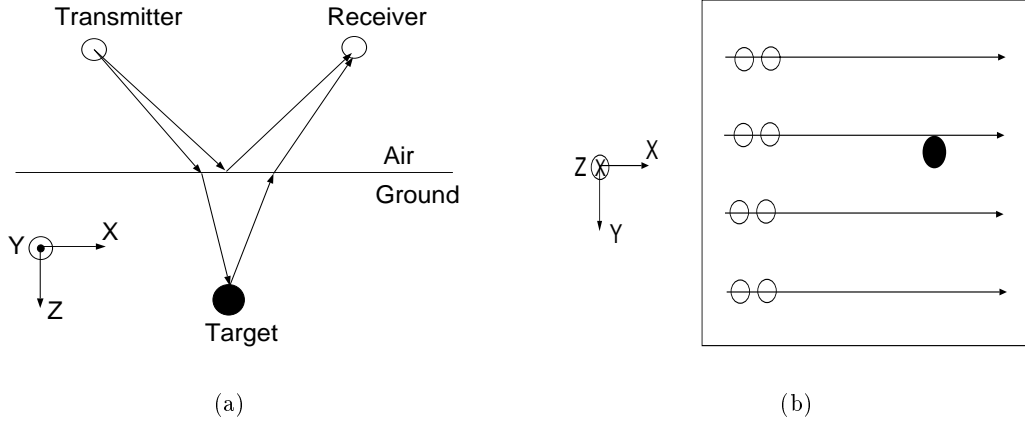


Fig. 1. Setup, a) a single GPR system, b) a GPR array (plane view).

for e.g. unknown ground structure is clearly infeasible at the current time.

Thus, here we consider detection methods which are less computationally demanding with an eye toward approaches that could be used in real-world scenarios. Our interests are in techniques possessing three important characteristics. First, to reflect the manner in which GPR data are acquired and the nature of the GPR mission, the algorithms should be causal in that they need only the data at the current and previous sensor position to determine whether an object is present in the field of view of the sensor. Second, they should be of low complexity. Preferably the number of calculations would grow linearly with the size of the data set. Finally, the processing schemes should be robust to uncertainties in the GPR environment and hence the particular detailed structure of the received signals.

Current signal processing methods with some or all of these characteristics fall into one of three categories. First, pattern matching methods [8] employ techniques such as fuzzy set theory and neural networks. Such methods can be fast but also require extensive training to function well. Second, image-then-detect techniques [9] employ a beamforming or backpropagation approach to build an image of the subsurface which is then post-processed to detect objects. These approaches generally require the data from the full GPR scan to

form an image and are thus not well suited to on-line computations in which information is processed sequentially as the array proceeds down-track. Finally, there has been much work done in statistical signal processing, where one can employ statistical tools to detect objects and examine quantities such as probability of detection and probability of false-alarm [10].

Here we consider a statistical, transient detection approach. By “transient” we mean that the signals of interest are manifest in the GPR data for a small number of sensor positions and for relatively few samples in any received waveform. For example, in Fig. 2 we plot raw observations obtained by one T/R pair from an EG&G GPR system [11], over an M20 metal mine. Each column of this image is a time-series of observations for a given stop of the array. It is seen that the received GPR signal is transient in two ways. First, for each time-series (i.e. for each column of the image) containing an object signal, the signal appears only in a brief window, roughly from samples 300 to 700. The reason is that the object signal always comes after the signal arising from the bounce off of the air-ground interface and attenuates quickly in lossy media. Second, the object signal shows up only at a few down-track positions of the GPR array, specifically locations 15 through 25. In both cases, the appearance of object signal changes the mean value of the data. Our method for object detection then is based on detecting change in this mean first in the cross-track direction and then in the down-track direction.

More specifically our approach consists of two parts. First, at each down-track position of the array, we process the data among all T/R pairs to generate one test statistic. We use high-dimensional analysis of variance (HANOVA) to test whether the data consists of reflected signal from a buried object. The HANOVA is a generalized version of standard analysis of variance (ANOVA), which is a method for testing hypothesis about means of random vectors [12, 13]. Second, a sequential probability ratio test (SPRT) is applied to

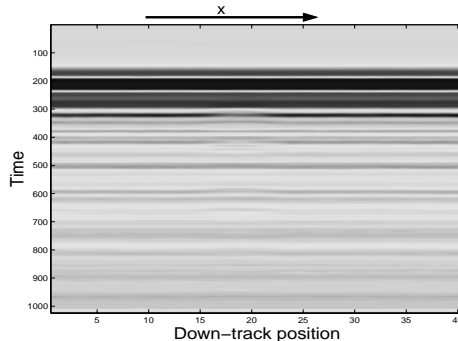


Fig. 2. Observation from one T/R pair, for a metal mine M20 buried at about position 20. Unit in down-track position is about 7.6 cm. Unit in time axes is 0.02 ns.

process the statistic of the HANOVA as the array moves down-track. The SPRT is a recursive statistical hypothesis testing technique that provides early indication of the onset of changes in a time series. The output of the SPRT is compared with a threshold. If it exceeds the threshold, a detection is declared, otherwise, the GPR array moves one more step down-track and new data are collected and processed in the above manner [14].

As explained in greater detail below, our approach does in fact satisfy the three requirements we discussed previously. It is causal and has computational complexity that grows linearly with the size of the data. Moreover, we show through real-data examples that it is robust, requiring little in the way of training and able to successfully address the object detection problem for a number of GPR systems operating in a wide range of environments. We do stress here that the algorithm in this paper is intended *only* to find anomalies beneath the GPR array and *not* to solve the far more challenging classification problem. Thus, from a practical perspective our approach will serve well as an efficient “pre-screener” in a larger automatic target detection algorithm suite. Finally, our method is motivated by landmine detection using GPR, however it can also be used in other detection application, such as laser-induced acoustic subsurface objects detection [15].

The paper is organized as follows. Section II discusses the problem formulation and our

method. Section III gives some examples of using the method in different situations. Field data from different radar configurations and test sites are used to show how the algorithm works. Conclusion and direction of future work are given in Section IV.

II. PROBLEM FORMULATION AND ALGORITHM

To begin, we consider a single GPR T/R pairs as shown in Fig. 1(a). After each transmission, the receiver collects an echo for a certain amount of time. Depending on the presence of an object, there are either two or three components in the echo. One is measurement noise, assumed to be white and Gaussian. Another is background, i.e., “nominal” signal observed in object-free regions. The third component is object signal, reflection from a buried object.

For the GPR array shown in Fig. 1(b), assume we have M GPR Transmitter/Receiver (T/R) pairs surveying an area in N steps, the task is to use present and previous array measurement to detect buried mines as the array moves down-track. At each down-track position, we model the array detection problem in a typical hypotheses testing framework [14],

$$H_0 : \quad \text{there is no object,}$$

$$H_1 : \quad \text{there is an object.}$$

The null hypothesis H_0 means that there is no buried object in the field of view of the GPR array, so the total received signal is comprised of nominal background and measurement noise. By nominal background, we mean any portion of the received waveform not sensor noise and not arising from the interaction of the transmitted pulse with the object. Reflection from the air-ground interface is the dominant component of this part of the signal. The alternative hypothesis H_1 indicates that there is buried object so that the received signal consists of nominal background, measurement noise, and an object signal.

In this paper we assume that the nominal background signal has been removed via a

preprocessing stage. The most used background removal methods include casual methods, such as subtraction of a moving average from the observation [16], and non-causal methods, such as subtraction of an ensemble average from the observation [17, 18]. Causal methods use data from previous and present collection, non-causal methods use data from previous and future collection. In this paper, a moving average (MA) filter is used to eliminate the nominal background.

In practice, the receiver collects time-samples of the reflection and stores it as a vector. For convenience, we use vector notation in our discussion, i.e., $\mathbf{y}(m, n)$ is a column vector representing observation of the m th T/R pair at the n th down-track position. The length of $\mathbf{y}(m, n)$ is K , the number of samples in time. Fig. 3 shows the received signal after the nominal background removal.¹ We then have the hypothesis test

$$\begin{aligned} H_0 : \quad \mathbf{y}(m, n) &= \mathbf{v} \\ H_1 : \quad \mathbf{y}(m, n) &= \mathbf{s}(m, n) + \mathbf{v} \end{aligned} \tag{1}$$

where $m = 1, \dots, M$, $n = 1, \dots, N$ are positions of GPR, $\mathbf{s}(m, n)$ is the assumed signal due to presence of buried object, \mathbf{v} is assumed to be a white Gaussian noise with a zero mean, and covariance matrix $\sigma_v^2 \mathbf{I}$, where \mathbf{I} is the identity matrix of size K and independent of (m, n) .

The statistical assumptions about \mathbf{v} are not strictly accurate in describing the noise in a GPR signal. For example the background removal process will not be perfect leaving a component of correlated “clutter” in the data which may or may not possess Gaussian statistics. Despite the mismatch, the use of the additive white Gaussian noise model is useful for a number of reasons. This model allows us to develop an algorithm for object

¹ For the purpose of illustration, in this section we use field data from a buried metal mine to illustrate clearly the concept under consideration. Examples which demonstrate better the utility of our approach on more challenging problems, including buried plastic mines, are given in Section III.

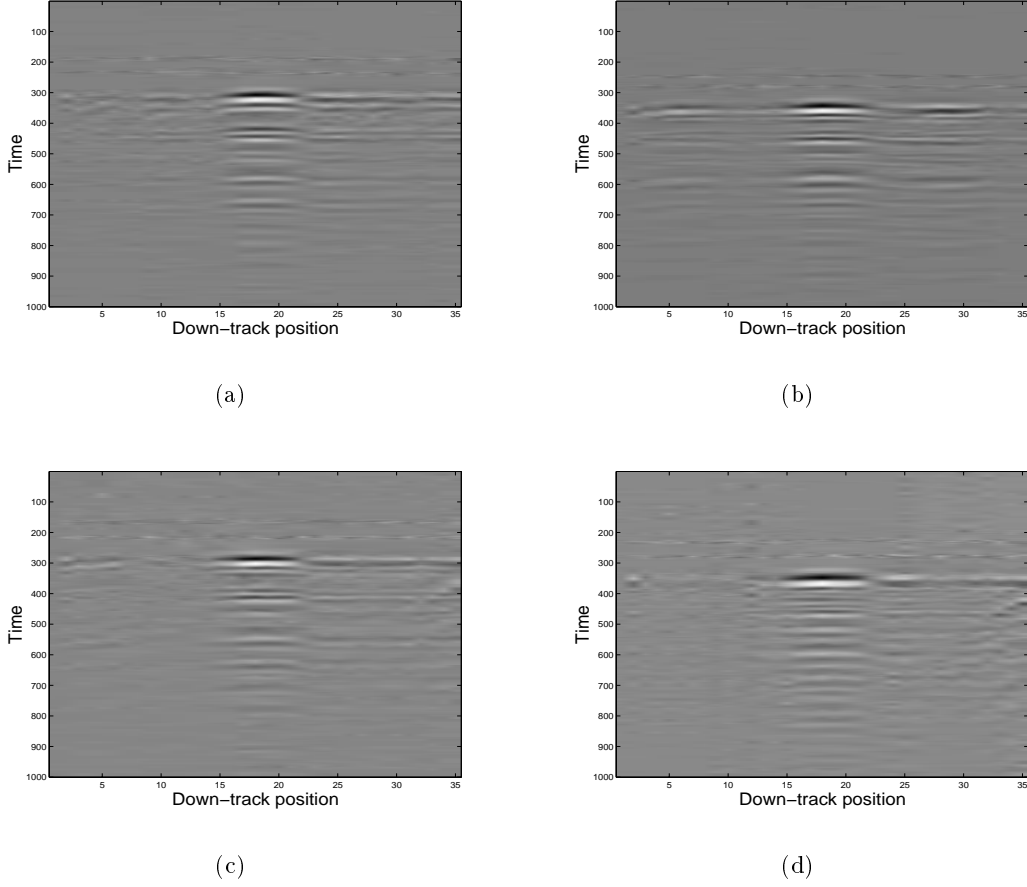


Fig. 3. Signals from four T/R pairs, after background removal, a) pair 1, b) pair 2, c) pair 3, d) pair 4. Unit in time axes is 0.02 ns.

detection which is firmly rooted in Gaussian-based statistical decision theory and which can be generalized in the future for more complex noise processes. Moreover, the complexity of such algorithms is quite low making them well suited for real-world implementation. Finally, test results in Section III from real field data demonstrate that the method is quite effective in detecting objects. Thus, the Gaussian noise model is shown to work in practice. While it may be interesting to explore other, more accurate models for the sensor noise to determine for example what can be gained in terms of performance and what would be lost in terms of computational complexity, such an effort is beyond the scope of the work in this paper.

Based on the previous discussion, after background removal the hypothesis test in (1) may be written as

$$\begin{aligned} H_0 : \quad & \mathbf{y}(m, n) \sim N(\mathbf{0}, \sigma_v^2 \mathbf{I}) \\ H_1 : \quad & \mathbf{y}(m, n) \sim N(\mathbf{s}(m, n), \sigma_v^2 \mathbf{I}), \quad m = 1, \dots, M, \quad n = 1, \dots, N \end{aligned} \quad (2)$$

where the notation $\mathbf{y} \sim N(\bar{\mathbf{x}}, \mathbf{R})$ indicates that \mathbf{y} is distributed as a Gaussian random vector with mean $\bar{\mathbf{x}}$ and covariance matrix \mathbf{R} .

As stated in the Introduction, we take a two-step approach to the processing of $\mathbf{y}(m, n)$. First for each n we use the HANOVA procedure to generate a single test statistic, $Y(n)$, from the data from all T/R pairs. Second, a recursive, sequential detection scheme is employed to process $Y(n)$ as we proceed down track in order to determine where objects are present.

A. Cross-track Processing

We begin by discussing the use of HANOVA to process data in the cross-track direction. HANOVA is a generalized version of analysis of variance (ANOVA). ANOVA is a body of methods to analyze the data with a view to test hypotheses about the effects of one or more factors [19]. To review the basics of ANOVA, we follow the notation established above for the GPR problem and for simplicity assume we have one data vector of size $K \times 1$ from a single T/R pair, $\mathbf{y} \sim N(\mathbf{s}, \sigma^2 \mathbf{I})$ and we wish to test $H_0 : \mathbf{s} = \mathbf{0}$ (i.e., no object) vs. $H_1 : \mathbf{s} \neq \mathbf{0}$ (i.e., an object present)². Standard ANOVA is essentially an “energy detection” scheme [12] where we estimate \mathbf{s} by \mathbf{y} , generate the test statistic $Y = \|\mathbf{y}\|^2$, and compare Y to a threshold, γ . If Y exceeds the threshold, H_1 is chosen, else H_0 is selected. The probability of detection of the standard ANOVA is

$$P_d(H_1|H_1) = Q\left(\frac{\gamma - \frac{\|\mathbf{s}\|^2}{\sigma^2 \sqrt{2K}}}{\sqrt{1 + \frac{2\|\mathbf{s}\|^2}{\sigma^2 K}}}\right) \quad (3)$$

² For notational simplicity, we drop the explicit dependence of all quantities on m and n in this discussion

where γ is the test threshold decided by setting an acceptable probability of false-alarm under H_0 and Q is the complementary cumulative distribution function and is strictly decreasing [20]

$$Q(Y) = \int_{\gamma}^{\infty} P_1(Y) dY. \quad (4)$$

Recently, Fan [12] and Fan and Lin [13], have noted that the performance of ANOVA suffers for problems when the signal of interest is limited to a window of the observation vector. The reason is that a full dimensional test loses its power due to accumulation of stochastic noise. To see why, suppose \mathbf{s} is different from $\mathbf{0}$ only for say the first k_0 samples of the full observation vector. Then on average as $K > k_0$ goes large, $\sum_{k=1}^K [\mathbf{s}]_k^2 / \sigma^2 \sqrt{2K}$ decreases due to the accumulation of zero mean noise samples and the term within the parenthesis of (3) increases, thus reducing P_d . Therefore, for higher probability of detection, we would like to confine the test on a window mostly containing the signal of the observation vector. The window we choose is a box window \mathbf{w} , defined as

$$[\mathbf{w}]_k = \begin{cases} 1, & k = k_1, \dots, k_2 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where $1 \leq k_1 < k_2 \leq K$. The k_1 and k_2 are chosen in a preset manner, as discussed later in this section. Multiplying each element in \mathbf{y} by the corresponding element of \mathbf{w} gives the windowed \mathbf{y}_w

$$[\mathbf{y}_w]_k = [\mathbf{y}]_k \times [\mathbf{w}]_k, \quad k = 1, \dots, K. \quad (6)$$

To demonstrate the utility of HANOVA, we test the time-series shown in Fig. 4(a). We choose to test the vector at its full dimension k_1 in (5) is 1 and $k_2 = 1000$, and two windowed sub-dimensions (each containing fewer and fewer noise components) $k_1 = 100$ and $k_2 = 900$,

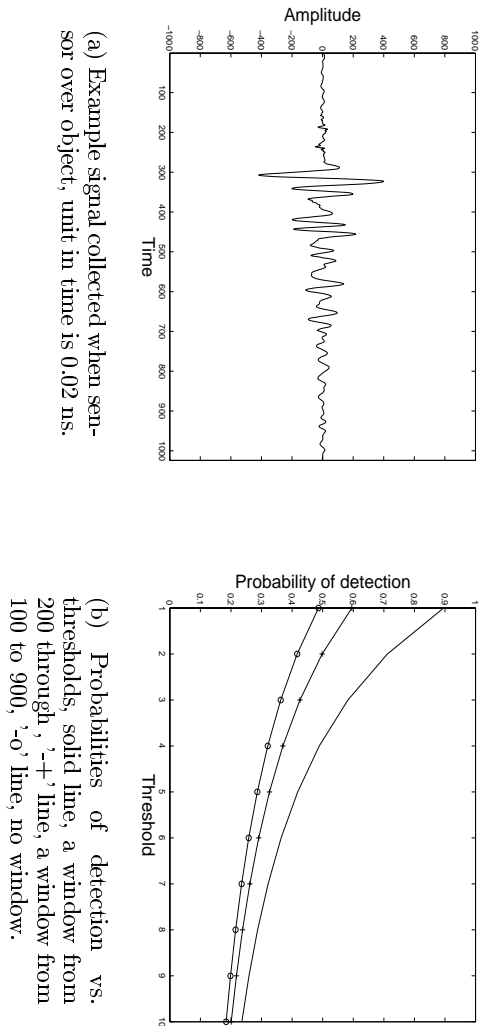


Fig. 4. Motivating example for use of HANOVA rather than ANOVA. (a) displays sample data collected by a GPR when over an object. The transient nature of the relevant portion of the signal is clear. Using this signal, in (b) we display the decreasing detection rates associated with including increasing numbers of “noise” samples in the processing.

and $k_1 = 200$ and $k_2 = 800$. From Fig. 4(b), it is seen that by setting the window properly, higher probability of detection is gained at different levels of detection thresholds, γ in (3). It demonstrates that when signal is not “full-dimensional”, looking for a window of signal-rich sub-dimensions to test will increase the probability of detection.

When the observation is a sequence of high-dimensional vectors whose components are mostly noise, as is the case for our GPR problem, it is desirable to adaptively choose the window to maximize the probability of detection. This kind of method of reducing a full dimensional test to a windowed version is called HANOVA [13]. Fan’s original work was limited to problems in which the first k_0 dimensions are believed to be signal-rich and used in HANOVA, with k_0 found from the data. Here we consider a generalization of Fan’s work to take into account the fact that for the GPR problem the transient signal is significant over a window not generally starting with the first dimension but in the middle of the observation vector. Moreover, this window will vary with (m, n) .

To choose this window we note that (3) indicates that the probability of detection achieves its maximum value when the term inside the parenthesis is minimized. Equally, one wants to maximize the quantity

$$\arg \max_{k_1, k_2} \frac{\sum_{k=k_1}^{k_2} [\mathbf{s}(m, n)]_k^2}{\sigma_v^2 \sqrt{k_2 - k_1 + 1}} - \sqrt{k_2 - k_1 + 1} \quad (7)$$

where $k_1 < k_2$ and $k_1, k_2 \in 1, \dots, K$. The difficulty for us is that in general, the precise structure of \mathbf{s} is not known. Hence, we use the data to form an estimate of \mathbf{s} as follows. Assume we are at the n th stop, then we estimate \mathbf{s} by the mean value of the previous l vectors

$$\hat{\mathbf{y}}(m, n) = \begin{cases} \frac{\sum_{j=1}^n \mathbf{y}(m, j)}{j}, & n = 1, \dots, l \\ \frac{\sum_{j=n-l}^n \mathbf{y}(m, j)}{l}, & n > l \end{cases} \quad (8)$$

where $m = 1, \dots, M$ and the corresponding window $\mathbf{w}(m, n)$ is decided based on $\hat{\mathbf{y}}(m, n)$ as k_1, k_2 are defined by (7). More will be said about choosing a proper l in Appendix A.

Rather than looking for the optimal window by searching over all k_1 - k_2 pairs, we pursue a suboptimal, but more efficient two-stage approach. First, we fix k_1 as 1, incrementally increase k_2 , and stop when (7) is maximized. Thus we determine the end point of the window k_2 . Starting from k_2 , working backward toward the first point, we similarly determine the starting point of the window, k_1 . Both searching steps can be computed in linear complexity, it takes $o(K)$ steps to find the k_2 and $o(k_2)$ steps to find the k_1 ³. In summary the steps for looking for windows at the n th stop of the GPR array are given in Fig. 5.

Having determined the window at the position (m, n) , the next stage of processing is to generate a single detection statistic at stop n . Here we generalize HANOVA to multiple vector observations, via

$$Y(n) = \frac{1}{\sigma_v^2} \sum_{m=1}^M \|\mathbf{y}_w(m, n)\|^2. \quad (10)$$

³ The notation $o(K)$ means that the computational complexity grows slower than or equally fast as K increases.

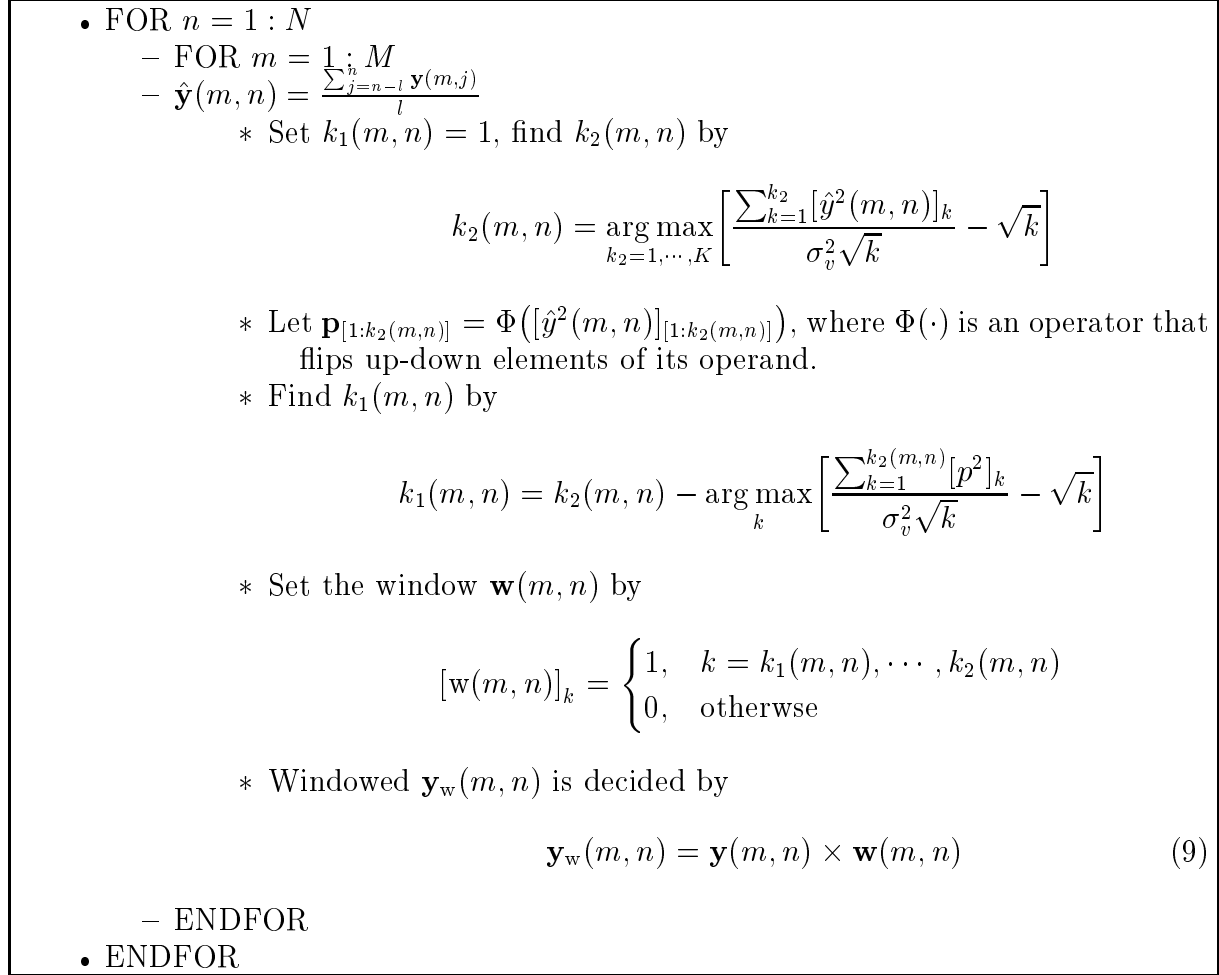


Fig. 5. Steps of deciding window $\mathbf{w}(m, n)$ and $\mathbf{y}_w(m, n)$.

Note $\mathbf{y}_w(m, n)$ can be of different length because of different window applied. Fig. 6(a) shows the result of applying HANOVA to the data in Figure 3. Where the HANOVA output is high, so too is the likelihood of an object being present. Thus in Fig. 6(a), the object is clearly detectable. More examples involving different types of objects will be given in Section III.

B. Down-track Processing

While HANOVA detects statistical significance at one stop of the array, it does not capture the object signal structure seen as the array moves down-track. To improve detection performance, we employ a sequential detection scheme to process $Y(n)$ recursively as n

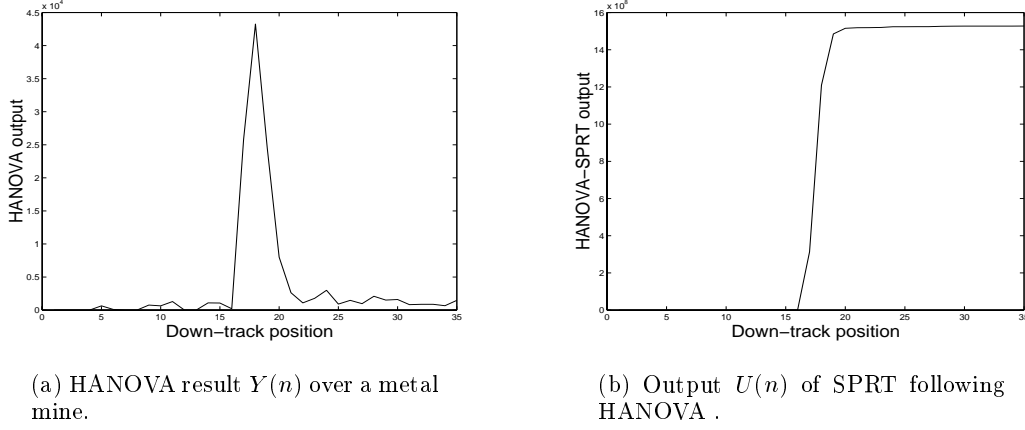


Fig. 6. HANOVA and SPRT processing results for metal mine data shown in Fig. 3

increases in order to identify the transient signal arising from the mine [14,21]. Specification of this sequential probability ratio test (SPRT) begins by noting that under our models $Y(n)$ takes on a χ^2 distribution under both H_0 and H_1 . Standard statistical analysis [12] yields

$$\begin{aligned}
 H_0 : \quad Y(n) &\sim \chi^2_{\sum_{m=1}^M \Delta k(m,n)}(0) \\
 H_1 : \quad Y(n) &\sim \chi^2_{\sum_{m=1}^M \Delta k(m,n)}(\delta^2(n))
 \end{aligned} \tag{11}$$

for $n = 1, \dots, N$ where the notation $x \sim \chi_p^2(\delta^2)$ indicates that the random variable x is distributed according to a χ^2 law of order p and non-centrality parameter δ^2 [20] and $\Delta k(m, n) = k_2(m, n) - k_1(m, n)$ is the length of the (n, m) th window. For the GPR problem it is easy to show that

$$\delta^2(n) = \frac{1}{\sigma_v^2} \sum_{m=1}^M \|\mathbf{s}(m, n) \times \mathbf{w}(m, n)\|^2. \tag{12}$$

For our problem, the length of each window, $\Delta k(m, n)$, is large (on the order of hundreds) and the central limit theorem permits us to approximate the χ^2 distribution using a Gaussian

distribution [20]. We then have

$$\begin{aligned} H_0 : Y(n) &\sim N(\mu_0, \sigma_0^2) \equiv N\left(\sum_{m=1}^M \Delta k(m, n), 2 \sum_{m=1}^M \Delta k(m, n)\right) \\ H_1 : Y(n) &\sim N(\mu_1(n), \sigma_1^2(n)) \equiv N\left(\sum_{m=1}^M \Delta k(m, n) + \delta^2(n), 2 \sum_{m=1}^M \Delta k(m, n) + 4\delta^2(n)\right). \end{aligned} \quad (13)$$

At stop n , the log likelihood ratio for the hypothesis testing problem in (13) is

$$u(n) = \ln \frac{p_n(Y(n))}{p_0(Y(n))}, \quad n = 1, \dots, N \quad (14)$$

where $p_n(Y(n))$ is the PDF of $Y(n)$ evaluated at the n th stop under H_1 and $p_0(Y(n))$ is the PDF of $Y(n)$ evaluated under H_0 . Under H_0 , μ_0 and σ_0^2 are estimated using data from an object-free area. Therefore, for this algorithm, the GPR array must start by collecting data in a calibration region to initialize these variables. Under H_1 , one difficulty with generating $u(n)$ is that $\mu_1(n)$ and $\sigma_1^2(n)$ are typically not known *a priori* since the underlying $\mathbf{s}(m, n)$ are not assumed known. It turns out that we only need to estimate $\mu_1(n)$, and $\sigma_1^2(n)$ can be found from the following relation

$$\begin{aligned} \sigma_1^2(n) &= 2 \sum_{m=1}^M \Delta k(m, n) + 4\delta^2(n) \\ &= 2 \sum_{m=1}^M \Delta k(m, n) + 4 \left[\mu_1(n) - \sum_{m=1}^M \Delta k(m, n) \right] \\ &= 4\mu_1(n) - 2 \sum_{m=1}^M \Delta k(m, n). \end{aligned} \quad (15)$$

At the n th stop, we estimate the mean of $Y(n)$ by its maximum likelihood estimator $\mu_1(n) = Y(n)$.

The *sequential probability ratio test* statistic $U(n)$ is a cumulative sum, changing with the acquisition of each new $u(n)$

$$U(n) = \max(0, U(n-1) + u(n)). \quad (16)$$

- $U(1) = 0$
- FOR $n = 2, \dots, N$
 - $\mu_1(n) = Y(n)$
 - Form $\sigma_1^2(n)$ according to (15)
 - Form $u(n)$ according to (14)
 - $U(n) = \max(0, U(n-1) + u(n))$
 - IF $U(n) > \alpha$, declare object, set $U(n) = 0$, ENDIF
- ENDFOR

Fig. 7. Sequential processing.

Because subsurface object detection is a binary hypothesis testing problem, e.g., we are only interested in knowing whether there is a buried object, the SPRT statistic is bounded from lower bound, zero. When $U(n-1) + u(n)$ is negative, $U(n)$ is reset to zero. For a preset threshold α , the SPRT will make one of two decisions at each n

$$U(n) \geq \alpha \Rightarrow \text{choose } H_1$$

$$U(n) < \alpha \Rightarrow \text{take another observation.}$$

The sequential detection is then essentially a repeated SPRT [22] and summarized in Fig. 7. Fig. 6(b) shows the sequential test statistic when the SPRT is applied to the data in Fig. 6(a). Because the SPRT in (16) has the form of a modified “integrator,” a typical time series for the SPRT statistic takes a step-like form. The larger and sharper the step, the more likely it is that a target is present. At the position where there is an object, the sequential test statistics has a clear upward change again indicating the existence of an object at about position 16.

III. EXAMPLES

In this section we use field data as examples to illustrate the performance of our method. The field data are collected by both single GPR and GPR arrays at different test sites. For each data set, we compare the results from using standard ANOVA, HANOVA, ANOVA

followed by SPRT, and HANOVA followed by SPRT. Comparison indicates that generally HANOVA performs better than ANOVA, and with SPRT, both ANOVA and HANOVA make fewer false-alarms. In other words, HANOVA with SPRT gives the best receiver-operating characteristics, as we shall see later in this section.

At first, we apply our method on data collected by single GPR at different test sites. Some data are taken under relatively favorable condition, while most are from more hostile test sites which involve rough ground surface and other clutter. Fig. 8 compares results of ANOVA and HANOVA on a buried steel object at position 50. For comparison, the outputs of ANOVA and HANOVA are normalized to one. It is observable that while both methods detect the object easily, the HANOVA is better in suppressing noise output where there is no object, e.g., at position 1 through 40 and 60 through 100, Fig. 8(c) and (d). Fig. 9 shows the results from detecting a plastic mine, M19, at position 50. Again, the HANOVA performs better in suppressing noise. At positions 20 through 40, the HANOVA creates a much lower noise level than the ANOVA does. Similarly, the HANOVA produces a cleaner output at the end of the run.

Fig. 10 shows the results of ANOVA and HANOVA in detecting an anti-tank mine, TM62, from a very “noisy” data set. The mine is buried at position 60. Outputs of both HANOVA and ANOVA consist of the correct detection and some false alarms. The HANOVA maintains a better performance than the ANOVA in the sense that, for a given detection threshold, the HANOVA would generally have a smaller number of false-alarms. For the HANOVA, no false alarms will be declared for a threshold greater than 0.5, while for the ANOVA, the threshold must be set above 0.8 to avoid making a wrong decision. Between threshold 0.5 and 0.8, the ANOVA will make two false-alarms while the HANOVA has zero false-alarm.

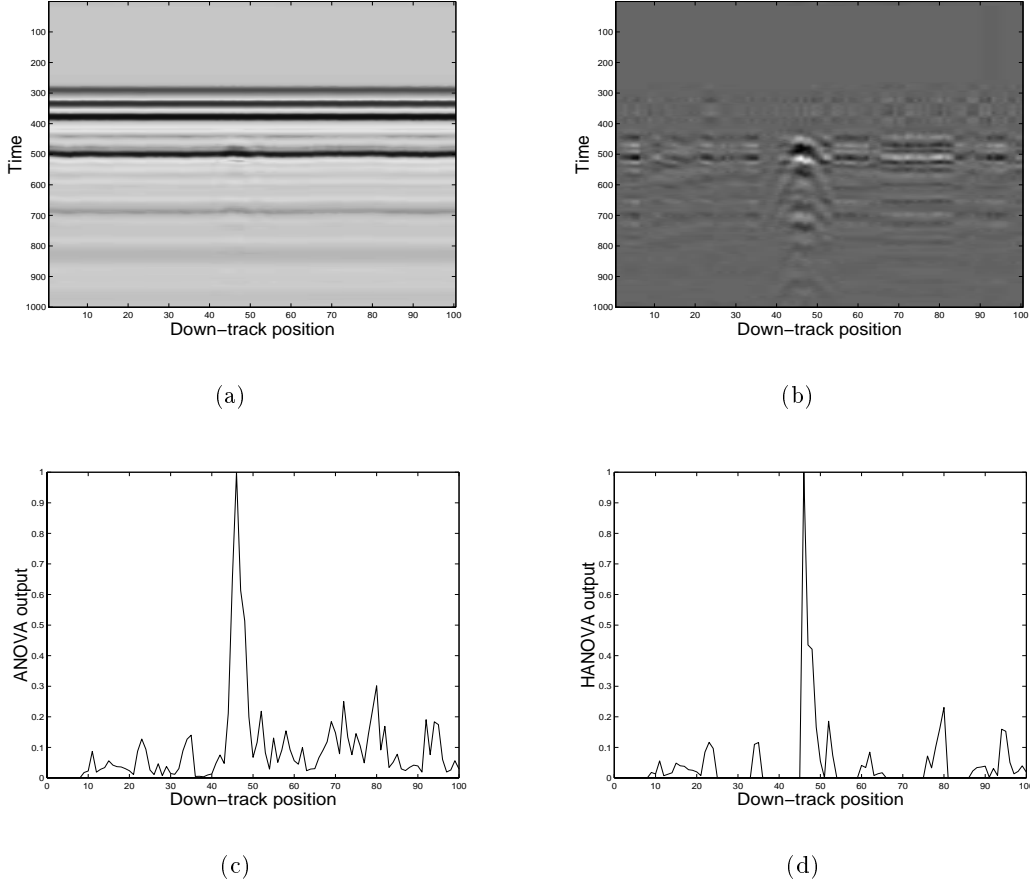


Fig. 8. Results of a single GPR measurement above a steel object around position 50, a) raw observation, b) observation after nominal background removal, c) ANOVA output, d) HANOVA output. In a) and b), each unit in time axes is 0.02 ns.

Next, by comparing the outputs of the SPRT in the above three examples, we see that sequential processing generally smoothes the output and generates fewer false-alarms than by using ANOVA (or HANOVA) only, Fig. 11. In all three examples, SPRT following HANOVA performs better than SPRT following ANOVA, in the sense that the output is more leveled off at object-free area and the jump at the position of the buried object is sharper.

To study the receiver operating characteristic (ROC) of the method, we test our method on multiple runs of different type of targets. Fig. 12(a) shows the ROC curves of ANOVA and HANOVA to detect metallic objects. The objects include metallic mines such as TM15,

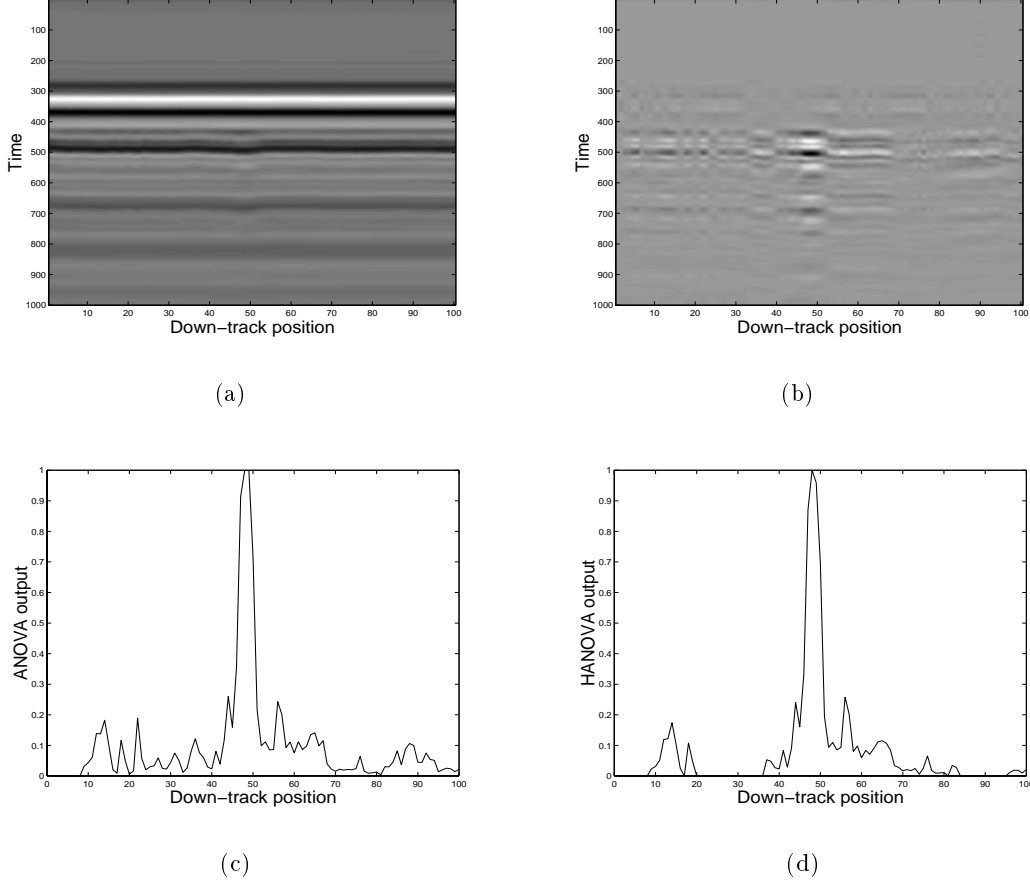


Fig. 9. Comparison between the ANOVA and HANOVA, a) raw data over an M19, anti-tank mine, buried at position 50, b) demeaned data, c) result of the ANOVA over the M19, d) result of the HANOVA.

TM46, and PMN. Fig. 12(b) shows the ROC of ANOVA-SPRT and HANOVA-SPRT. Compared with Fig. 12(a), SPRT improves the performance of both ANOVA and HANOVA. In generating these curves a correct identification of any of the objects was taken to be a “detection” whether or not the object itself was a mine. Indeed, as noted in the Introduction, the algorithm in this paper is intended *only* to detect the presence of objects below the array and not to solve the classification problem. Still, given the “real-world” conditions under which the data were taken, the low false alarm rates here point to the robustness of our approach.

Next, we compare the performance of ANOVA, HANOVA, ANOVA-SPRT, and HANOVA-
April 29, 2001—3 : 44 pm DRAFT

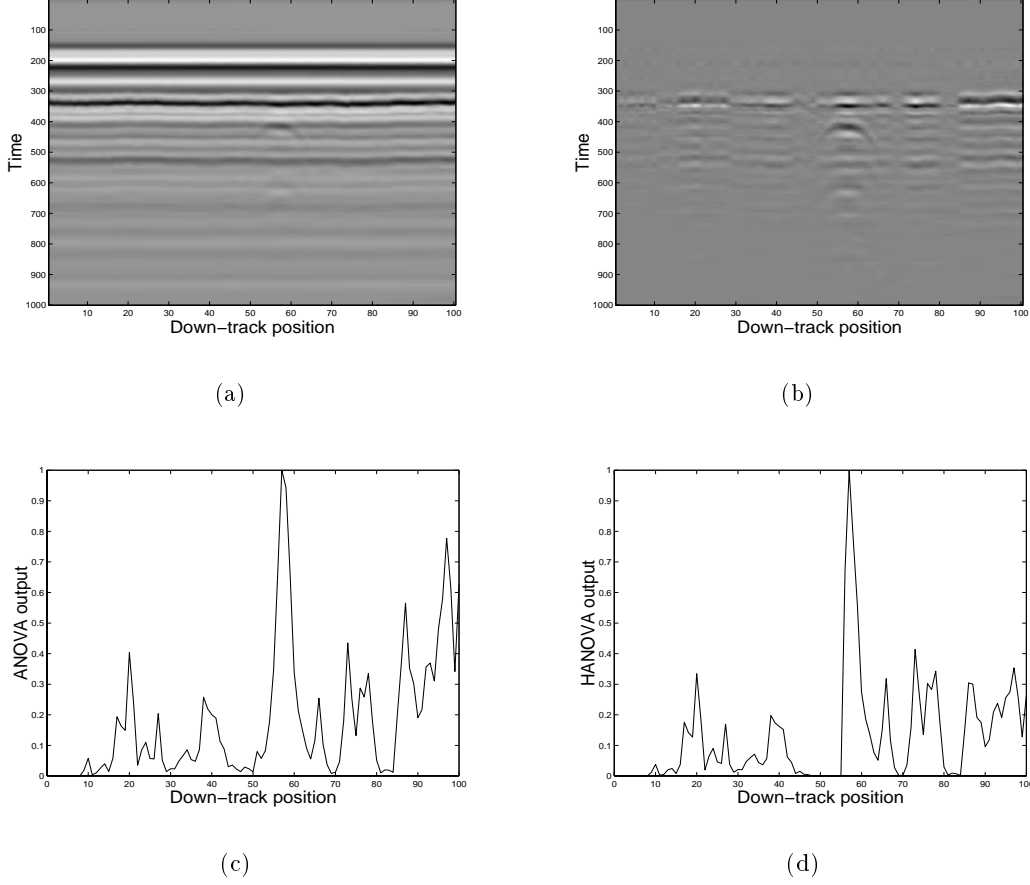


Fig. 10. More comparison between the ANOVA and HANOVA, a) raw data over a TM62, anti-tank mine, buried to the side of the track at position 58, b) demeaned data, c) result of the ANOVA, d) result of the HANOVA.

SPRT in detecting plastic mines. The mines are M19, VS-1.6, T72, and C4A1. Fig. 13 shows the ROC curves of the above four methods. It is seen that both the ANOVA-SPRT and HANOVA-SPRT perform better than the ANOVA and HANOVA, respectively.

As another example, we test our method on a different array radar system at another test site. The setup of the GPR array is shown in Fig. 14. There is one transmitter in this system. In front of the transmitter, four receivers are positioned in a 2×2 pattern. Above the transmitter and the receivers there is a hyperbolic reflection plate, it is set so that the transmitter is at the focal point of the reflection plate. The array moves on a linear track

to collect data. At each step, the transmitter sends a spherical wave to the reflection plate and after reflection, the sphere wave becomes plane wave. The four receivers then collect reflection of this plane wave from the ground. The system has the advantage of generating plane wave and points it forward to reduce ground reflection. Fig. 15 displays collected data from the two front receivers at the Dedham test site of Northeastern University and the corresponding signal after background removal. In an area of 58 m², there are 12 buried landmines of different types, such as M19, PMN, VS-2.2, and so on. Using our method we are able to detect all 12 mines with a few false-alarms, Fig. 16. The results are similar to those obtained by a single GPR. For a detection rate above 90%, the HANOVA has a significantly smaller number of false-alarm.

IV. SUMMARY

In this paper, we have proposed a sequential, high-dimensional ANOVA to process GPR returns. The method is tested on real data and has a relaxed requirement on the physical model used in the processing routine. The method is on-line implementable and has a linear computational load. The method works in two stages: first it looks for statistically significant difference from array observations, second, it applies a sequential detection as new data are obtained. HANOVA is powerful in the sense of maximizing probability of detecting statistically significant difference among sub-dimensions of a full vector of observations. Sequential detection recursively processes the result of the HANOVA and enables real-time processing as new data are collected. We have demonstrated the performance of this technique on samples of field data.

Future research will focus on classification and localization. Classification consists of two steps. First, a feature extraction scheme is applied on reflected signals to generate intermediate result, second, the output from feature extraction is fed into a Bayesian classifier

to make the classification. Localization is based on optimized frequency-wavenumber (F-K) migration. F-K migration is an inversion method that back-propagates wavefield from the ground surface to subsurface and construct an image of subsurface reflectivity. Regular F-K migration is well modeled for seismic signal processing. Though GPR signal is different from seismic signal, F-K migration can still work very well in processing GPR signals. Nonetheless, improvement in terms of resolution and accuracy can be achieved by considering optimization in F-K migration. In our future work, an optimization method will be used to improve the F-K migration.

APPENDIX

I. WINDOW SELECTION IN HANOVA

Ideally, we want to find a window that is sensitive to the presence of a signal and provides little response in the test statistic when there is noise only. But these two requirements are often in conflict with each other. From (8), we can change the order l of the MA process to control the window we use. The smaller is l , the more sensitive the window is to the presence of signal and strong noise. On the other hand, the larger is l , the more robust will the statistic be to noise, which translates into a smaller probability of false-alarm. But a large l reduces sensitivity of the HANOVA to signal. Fig. 17 shows the effect of l on window selection and the corresponding HANOVA results. Three different l are used, i.e., $l = 1, 4, 9$. In the data, there are three mine objects, two metal mines at the position 110 and 170. A weak mine object is at position 25. For comparison, we normalize the HANOVA outputs in each case by its maximum value, which corresponds to the strong metal mine buried at position 110. Fig. 17(a) and (b) show the window chosen by a MA of order 1 and the resulting HANOVA output. The two strong objects can be detected at a threshold of 0.7, the weak object can only be found at a threshold of 0.2. Fig. 17(c) shows the window

chosen by a MA of order 4. The window oscillates much less than the window in Fig. 17(a). From the HANOVA result, Fig. 17(d), we can find all the three objects at a threshold of 0.3. Increasing the order of MA process can make the results worse, Fig. 17(e) and (f). A large window reduces the sensitivity of the HANOVA to signal and actually makes detection more difficult. Now the weak object at position 25 can not be detected at a threshold greater than 0.3. As a guideline, we find that MA processes of order between 3 and 10 yield good windows both in sensitivity to signal and robustness to noise. This selection is affected by the step-size of the array. An array moving at small step-size will allow an MA process of large l in selecting windows, and vice versa.

REFERENCES

- [1] L. Peters Jr., J. J. Daniels, and J. D. Young, “Ground penetrating radar as a subsurface environmental sensing tool”, *Proceedings of the IEEE*, vol. 82, no. 12, pp. 1802–1822, Dec. 1994.
- [2] C.M. Rappaport, S.G. Azevedo, T. Rosenbury, and J. Gough, “Handheld forward-looking focused array mine detection with plane wave excitation”, in *Detection Remediation of Mines and Minelike Targets V*, 2000, vol. SPIE 4038.
- [3] J. Kositsky and P. Milanfar, “A forward-looking high-resolution GPR system”, in *Detection Remediation of Mines and Minelike Targets IV*, 1999, vol. SPIE 3710, pp. 1052–1062.
- [4] J. M. Bourgeois and G. S. Smith, “A fully three-dimensional simulation of a ground-penetrating radar: FDTD theory compared with experiment”, *IEEE Trans. Geosci. Remote Sensing*, vol. 34, no. 1, pp. 36–44, Jan. 1996.
- [5] J. R. Wait, *Geo-electromagnetism*, Academic Press, Inc., New York, 1982.
- [6] G. R. Olhoeft, “Application of ground penetrating radar”, in *GPR’96*, Sendai, Japan, April 29, 2001—3 : 44 pm

- Sept. 1996, pp. 1–3.
- [7] K. Demarest, Z. Huang, and R. Plumb, “An FDTD near- to far-zone transformation for scatterers buried in stratified ground”, *IEEE Trans. Antennas and Propagation*, vol. 44, no. 8, pp. 1150–1157, Aug. 1996.
- [8] P. G. Gader, H. Hichem, B. N. Nelson, G. Vaillette, and J. M. Keller, “New results in fuzzy set based detection of landmines with GPR”, in *Detection and Remediation of Mines and Minelike Targets IV*, 1999, vol. SPIE 3710, pp. 1075–1084.
- [9] M. R. Bradley, T. R. Witten, R. McCummins, M. Crowe, S. Stewart, and M. Duncan, “Mine detection with a multichannel stepped-frequency ground-penetrating radar”, in *Detection and Remediation of Mines and Minelike Targets IV*, 1999, vol. SPIE 3710, pp. 953–960.
- [10] T. Dogaru and L. Carin, “Time-domain sensing of targets buried under a rough air-ground interface”, *IEEE Trans. Antennas and Propagation*, vol. 46, no. 3, pp. 360–372, Mar. 1998.
- [11] “UXO Signature Data”, <http://www.denix.osd.mil/denix/Public/News/UXOCOE/Sigdata/sigdata.html>, 1999.
- [12] J. Fan, “Test of significance based on wavelet thresholding and Neyman’s truncation”, *Journal of American Statistical Association*, vol. 91, pp. 674–688, 1996.
- [13] J. Fan and S. Lin, “Test of significance when data are curves”, *Journal of American Statistical Association*, vol. 93, pp. 1007–1021, 1998.
- [14] C. W. Helstrom, *Elements of Signal Detection and Estimation*, Prentice-Hall, New York, 1995.
- [15] P. Shi and E. L. Miller, “Baseband Wiener filter processing for mine detection from scanned laser induced acoustic data”, in *Detection and Remediation of Mines and Mine-*

- like Targets IV*, 1999, vol. SPIE 3710, pp. 1373–1384.
- [16] X. Xu, E. L. Miller, and C. M. Rappaport, “Combined high-dimensional analysis of variance (HANOVA) and sequential probability ratio test (SPRT) to detect buried mines”, in *Detection and Remediation of Mines and Minelike Targets V*, Apr. 2000, vol. SPIE 4038.
- [17] X. Xu, E. L. Miller, and C. M. Rappaport, “Statistically-based sequential detection of buried mines from array ground penetrating radar data”, in *Detection and Remediation of Mines and Minelike Targets IV*, 1999, vol. SPIE 3710, pp. 1063–1075.
- [18] M. El-Shenawee and C. M. Rappaport, “Quantifying the effects of different rough surface statistics for mine detection using the FDTD technique”, in *Detection Remediation of Mines and Minelike Targets V*, 2000, vol. SPIE 4038, pp. 966–975.
- [19] B. K. Ghosh, *Sequential Tests of Statistical Hypotheses*, Addison-Wesley, Reading, MA, 1970.
- [20] S. C. Port, *Theoretical Probability for Applications*, John Wiley & Sons, New York, 1994.
- [21] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlag, New York, 1985.
- [22] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall, New York, 1993.
- [23] C. Han, P. K. Willett, and D. A. Abraham, “Some methods to evaluate the performance of Page’s test as used to detect transient signals”, *IEEE Trans. Signal Processing*, vol. 47, no. 8, pp. 2112–2127, Aug. 1999.
- [24] E. Bahar, “Full-wave solutions for the depolarization of the scattered radiation fields by rough surfaces of arbitrary slope”, *IEEE Trans. Antennas and Propagation*, vol. AP-29,

pp. 443–454, 1981.

- [25] D. A. Kapp and G. S. Brown, “A new numerical method for rough-surface scattering calculations”, *IEEE Trans. Antennas and Propagation*, vol. 44, no. 5, pp. 711–721, May 1996.

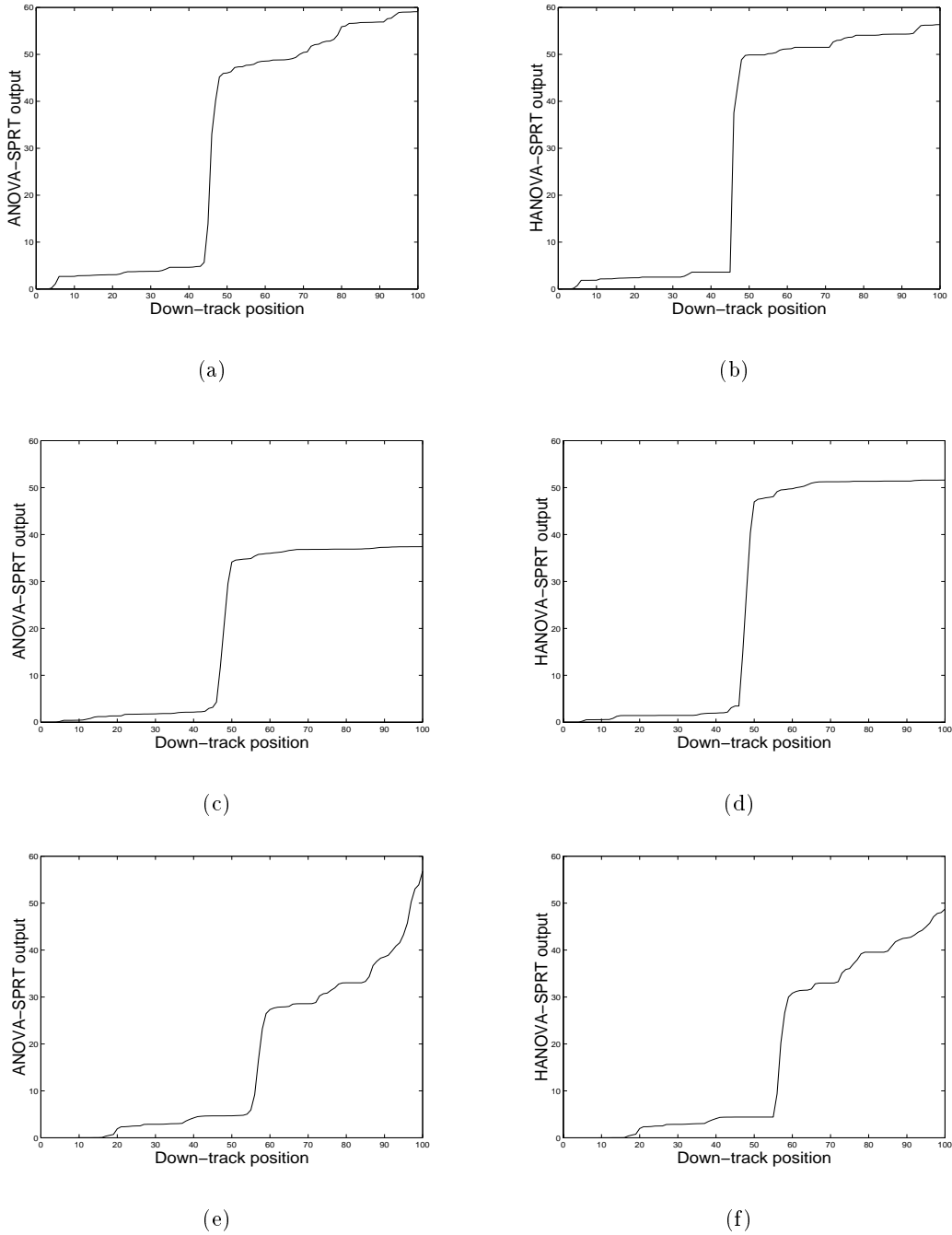


Fig. 11. Results of the SPRT, a buried steel object, a) output of ANOVA-SPRT, b) output of HANOVA-SPRT; a buried M19, c) output of ANOVA-SPRT, d) output of HANOVA-SPRT; a buried TM62, e) output of ANOVA-SPRT, f) output of HANOVA-SPRT.

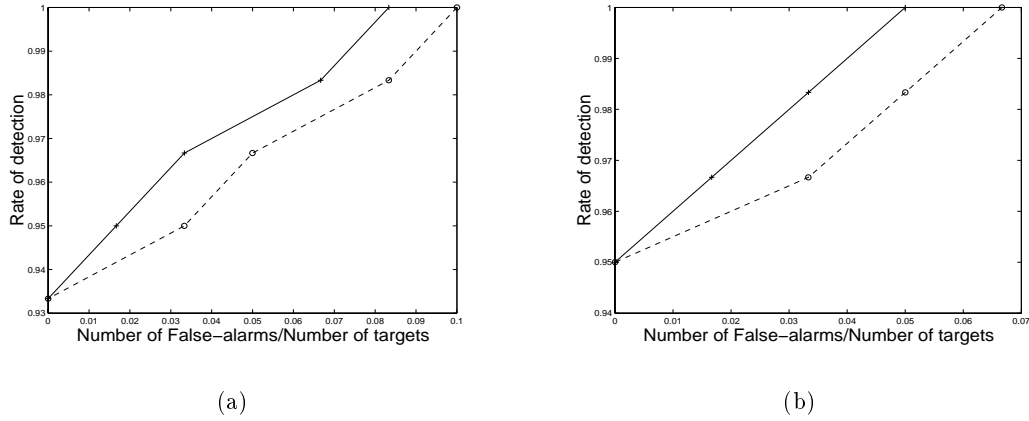


Fig. 12. Rate of detection and rate of false-alarms in detecting metallic objects, solid line is the result of HANOVA, dashed line is the result of ANOVA, a) ANOVA vs. HANOVA, b) ANOVA-SPRT vs. HANOVA-SPRT.

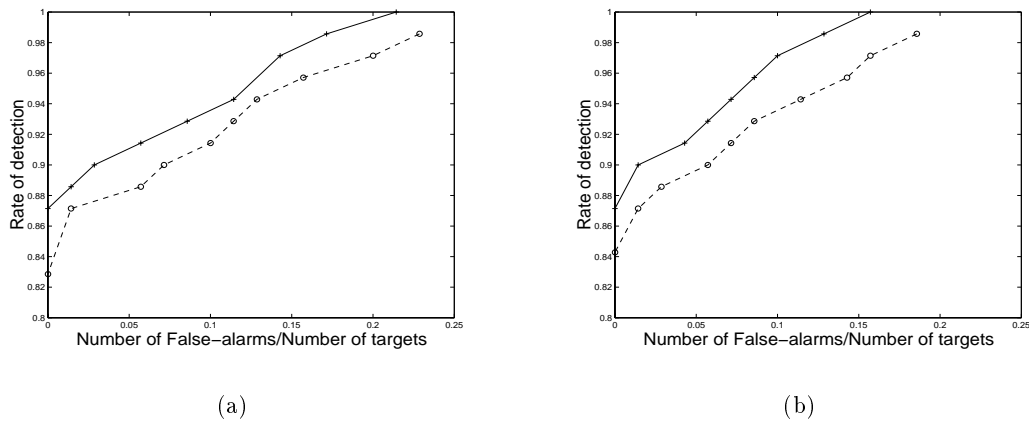


Fig. 13. Rate of detection and rate of false-alarms in detecting plastic mines, solid line is the result of HANOVA, dashed line is the result of ANOVA, a) ANOVA vs. HANOVA, b) ANOVA-SPRT vs. HANOVA-SPRT.

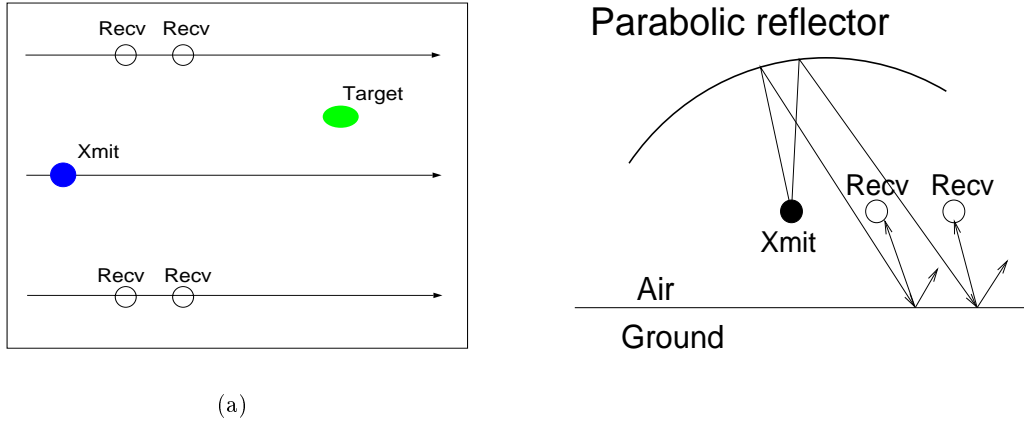


Fig. 14. GPR array used at Dedham test site of Northeastern University, a) plane view, b) side view.

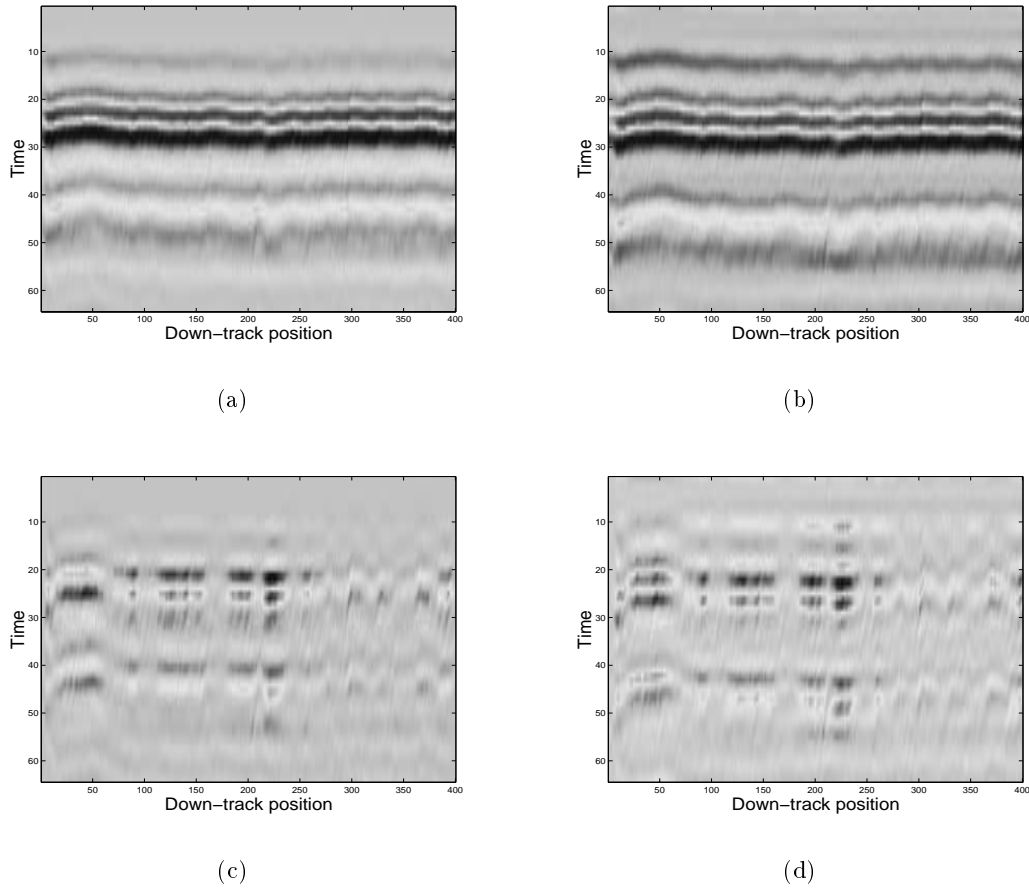


Fig. 15. GPR data from the Dedham test site, (a) from the left front receiver, (b) from the right front receiver, (c) signal (a) after background removal, (d) signal of (b) after background removal. Unit in time axes is 120 ps.

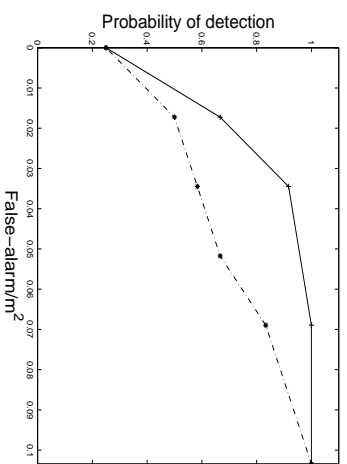


Fig. 16. ROCs of the Dedham test, solid line, HANOVA-SPRT, dash line, ANOVA-SPRT.

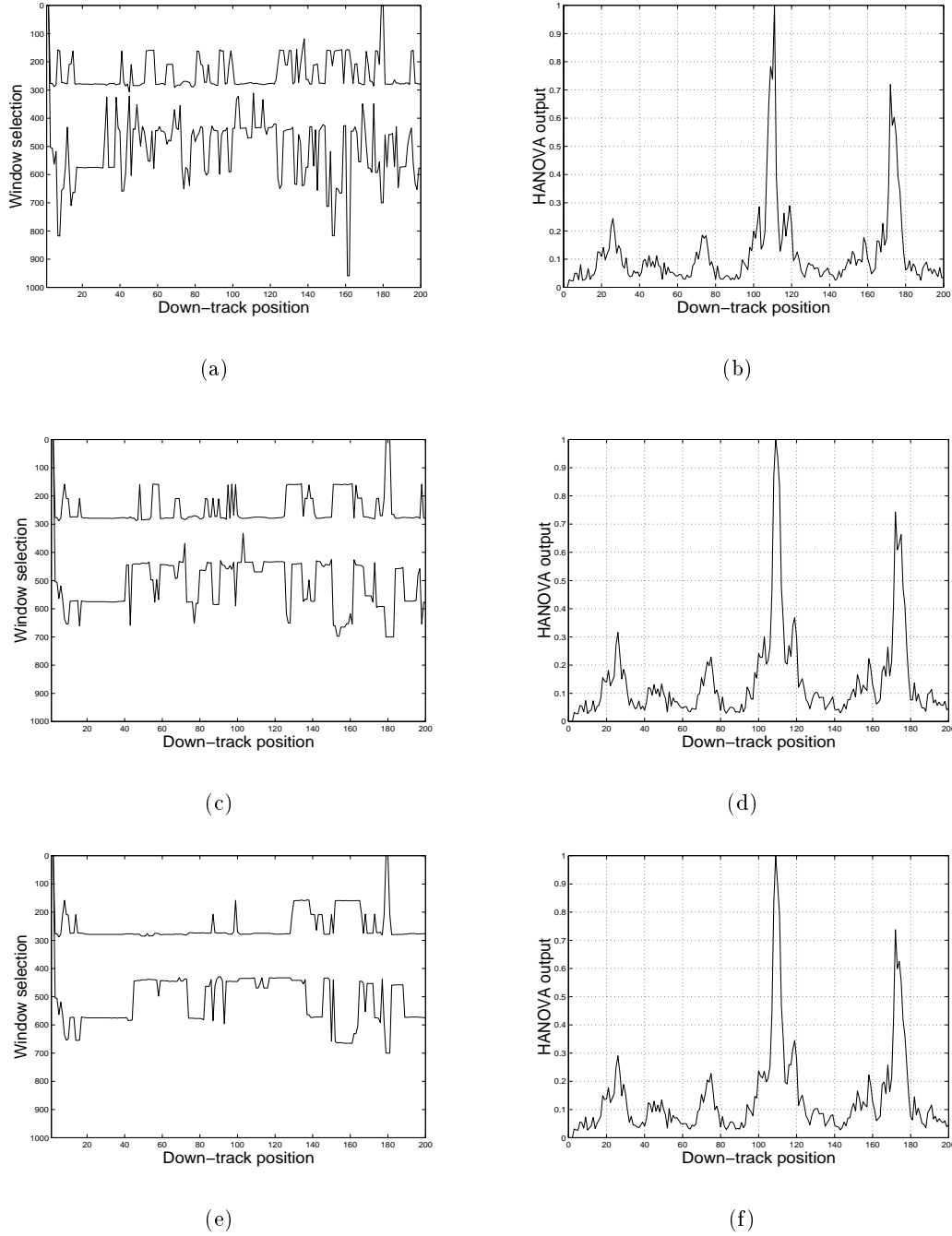


Fig. 17. Choices of window and effect on HANOVA, lower line is $k_1(m, n)$, upper line is $k_2(m, n)$, a) window selected by an order 1 MA process, b) HANOVA result from the window to the left, c) window selected by an order 4 MA process, d) HANOVA result from the window to the left, e) window selected by an order 10 MA process, f) HANOVA test result from the window to the left.

Baseband Wiener Filter Processing for Mine Detection from Scanned Laser Induced Acoustic Data

Pengyu Shi ^a and Eric L. Miller^a

^a Center for Electromagnetics Research,
Department of Electrical and Computer Engineering
Northeastern University, Boston, MA.

ABSTRACT

We consider the problems of detecting, localizing, and characterizing the shapes of buried mines from acoustic transducer data. A multipath model is used to describe the contributions in the data from fields scattered both by the ground as well as the object. By identifying the parameters in the model, we can successfully solve the target identification problem. Unfortunately, the narrow band and bandpass characteristics of the transducer in our system prevent a straightforward application of a Wiener filter as a means of extracting these parameters. Because of the bandpass nature of the transducer it is necessary to develop a base-banding procedure as a preprocessing stage for the Wiener filter. Due to the very narrow bandwidth, a detect and subtract method is constructed to extract the weak signal arising from the buried object which otherwise is drowned out in the sidelobes generated by the processing of the ground bounce. We demonstrate the utility of our approach on real experimental data collected at Northeastern University.

Keywords: Mine detection, laser induced acoustic scanner, Wiener filter

1. INTRODUCTION

The problem of detecting and localizing buried mines from observation of reflected radiation has received considerable attention in recent years. In this paper, we consider a statistical signal processing approach to solving such problems given data collected using a scanning laser-based acoustic sensor developed in [1]. The experimental system is shown in Fig. 1. The transmitter sends a laser pulse into the ground where an unknown target is buried. An acoustic wave is generated by the sand as it is heated by laser. Part of this acoustic wave is reflected by the ground and seen by the acoustic transducer. The other part is transmitted into the earth and reflected by a landmine. Thus the received signal will be the mixture of reflected waves from both the ground and the target. This procedure is repeated on a regular basis as the apparatus scans a 2D grid. The objective of the processing is the localization and characterization of the buried object given this collection of time traces.

Because the propagation and scattering of acoustic waves through the earth is a highly complex physical process even when the constitutive parameters of the medium are known, it would be quite difficult to employ sophisticated inverse scattering type methods to image the subsurface. Thus, in this paper we consider a simplified model for describing the data which lends itself to accurate yet tractable information extraction algorithms. We assume that the problem can be described using a so-called *multipath model* commonly employed in the field of communications² wherein the received signal at location i of the scan, $r_i(n)$, can be expressed as

$$r_i(n) = s(n) * h_i(n) + w(n) \quad (1)$$

Other author information: PS: peshi@ece.neu.edu, ELM: elmiller@ece.neu.ed. This work was supported by the Army Research Office Demining MURI under Grant DAAG55-97-1-0013

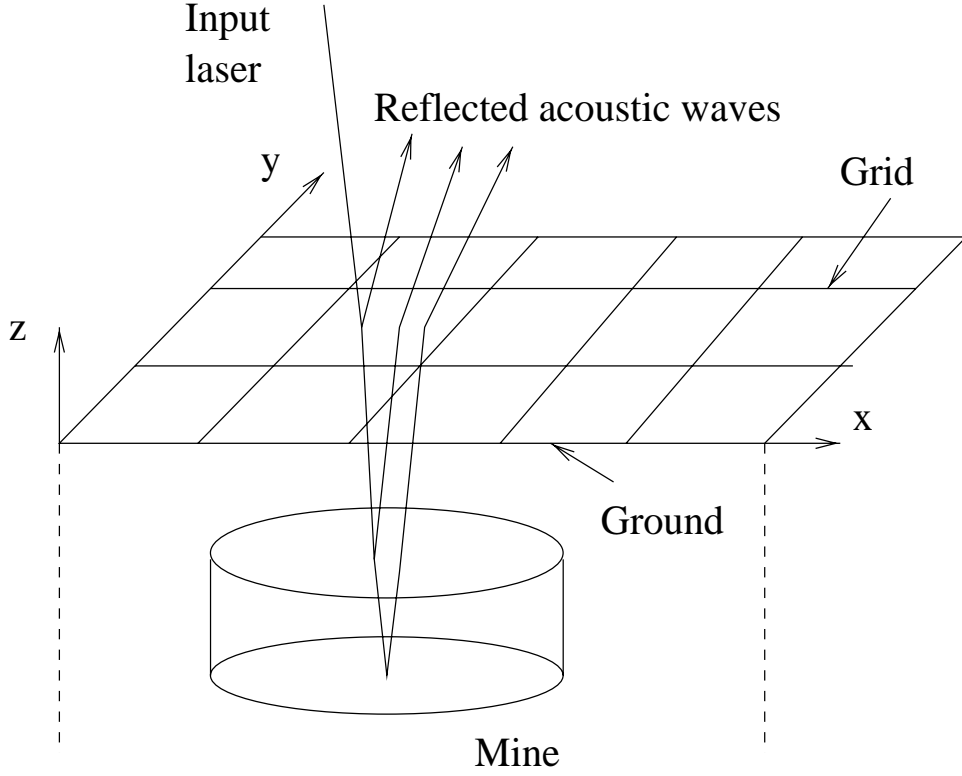


Figure 1. Experimental system

where $s(n)$ is taken to be a nominal template signal, $w(n)$ is additive noise and $*$ denotes convolution. The multipath channel impulse response, $h_i(n)$, characterizes the structure of the reflection from the ground and the mine when the sensing system is at location i and is written as

$$h_i(n) = \sum_{j=1}^{N_i} \alpha_{i,j} \delta(n - \tau_{i,j}) \quad (2)$$

with N_i the number of paths seen at location i . In this work, N_i is taken to be one or two depending on whether only the ground bounce is seen in the receiver ($N_i = 1$) or if one observes both the ground bounce as well as a response from the buried object ($N_i = 2$). The quantities $\tau_{i,j}$ are the time delays associated with each path and the $\alpha_{i,j}$ are amplitude reduction factors caused by the propagation and scattering processes.

The goal of the processing then is the determination of the multipath parameters, $\{N_i, \alpha_{i,j}, \tau_{i,j}\}$ for each position of the sensor. As we discuss in § 3, examination of these quantities can be of great use in localizing and characterizing the shape of a buried object. From a processing perspective then, the challenging feature of this problem is the detection of the multipath parameters. This is a common statistical signal processing problem and for which various solution methods have been proposed including generalized cross correlation,³ Wiener-type filtering,⁴ and nonlinear estimation techniques.⁵ In this paper, we focus our attention on Wiener filtering processing due to its simplicity and strong performance for the problem at hand.

2. DETECTION METHOD

The Wiener filter solution for a discrete time channel identification problem given by (2) is⁴:

$$\hat{H}(e^{j\omega}) = \frac{R(e^{j\omega})S^*(e^{j\omega})}{|S(e^{j\omega})|^2 + \epsilon} \quad (3)$$

where $\hat{H}(e^{j\omega})$, $R(e^{j\omega})$ and $S(e^{j\omega})$ are the discrete-time Fourier transforms (DTFT) of $\hat{h}(n)$, $r(n)$ and $s(n)$ respectively, ϵ is the variance of the channel distribution, and a hat over a variable indicates estimated quantity. The desired multipath parameters can then be determined by taking the inverse DTFT of (3) to obtain $\hat{h}(n)$ and finding the peaks. The locations of the peaks provide estimates of the delay parameters and the amplitudes of the peaks yield the α 's. There are however three difficulties that need to be addressed before applying the Wiener filter to our problem:

- The received signal in our system is bandpass. As is discussed at length in § 2.1 this introduces a number of phase shifts which impact the performance of a straightforward Wiener filter implementation.
- The ability to clearly distinguish the peaks in \hat{h} requires that the data be broadband. Unfortunately, the transducer in our system has a very narrow bandwidth so $\hat{h}(n)$ will display large side-lobes and a broad main-lobe. The delay and amplitude estimation problem becomes much more difficult in this case especially since the second path (indicating the presence of an object) will be highly attenuated with a delay quite close to that of the ground bounce.
- The template $s(n)$ for our problem is unknown while the above processing assumes it is given. The precise structure of the template signal depends on the nature of the transmitter and receiver as well as the detailed physics of the propagation problems. For none of these quantities do we currently have accurate models.

To solve this problem we examine two approaches. First, we will determine $s(n)$ empirically by taking it to be the signal received from a known no-mine location. That is, the observed ground path will represent the template. Then all other traces will be processed based on this template. Second, we explore methods for adaptively updating the template as the sensor scans to combat the spatial variability of the template and reduce the effects of additive noise.

These three problems will be addressed in the following three subsections respectively.

2.1. Baseband processing with Wiener filter

Because the transducer acts as a bandpass filter, the Wiener filter output is bandpass with a non-zero carrier. Since there is normally a phase shift between the baseband Wiener filter output and this carrier, the peak location of the bandpass Wiener filter output will generally differ from that of the baseband case. More importantly, there is also a phase shift between the reflected waves from the ground and the buried object. In this case, the bandpass signal can not be used directly as a template for the Wiener filter because the reflected wave of the second path is no longer a scaled and delayed version of the first. This fact makes direct use of the Wiener filter in (3) highly inaccurate. However, this problem can be solved by the following baseband processing.

Dropping the explicit dependence on i , the position of the sensor, the bandpass signal can be written as $s(n) = s_c(n) \cos(\omega_0 n + \phi) - s_s(n) \sin(\omega_0 n + \phi)$, where s_c and s_s are the in-phase and quadrature parts of s respectively. Denote the baseband expression of $s(n)$ as $s_b(n) = s_c(n) + js_s(n)$. Assuming there are

only two paths in the channel, we have $r(n) = \alpha_1 s(n - \tau_1) + \alpha_2 s(n - \tau_2) + w(n)$, where α_1 and α_2 are real and $w(n)$ is additive Gaussian noise. After substitution, we have

$$r(n) = \sum_{j=1}^{N_i} \alpha_j s_c(n - \tau_j) \cos(\omega_0(n - \tau_j) + \phi_j) + \alpha_j s_s(n - \tau_j) \sin(\omega_0(n - \tau_j) + \phi_j) + w(n) \quad (4)$$

The signal $r(n)$ then is down converted to baseband by multiplying by $\cos(\omega_0 n)$ and $\sin(\omega_0 n)$ respectively. After some algebra, it can be shown that at baseband,

$$r_b(n) = \sum_{j=1}^{N_i} \alpha_j \exp(j\phi'_j) s_b(n - \tau_j) + w_b(n) \quad (5)$$

where $w_b(n)$ is the baseband expression for $w(n)$ and is complex. Thus, it can be seen that by using $s_b(n)$ as a template, path information such as time delay τ , amplitude α and phase shift ϕ' can be detected by a Wiener filter which is employed at baseband. Following this algorithm, our baseband Wiener filtering procession can be implemented as shown in Fig. 2.

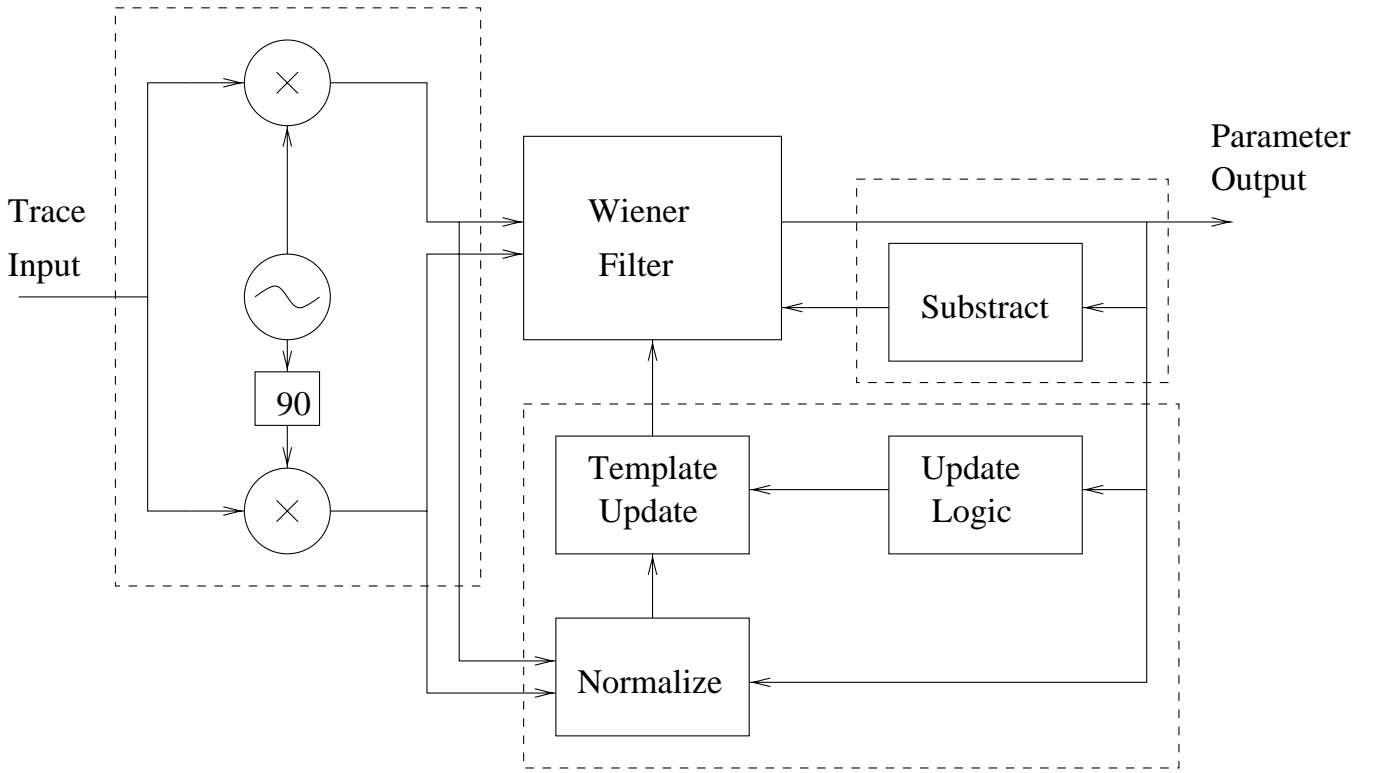


Figure 2. Baseband Wiener filter processing

2.2. Template Determination

The next problem we address is the determination of the template signal $s_b(t)$. We begin by using a fixed template obtained as the trace at a known no-mine location. By doing this, it is assumed that the noise level is not very high which is valid for our system. While the assumption of a fixed template may be

accurate over some limited spatial region, the changing characteristics of ground (temperature, moisture, material composition, etc.), imply that one should consider methods for adaptively updating the template. Thus, the problem at hand is how to decide the template for a new trace before we do the processing. To address this difficulty, first we assume that the system is slowly space-varying system, which implies the template at a location will be highly correlated with that of its neighbors. Thus we may predict the new template based on the previous templates.

One issue of concern here is that we update the template using only data from positions in which there is no buried object. While one could extract templates from neighboring scans for which a mine is present, the utility of such a procedure is heavily dependent on the accuracy of the estimated multipath parameters. While the experiments in § 3 demonstrate that we can determine roughly the location and shape of the buried object, at this point in time, the parameter estimates are not sufficiently accurate for template extraction. Thus, we prefer to take an approach where, at each location of the sensor, we first determine whether or not there is a buried object present. This determination is made by examining the estimate of the amplitude coefficient associated with the second path. If this estimate is above a given threshold, we say that a mine is present, otherwise we assume there is no object at this position. A couple of remarks are in order concerning this process:

1. At the current time, the threshold is determined assuming we have a rough idea as to the size of the second-path amplitude coefficient for the class of buried objects of interest. Clearly, a key area of future research is the adaptive determination of this quantity as the sensor scans the region.
2. Given knowledge of the second-path amplitude, we currently set the mine/no-mine threshold relatively low to ensure that few traces associated with the presence of a mine are used in the updating procedure. This low threshold does cause some cases where we should update the template but do not because we mistakenly say that a mine is present. We have found empirically that such “mistakes” have little impact on the overall results of our processing as the template itself generally changes quite slowly so that a small number of missed updates cause little degradation in performance.

Finally our updating algorithm is

$$s_{new}(n) = (1 - \beta)s_{old}(n) + \beta r'(n) \quad (6)$$

where $s_{new}(n)$ is the new template, $s_{old}(n)$ is the old template, β is updating factor and $r'(n)$ is a normalized version of current trace $r(n)$ with $r'(n) = r(n - \tau_1)/\alpha_1$. The received current trace cannot be used directly because the ground is normally not flat. The ground-reflected trace will have a different arrival time thus causing error if added directly. Further, the power of the template needs to be fixed so that the detected amplitude information of the second path can be employed to make a decision at both the updating stage and the post-detection stage. Moreover, by averaging the signals using (6) this algorithm is seen to be capable of alleviating the effects of additive noise.

So far what has been achieved is a Wiener filter output which contains all the information of the channel while the particular characteristic of template are largely eliminated. The next question we want to answer is how to extract channel information from this output. This issue is highly nontrivial when considering the extremely narrow bandwidth of the received signal (see Fig. 3). As discussed in the next section, we consider a detect and subtract method for solving this problem.

2.3. Detect and Subtract

The acoustic transducer used in our system has a very narrow bandwidth (Fig. 3). Thus, in the time domain, all the spikes (i.e. multipath components) in the channel response are “smeared out” which makes

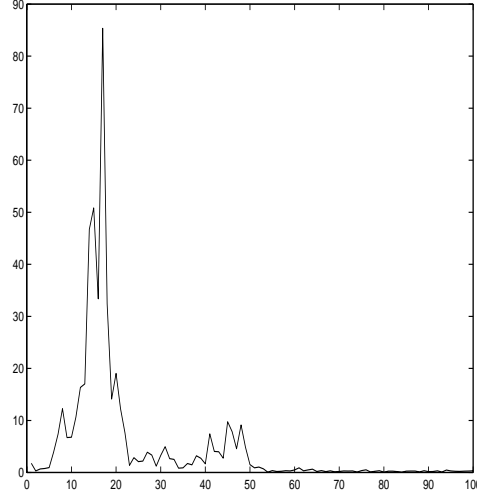


Figure 3. Frequency spectrum of received signal

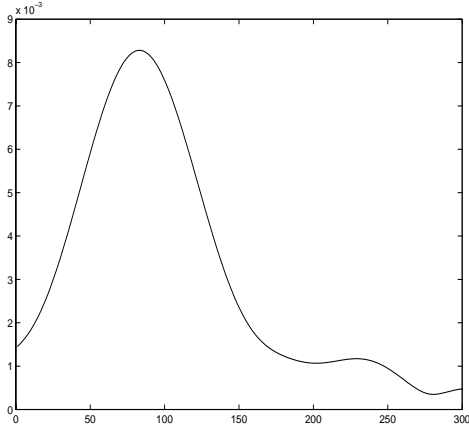
the distinction of closely separated paths very difficult. More problematic, it also introduces large side lobes at the output of the Wiener filter which makes weak paths vanished in the side lobes of a strong path. Unfortunately, the amplitude of the desired second path (coming from the mine) is much smaller than that of main path (coming from the ground) in our system. Thus the second path is very difficult to detect from the original Wiener filter's large-rippled output. Figs. 4(a) and (c) are typical traces obtained over a no-mine location and a mine location respectively. Clearly, they are very similar and the second path can not be detected directly.

Our method to overcome this problem is as follows. First we extract the information (amplitude, time delay and phase shift) from the ground-reflected path. Due to the strength of this signal, this process is generally quite accurate. Next, rather than continuing by looking for the second path from this original Wiener filter output, we use our estimated parameters to mathematically subtract the ground bounce portion of the signal from the current time trace. By using this modified trace as input to a new Wiener filter, it is then much easier to determine the parameters of the signal component coming from the buried object. In Figs. 4(b) and (d) we show the Wiener filter output after subtraction of the first path for Figs. 4(a) and (c) respectively. It can be seen that the second path is clear in Fig. 4(d) while the second path in Fig. 4(b) will be eliminated due to its small amplitude.

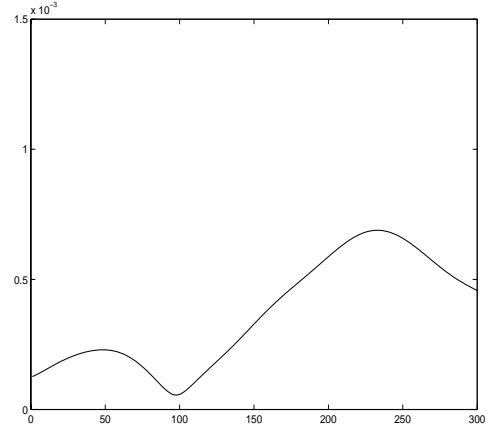
2.4. Mine Detection with Post Processing

By applying the above methods, we get estimates of the amplitude, time delay and phase shift of the ground-reflected wave (first or main path) and the target-reflected wave (second path). The issue now is how best to use the information of the second path to decide if there is a mine and determine the mine's contour and depth in the sand.

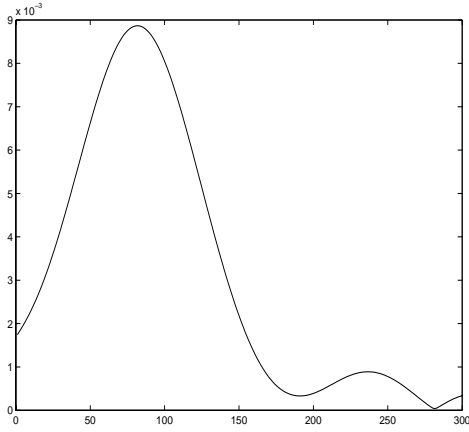
Note that no matter whether there is a mine, the second path parameters are always estimated. However it can be seen from the results that the mine's information can be identified by noticing the fact that at positions where a mine is present: (a) the amplitudes of the second path will be relatively large and (b) the time delay of the second path will be pretty close (flat target surface) or vary slowly (non-flat



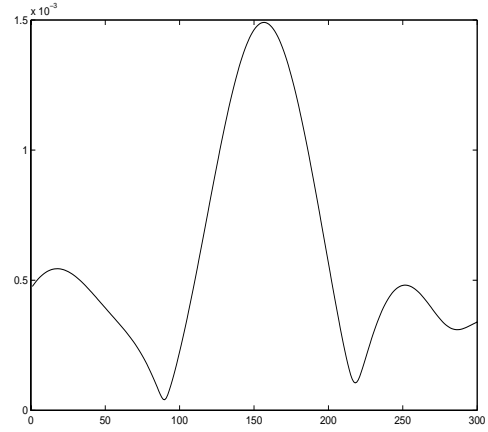
(a) No 2nd path, before DAS



(b) No 2nd path, after DAS



(c) With 2nd path, before DAS



(d) With 2nd path, after DAS

Figure 4. Output of Wiener filter

target surface). Based on these observation, simple post processing is employed by checking the amplitude of second path and comparing the time delay with its neighbors. First, for locations with the largest second path amplitudes, we say there is a mine. Then we reduce the amplitude threshold and examine the locations which are adjacent to the currently-detected mine-locations. If their delays are close enough (a delay threshold is used here), we say these adjacent locations are also mine-locations. This process is then repeated until a predetermined amplitude threshold is reached at which point we stop. As was the case with the template updating process, the value of this lower amplitude limit is determined under the assumption that we have a rough idea as to the second-path amplitude coefficient for the class of object under consideration.

3. RESULTS

The sand box is grided into 25x15 unit squares and a half hockey puck is buried horizontally. The results of the fixed template method are shown in Fig. 5, (a) is the ground delay (relative to the initial template which is chosen to be (1,1) location), from which we can see that the ground is not flat as assumed and has a small slope. Fig. 5(b) and (c) are the amplitude and time delay of detected second paths respectively. It can be seen that the amplitudes of the second path when a mine is present are relatively larger than those of no-mine locations and the time delays for mine locations are pretty close to each other (due to flat upper surface of the target). After post-detection processing, the mine's support is shown in Fig. 5(d).

The updating template method is also employed with several different updating factors, β in (6). It is found that when β is large (eg. 0.8 or 1.0), the system is sensitive to the detection error and the performance is worse. If β is 0 then we have the non-adaptive processing scheme. We have experimentally determined that an updating factor of 0.2 is found to be a good choice and its results are shown in Fig. 6. We can see that ground delay is similar to what we had before but the amplitude and time delay of the second path are much clearer. Finally, the detected mine's support is shown in Fig. 6(d). Relative to the fixed template case, we see that the holes have been filled and the outliers entirely suppressed.

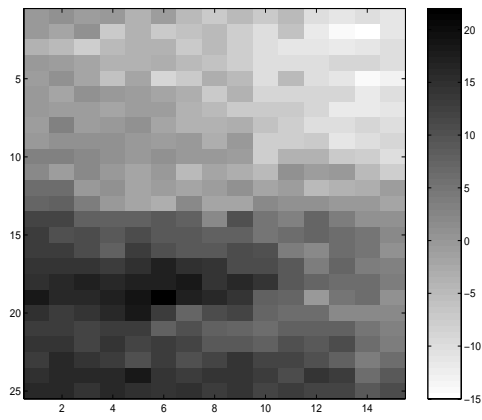
Fig. 7 shows the results of the updating template method (with $\beta = 0.2$) when two rectangular targets are buried in the sand. As can be seen, these two objects can also be detected clearly.

4. CONCLUSIONS AND FUTURE WORK

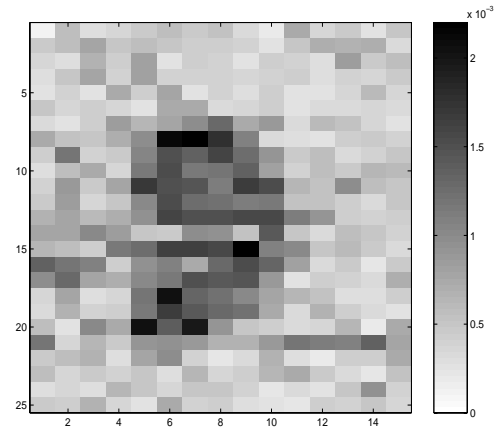
In this paper, we have presented an approach for the detection and localization of buried mines from acoustic transducer data via parameter estimation using a multipath propagation model. Since the received signal is bandpass, a baseband Wiener filter is introduced to solve the problem of phase shifts among the multipath components in the signal. Due to the extremely narrow bandwidth of the transducer, distinct multipath spikes are smeared out and large side lobes are generated in the output of the Wiener filter. A detection and subtraction method then is employed to deal with this bandwidth problem. Both fixed and adaptive template methods are considered in this paper. It can be seen from the result that the later works better by properly choosing the update factor.

The above processing is based on a linear multipath model assumption. Specifically, the characteristics of acoustic wave propagation are entirely omitted to simplify the analysis and computation. Thus the detection performance is expected to be better if the model can be improved to include more specific features of the acoustic wave properties. In particular, we conjecture that there is a third signal component caused by reflection of the input signal from the bottom of the mine. Our current processing methods however are insufficient to pull this very weak signal from the data. The solution may lie on the improvement of channel and noise model but also on the more powerful parameter estimation methods. All told, we consider the following topics to be of great interest for future work

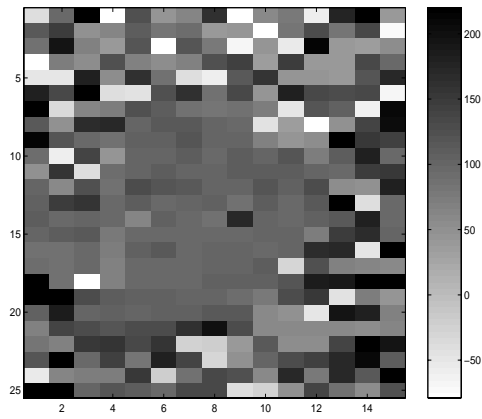
1. Propagation model improvement.
2. Noise characteristic modeling.
3. Adaptive and robust methods for determining the thresholds in the template updating and post detection processing stages.
4. Performance analysis for all detection and estimation stages of our approach.
5. Post detection algorithm modification to further sharpen the boundaries of the object.



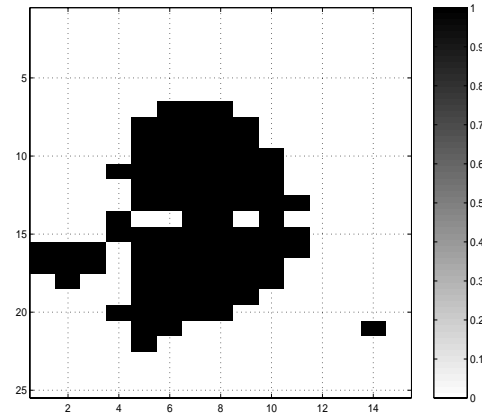
(a) Time delay of first path



(b) Amplitude of second path

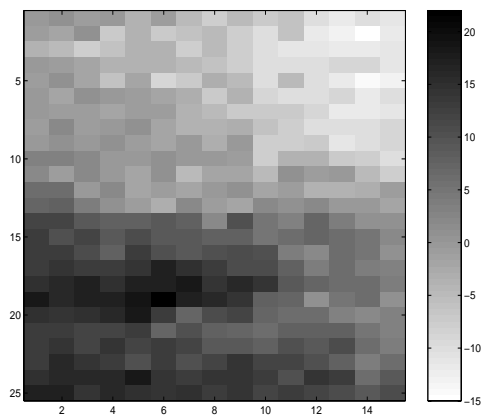


(c) Time delay of second path

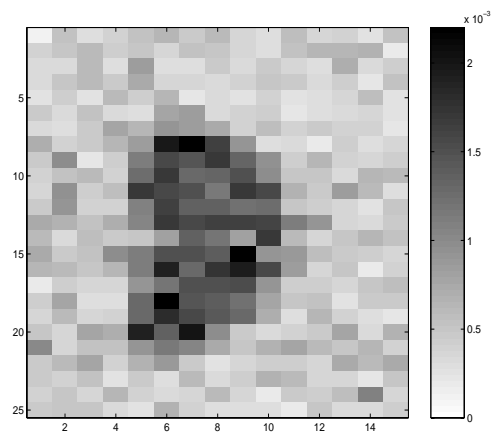


(d) Mine detection after post processing

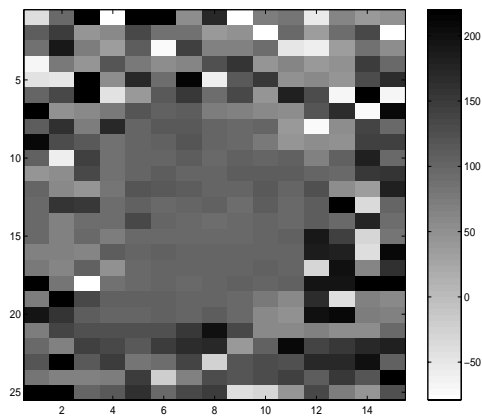
Figure 5. Results for fixed template



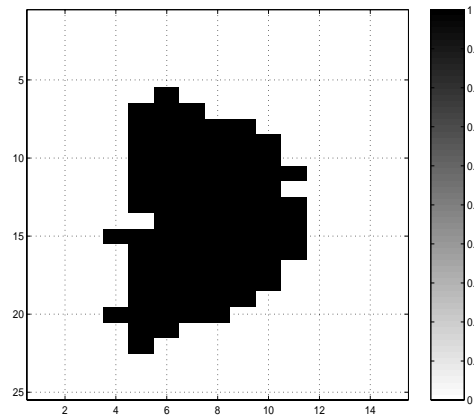
(a) Time delay of first path



(b) Amplitude of second path

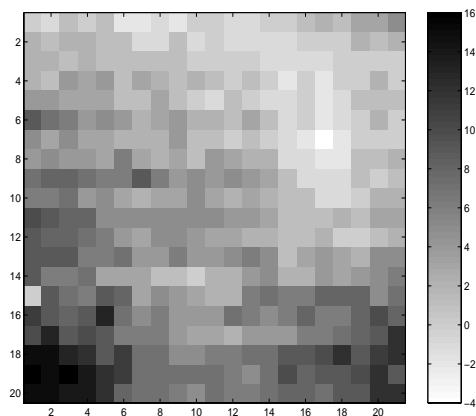


(c) Time delay of second path

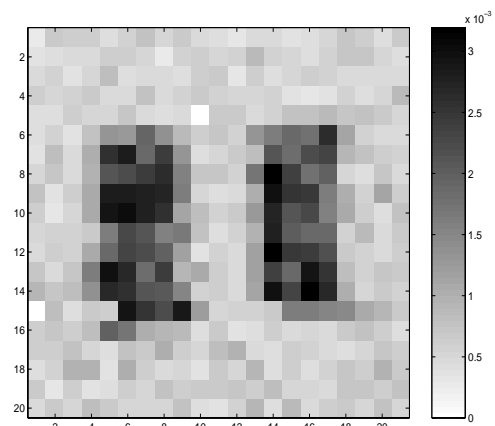


(d) Mine detection after post processing

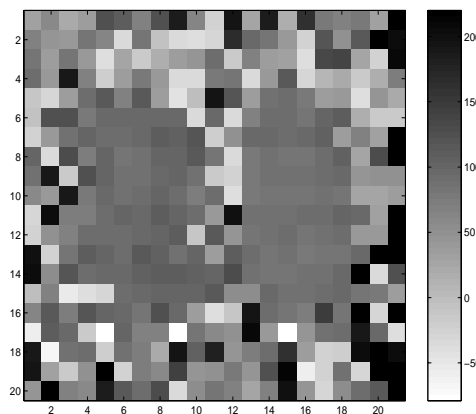
Figure 6. Results for updating template



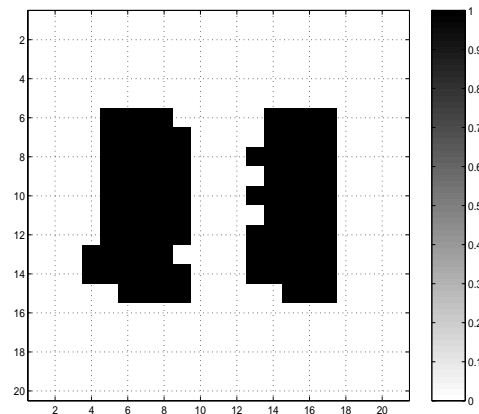
(a) Time delay of first path



(b) Amplitude of second path



(c) Time delay of second path



(d) Mine detection after post processing

Figure 7. Results for updating template (two objects)

5. ACKNOWLEDGMENTS

The authors would like to thank Mr. Wen Li, Dr. Charles DiMarzio and Prof. Stephen McKnight, all of Northeastern University, for helping to provide the data as well as the motivation for the work in this paper

REFERENCES

1. W. L. Stephen W. McKnight, Charles A. Dimarzio, D. O. Hogenboom, and G. O. Sauermann, "Laser-induced acoustic detection of buried objects," *Proc. SPIE, Detection and Remediation Technologies for Mines and Mine-like Targets III* **3392**, pp. 841–847, 1998.
2. J. Proakis, *Digital Communications, 3rd edition*, McGraw-Hill, 1995.
3. G. C. Carter, *Coherence and time delay estimation: an applied tutorial for research, development, test and evaluation engineers*, IEEE Press, 1993.
4. C. W. Therrien, *Discrete random signals and statistical signal processing*, Prentice Hall, 1992.
5. T. Koh and E. Powers, "Second-order volterra filtering and its application to nonlinear system identification," *IEEE Trans. Acoustic, Speech and Signal Processing* **35**, pp. 1492–1494, Oct. 1985.

Statistically-based Sequential Detection of Buried Mines from Array Ground Penetrating Radar Data

Xiaoyin Xu, Eric L. Miller, and Carey M. Rappaport
Center for Electromagnetics Research
Dept. of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115
Telephone: (617) 373-8386
Telefax: (617) 373-8627

ABSTRACT

We consider the problem of detecting and localizing buried landmines from a ground penetrating radar (GPR) array. A simplified, ray-optics-based physical model for time domain GPR returns is presented. Under this model in the absence of an object from the field of view of the array, there exist well defined symmetries in the structure of the radar returns. In particular, for a bistatic system composed of one length M transmit array and a second length M array of receivers, we identify M subsets of signals from the M^2 total transmitter/receiver pairs such that the mean value of the signals within each subset should be the same when no object is present. This relationship then forms the basis for a modified Hotelling's T^2 -test to detect the presence of objects when there is noise in the signal. Simulation results demonstrate the validity of these methods.

1. INTRODUCTION

Mines kill or maim hundreds of people every week, mostly innocent and defenseless civilians. Among the various demining methods developed in recent years, ground penetrating radar (GPR) holds substantial promise because of its sensitivity to non-metallic, plastic objects which traditional metal detectors are largely incapable of finding. Nevertheless, using GPR to detect and localize mines is both difficult and complex. One key problem is the rejection of interference caused by the signal arising from scatter off the air-earth interface. Both the magnitude of the ground bounce and its timing are such that they can easily swamp the relatively small signal arising from the interaction of the transmitted GPR waveform with the buried mine. To help overcome this difficulty, we consider the use of a GPR array to provide a richer and more diverse set of data thereby making accurate detection and localization possible in circumstances where a single GPR is unable to perform well.

In this paper we discuss a method to detect mines which exploits both the physics of the problem as well as the geometry of the array system. We assume that the GPR array is deployed as shown in Fig. 3 with one linear array of transmitters and a second array of receivers both traveling down a track. As described in Section 3, the geometric symmetry inherent in this configuration introduces a “statistical symmetry” in the received signals. In particular, this symmetry is preserved precisely when there is no mine. On the other hand, the presence of a mine will break this symmetry and therefore provide information as to the existence of an object. Based on this relationship, we develop a statistical test of homogeneity to ascertain the presence of an object in the field of view of the array.

The organization of this paper is as follows. Section 2 introduces the physical model of GPR signals and the configuration of the GPR array. In Section 3 we present the detection algorithm, using a modified

Other author information: XX: xxu@cdsp.neu.edu, ELM: elmiller@ece.neu.edu, CR: rappaport@neu.edu. This work was supported by the Army Research Office Demining MURI under Grant DAAG55-97-1-0013

form of Hotelling's T^2 -test. Examples are given in Section 4 to demonstrate validity of the algorithm. Section 5 summarizes present work and points out future research direction.

2. PHYSICAL MODEL AND PROBLEM FORMULATION

2.1. A Single GPR System

To detect and localize mines, a ground penetrating radar array is implemented. Fig. 1 shows a typical single GPR system with the signals it generates.¹ In this paper we assume a simplified model where the signal seen by the receiver is composed of at most two components. The first signal is the reflected signal from the ground and is always present in the data. The second component (if it exists) is the reflected contribution from an object in the field of view of the array.

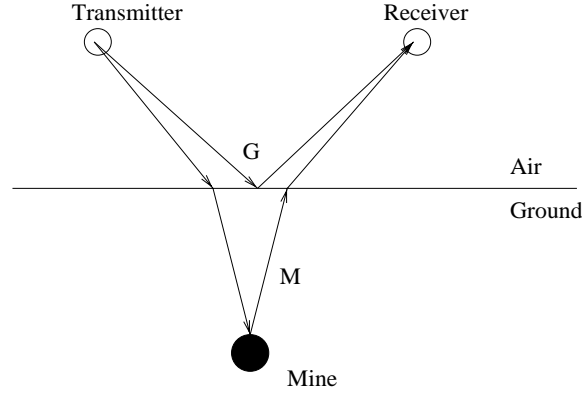


Figure 1. Schematic drawing of a single GPR, transmitter and receiver.

The received signal, $\phi(t)$, is taken to be the sum of delayed and attenuated versions of two “template” signals indicating the nominal behavior of the ground bounce signal and the nominal behavior of a signal arising from scattering from a mine. Mathematically we have

$$\phi(t) = a\psi_g(t - \tau_g) + b\psi_m(t - \tau_m) \quad (1)$$

where ψ_g and ψ_m are the nominal ground bounce and mine reflected signal, a and b are attenuation factors, τ_g is the delay of the ground reflection, and τ_m is the delay of the mine signal. Note that if no mine is present, ϕ is just equal to the first term of (1).

To find the delays and the attenuation factors we assume that the propagation of the signal from the transmitter to the receiver can be described using a ray-optics-type model shown in Fig. 1. That is, the ground bounce is composed of signal reflected from the interface at the specular point midway between the transmitter and receiver while the four-part path of the mine component of the signal can be determined via the judicious use of Snell's law.

To begin, the τ_g and τ_m are determined by the travel time of two-way paths and can be calculated as

$$\text{Delay} = \frac{\text{2-way path length}}{\text{velocity of the wave}}. \quad (2)$$

To find τ_g and τ_m , we need to locate reflecting point and refracting point shown in Fig. 2. Let media 1 be air and media 2 be soil, with electric permittivity ϵ_0 and ϵ_1 , respectively, Fig. 2.(a), the reflecting point on the boundary between two points (x_1, y_1) and (x_2, y_1) in media 1 is simply the mid-point $(x_4, 0)$, where $x_4 = \frac{x_1 + x_2}{2}$.

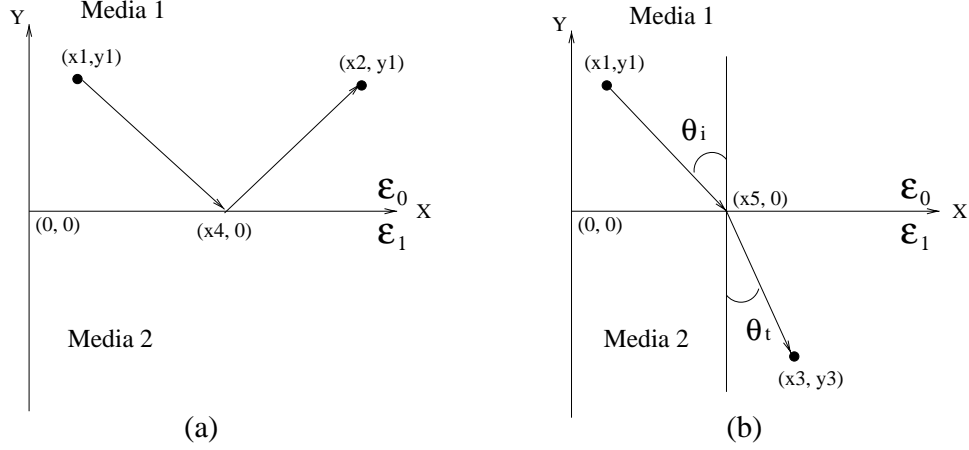


Figure 2. Geometries for determining (a) the reflecting point and (b) the refracting point, $\epsilon_1 > \epsilon_0$.

For the refracting point, according to Snell's law, for a source located at (x_1, y_1) in media 1 and target at (x_3, y_3) in media 2, the refracted ray from source to target must intersect the boundary at a point $(x_5, 0)$, Fig. 2.(b), such that²

$$\frac{Re\{\epsilon_1\}}{\epsilon_0} = \frac{\frac{(x_1 - x_5)^2}{(x_1 - x_5)^2 + y_1^2}}{\frac{(x_3 - x_5)^2}{(x_3 - x_5)^2 + y_3^2}}. \quad (3)$$

Solution of this quartic equation has four roots. By Fermat's principle, which states that of all possible paths joining two given points on a wave path, the wave path has actual least travel time, we can discard three physically impossible roots and retain the true refracting point. Once the reflecting point and the refracting point are established, the delay τ_g and τ_m can be found as,

$$\tau_g = 2 \frac{\sqrt{(x_1 - x_4)^2 + y_1^2}}{c} \quad (4)$$

$$\tau_m = 2 \frac{\sqrt{(x_1 - x_5)^2 + y_1^2}}{c} + 2 \frac{\sqrt{(x_3 - x_5)^2 + y_3^2}}{c/Re\sqrt{\epsilon_1}} \quad (5)$$

where c is the speed of light in air and $\frac{c}{Re\sqrt{\epsilon_1}}$ is the speed of the wave in soil.

In addition to the time delays, the received signal $\phi(t)$ has an amplitude reduction caused by propagation through the soil as well as geometric spreading as it traverses both the air and the earth. In soil, the wave attenuates exponentially with the distance it travels, $e^{-\alpha_s d}$. The quantity α_s is the attenuation constant of the soil which is related to the conductivity and permittivity of the medium³ while d is the distance the wave travels in the earth. We assume geometric spreading results in an inverse path length amplitude reduction. Referring to the setup of Fig. 2, then we have the overall amplitude reduction factors given by

$$a = \frac{1}{2\sqrt{(x_1 - x_4)^2 + y_1^2}} \quad \text{and} \quad b = \left(\frac{e^{-\alpha_s \sqrt{(x_3 - x_5)^2 + y_3^2}}}{\sqrt{(x_1 - x_5)^2 + y_1^2} + \sqrt{(x_3 - x_5)^2 + y_3^2}} \right)^2.$$

2.2. GPR Array

In this work, the GPR array is assumed to consist of M pairs of transmitters and receivers. Data are collected by the GPR array as it travels step by step down track, Fig. 3. At each stop of the array, M^2 signals (time-traces) are collected; one for each transmitter/receiver pair.

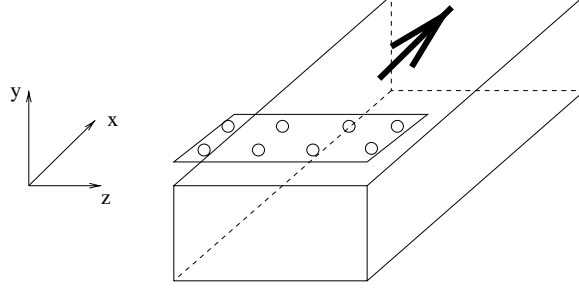


Figure 3. Array of GPR, moves in x-direction(down-track).

Using the model developed in the previous section and assuming that the GPR array is at its k -th stop, the signal seen at receiver j due to input from transmitter i is written as

$$\phi_{ij}^k(t) = a_{ij}\psi_g(t - \tau_{g,ij}) + b_{ij}^k\psi_m(t - \tau_{m,ij}^k), \quad i, j = 1, \dots, M, \quad k = 1, \dots, K. \quad (6)$$

Note that a and τ_g do not depend on k , the down-track GPR array position. This is easy to understand because the ground-reflected signal only depends on the relative position of transmitter and receiver*. To simplify matters, in the future we use the following shorthand

$$g_{ij}(t) = a_{ij}\psi_g(t - \tau_{g,ij}) \quad s_{ij}^k(t) = b_{ij}^k\psi_m(t - \tau_{m,ij}^k).$$

Each $\phi_{ij}(t)$ is densely sampled P times over a time interval. The interval is chosen to be long enough to embrace both ground bounce and mine signal. When no ambiguity will arise, we refer to the vector of samples, $\underline{\phi}_{ij}^k$, rather than the temporal signal, $\phi_{ij}^k(t)$ with a similar interpretation holding for \underline{g}_{ij} and \underline{s}_{ij}^k . Note $\underline{\phi}_{ij}^k$, \underline{g}_{ij} , and \underline{s}_{ij}^k are column vectors of size P .

For a given location of the GPR array, to detect mines, we carry out a binary hypothesis test. Under the null hypothesis, H_0 , the received signal $\underline{\phi}_{ij}^k$ is comprised of ground bounce \underline{g}_{ij} plus measurement noise, which is assumed to be a white Gaussian vector, $\underline{w} \sim N(0, \sigma^2 \mathbf{I})$. Under the alternate hypothesis, H_1 , $\underline{\phi}_{ij}^k$ consists of ground bounce, noise, and mine signal, \underline{s}_{ij}^k . Mathematically we have

$$\begin{aligned} H_0^k : \quad & \underline{\phi}_{ij}^k = \underline{g}_{ij} + \underline{w} \\ H_1^k : \quad & \underline{\phi}_{ij}^k = \underline{g}_{ij} + \underline{s}_{ij}^k + \underline{w}. \end{aligned} \quad (7)$$

Our processing method is based on the observation that under H_0 , the M^2 received signal should display certain symmetries, as illustrated by Fig. 4. For example if no mine is present then $\underline{\phi}_{12}$ should be “statistically equal” to $\underline{\phi}_{21}$, $\underline{\phi}_{23}$, $\underline{\phi}_{32}$, $\underline{\phi}_{34}$, and $\underline{\phi}_{43}$ because the ground bounce in each case depends only on the relative spacing of the sensors which is identical for these six pairs. Similarly, $\underline{\phi}_{13}$ should be statistically equal to $\underline{\phi}_{31}$, and so on. By “statistically equal” we mean that any variations in these signal are caused by random sensor noise. In other words, signals from these sets will, on average, possess the same means with some variability (variance) caused by the noise. Thus statistical tests designed to determine homogeneity of a population (i.e. equality of mean vectors) can be used to test whether an object is present (lack of homogeneity) or absent (all the data vectors are about the same). Finally, for the M^2 received signals it is not hard to show that there are only M sets of statistically different signals because of this symmetry. Note the M sets of signals are not of equal size, some sets consist of more signals than the others.

*Here we are assuming that the ground is locally flat over the extent of the sensing system. Extension of the results in this work to smoothly changing ground is an area of current work.

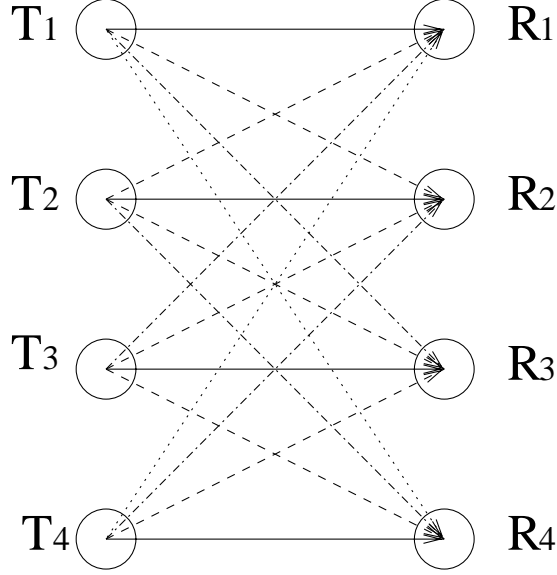


Figure 4. Symmetry of GPR signals, 4 different path lengths are shown in solid line, dashed line, dashed-dot line, and dot line.

3. INFORMATION EXTRACTION ALGORITHM

From the above discussion, we know that if a mine is present, the symmetry of received signals will be disrupted. Therefore, to detect mines, we can look for asymmetry of received signals. Basically, our approach is to sequentially detect any changes in the means of the received signals. For the purpose of illustration, we use $M = 4$ pairs of transmitter and receiver. At each stop k of the GPR array, we then have 4 sets of statistically different signals, $S_1 = \{\phi_{11}^k, \phi_{22}^k, \phi_{33}^k, \phi_{44}^k\}$, $S_2 = \{\phi_{12}^k, \phi_{21}^k, \phi_{23}^k, \phi_{32}^k, \phi_{34}^k, \phi_{43}^k\}$, $S_3 = \{\phi_{13}^k, \phi_{31}^k, \phi_{24}^k, \phi_{42}^k\}$, and $S_4 = \{\phi_{14}^k, \phi_{41}^k\}$.

3.1. Homogeneity Test of Single Set

Here we discuss the homogeneity test of a single set, say $S_1 = \{\phi_{11}^k, \phi_{22}^k, \phi_{33}^k, \phi_{44}^k\}$. We consider the acceptance test of the hypothesis

$$H_0: \underline{\phi}_{ii}^k \text{ in the set are from same multivariate Gaussian distribution, with equal and known covariance matrices, and identical, unknown means, } i = 1, \dots, 4.$$

Because the test is identically the same at each stop k for the remainder of this subsection we drop the k notation.

As the problem currently stands, we do not assume that the ground bounce signals are known. In fact, a simple transformation of the data allows us to perform the test without ever having to know the ground bounces or estimate them. To see this, we begin by forming the collection of all pairwise differences of the signals within the set under consideration. For S_1 we get the set D_1 defined as

$$D_1 = \left\{ \phi_{11} - \phi_{22}, \phi_{11} - \phi_{33}, \phi_{11} - \phi_{44}, \phi_{22} - \phi_{33}, \phi_{22} - \phi_{44}, \phi_{33} - \phi_{44} \right\} \equiv \{\underline{\delta}_1, \underline{\delta}_2, \underline{\delta}_3, \underline{\delta}_4, \underline{\delta}_5, \underline{\delta}_6\}.$$

Testing H_0 using the original set of $\underline{\phi}_{ii}$ signals then is equivalent to testing the following hypothesis using the $\underline{\delta}_n$ vectors:

$$H'_0: \underline{\delta}_n \text{ in the set are from same multivariate Gaussian distribution, with equal and known covariance matrices and means equal to zero}$$

To test this equality of means, we carry out a Hotelling's T^2 -test.⁴ As described more fully in [5], this test amounts to the following comparison

$$\xi_1 = \sum_{n=1}^6 \frac{1}{2} \underline{\delta}_n^T \mathbf{R}^{-1} \underline{\delta}_n \leqslant threshold \quad (8)$$

with $\mathbf{R}(= \sigma^2 \mathbf{I})$ is the covariance matrix of the measurement noise and *threshold* is chosen to ensure an *a priori* specified probability of false acceptance. The hypothesis is accepted if the left hand side is less than or equal to *threshold* and rejected otherwise. Noting that \mathbf{R} is diagonal, we write ξ_1 as

$$\xi_1 = \frac{1}{2\sigma^2} \Phi_1^T \mathbf{A}_1 \Phi_1 = \frac{1}{2\sigma^2} \begin{bmatrix} \underline{\phi}_1^T & \underline{\phi}_2^T & \underline{\phi}_3^T & \underline{\phi}_4^T \end{bmatrix} \begin{bmatrix} 3\mathbf{I} & -\mathbf{I} & -\mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & 3\mathbf{I} & -\mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & -\mathbf{I} & 3\mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & -\mathbf{I} & -\mathbf{I} & 3\mathbf{I} \end{bmatrix} \begin{bmatrix} \underline{\phi}_1 \\ \underline{\phi}_2 \\ \underline{\phi}_3 \\ \underline{\phi}_4 \end{bmatrix}. \quad (9)$$

For a set S_i of size N_i , it is easy to show that \mathbf{A}_i is an $N_i \times N_i$ block matrix. Its block diagonal elements are $(N_i - 1)\mathbf{I}_{P \times P}$ and the off diagonal elements are $-\mathbf{I}_{P \times P}$.

3.2. Homogeneity Test of Multiple Sets

For the application of interest in this paper, we have multiple sets (four for our sample system) for which we wish to test homogeneity. Here we adopt a simple, generalized form of the T^2 test in which the four ξ_i 's are weighted and added to yield one final test statistic,

$$\xi = \xi_1 \eta_1 + \xi_2 \eta_2 + \xi_3 \eta_3 + \xi_4 \eta_4. \quad (10)$$

The η_i are weight factors, defined by

$$\eta_i = \frac{1}{l_i} \left(\sum_{j=1}^4 l_j^{-1} \right)^{-1} \quad (11)$$

where l_j are the path lengths of ground bounces for set S_j . More will be said about this weight factor in Section 3.3. Stacking all the 16 signals ϕ_{ij} to form a long column vector, we obtain a new vector

$$\Phi^T = [\Phi_1^T \ \Phi_2^T \ \Phi_3^T \ \Phi_4^T]$$

and

$$\xi = \frac{1}{2\sigma^2} \Phi^T \mathbf{A} \Phi = \frac{1}{2\sigma^2} \begin{bmatrix} \Phi_1^T & \Phi_2^T & \Phi_3^T & \Phi_4^T \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 \eta_1 & 0 & 0 & 0 \\ 0 & \mathbf{A}_2 \eta_2 & 0 & 0 \\ 0 & 0 & \mathbf{A}_3 \eta_3 & 0 \\ 0 & 0 & 0 & \mathbf{A}_4 \eta_4 \end{bmatrix} \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \\ \Phi_4 \end{bmatrix}. \quad (12)$$

Matrix \mathbf{A} is a block diagonal matrix. Obviously, \mathbf{A} is symmetric. Note \mathbf{A} is positive semidefinite by its buildup. According to Mathai,⁶ the first two moments of ξ are given by,

$$\begin{aligned} \mu_\xi &= (6\eta_1 + 15\eta_2 + 6\eta_3 + \eta_4)P \\ \sigma_\xi^2 &= (24\eta_1^2 + 90\eta_2^2 + 24\eta_3^2 + 2\eta_4^2)P.^\dagger \end{aligned} \quad (13)$$

It has been shown that ξ asymptotically has a Gaussian distribution⁶ for large NP , i.e., $\xi \sim N(\mu_\xi, \sigma_\xi^2)$. The generalized T^2 -test amounts to $\xi \leq \alpha$ where α is the threshold chosen to ensure an *a priori* specified probability P_{fa} of false acceptance. The α is determined by

$$P_{fa} = \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\xi^2}} e^{-\frac{(\xi-\mu_\xi)^2}{2\sigma_\xi^2}} d\xi. \quad (14)$$

3.3. Sequential Detection

Now we consider the actual GPR array operation as the array moves down-track. At each stop k , we calculate a corresponding ξ^k . When there is no mine in the field of view of the array, we record ξ^k only. That is, $\nu^k = \xi^k$ under hypothesis H_0 . When there is a mine, the mine signal \underline{s}_{ij}^k adds to the ground bounce. Passing \underline{s}_{ij}^k to the generalized Hotelling's T^2 -test produces an output signal, denoted by ζ^k . So the problem is to detect signal ζ^k , given the observed signal sequence ν^k and known “noise” ξ^k , using the additive noise model $\nu^k = \xi^k + \zeta^k$ under the alternate hypothesis H_1 . Though ζ^k is unknown and changes with k , it is always *positive* because of the very nature of quadratic form of the generalized T^2 -test. Fig. 5 shows separate ξ_i^k and $\xi_i^k + \zeta_i^k$. Signals on the left side of Fig. 5 are typical “noise” sequences, taken from four different sets of transmitter-receiver combinations. Signals on the right side are “noise” and mine signals, corresponding to the respective sets of transmitter-receiver combinations. It is seen that ζ_i^k are much weaker in set S_3 and S_4 . This observation can be explained by the increased attenuation associated with the longer distances the mine signals in sets S_3 and S_4 travel. For this reason, we introduced the weight factors in Eq. 11.

When the mine is buried deep, the mine signal attenuates exponentially as explained in Section 2.1. Detecting this unknown low power signal can be aided quite a bit using sequential detection methods which retain information from previous scans to improve the SNR. We therefore employ a sequential detector as the GPR array moves down track. At each stop k , the detector makes one of two decisions⁷: (1) Hypothesis H_0 is true, no mine signal is present, (2) Reject H_0 . Because ζ^k causes a positive displacement of the mean of ν^k , we choose a running average of ν^k as a statistical test^{8,9}

$$g^k = \frac{1}{N} \sum_{j=k-N+1}^k \nu^j, \quad k = N, N+1, \dots, K \quad (15)$$

and make a decision by checking $g^k \leq \beta$ where β is a threshold. At each stop k , g^k is compared to the threshold to make a declaration of mine presence. Fig. 6 illustrates relationship between threshold settings and declaration of mine presence as the GPR arrays move down-track. Each filled dot indicates a location of the array where we say that a mine is present. Such a declaration is made when g^k is above the threshold for consecutive 5 stops of k . A lower threshold allows us to make early declaration of mine. The disadvantage is a high false-acceptance rate. The false-acceptance probability P_{fa} equals the probability under hypothesis H_0 that the g^k crosses the threshold. The detection probability P_d equals to $1 - P_{fa}$. By virtue of the generalized T^2 -test, ν^k are statistically independent. Hence, g^k has the Gaussian distribution, $N(\mu_\xi, \frac{1}{N}\sigma_\xi^2)$. Thus, in principle, for a given P_{fa} , β can be determined similarly as in Eq. 14.

[†]See appendix

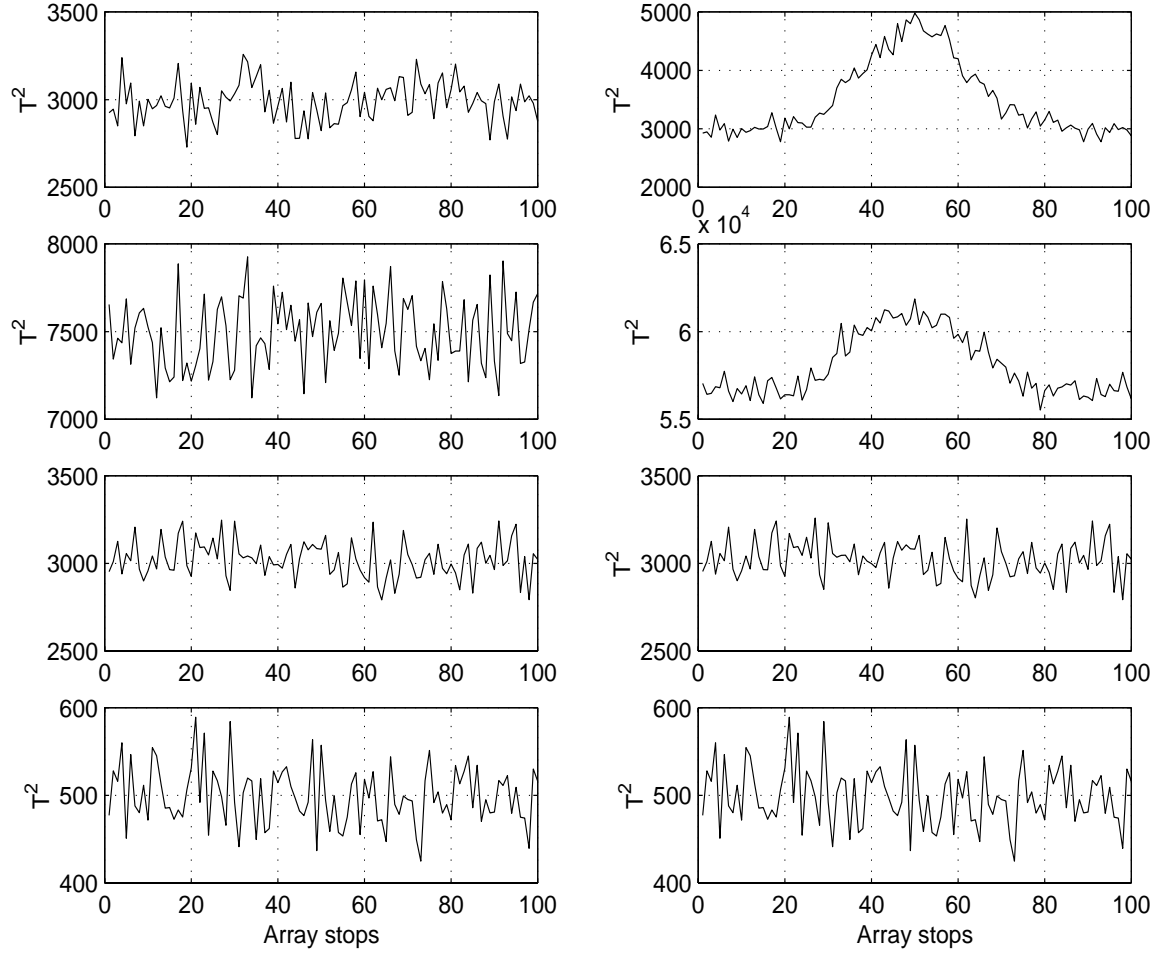


Figure 5. First row of figures corresponds to the T^2 statistic for set S_1 , the left figure is “noise” only, the right figure has both “noise” and mine reflected signal, second row of figures corresponds to set S_2 , third row of figures corresponds to set S_3 , and, fourth row of figures corresponds to set S_4 .

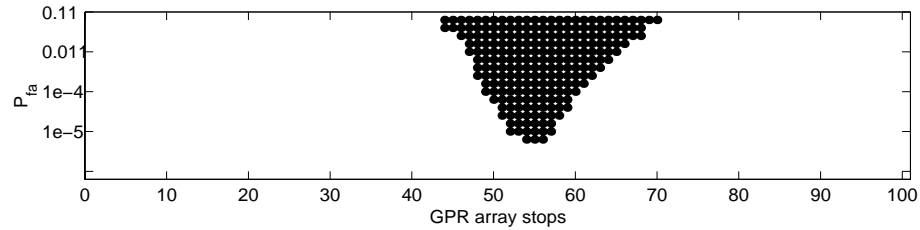


Figure 6. Detection results versus prescribed false alarm probability, P_{fa} . Note that lower P_{fa} results in slower detection while a higher allowable false alarm rate allows for detection prior to the array passing over the mine.

4. EXAMPLES

In this section we consider some examples of the above procedure. Monte-Carlo runs were used to determine the probability of detection and probability of false-alarm for various noise levels and depth of buried mines. In this work, the signal to interference plus noise ratio is defined as

$$\text{SINR} = 10 \log_{10} \frac{\underline{s}^T \underline{s}}{\underline{g}^T \underline{g} + P\sigma^2}. \quad (16)$$

In all cases, we generate synthetic data with an object located around the 50th stops of the GPR array and buried 10 cm underground. The GPR array is composed of four pairs of transmitters and receivers evenly spaced along a baseline width of 80 cm. A transmitter and its corresponding receiver (e.g. transmitter 1 and receiver 1) are 20 cm apart. The GPR array is 40 cm above the ground. For simplicity we assume that the nominal ground bounce and mine-bounce signal take the form of a second derivative of a Gaussian shown in Fig. 7(a). Current work in our group is aimed at developing more sophisticated models for these signals. Fig. 7(b) and (c) show the received signals of two pairs of T/R combinations. Because of the domination of the ground bounce, it is difficult to see any mine signature. Even after pairwise subtraction, the mine signal can hardly be observed in Fig. 7(d). But the generalized T^2 -test can pick up this difference and declare a mine.

Fig. 8(a) shows the receiver operating characteristics of detecting a mine buried 10 cm underground. Fig. 8(b) shows two ROC's for mines buried at different depth under the same noise power. Because of the fast attenuation in soil, the deeper buried mine has a significantly smaller SINR thereby leading to the degradation in performance.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we discuss a statistical approach to detect mines using a GPR array. Exploiting the geometric symmetry of GPR array setting, we have looked into tests for statistical homogeneity of GPR returns as a tool for performing detection. Basically, by treating the received signal as a multivariate Gaussian distribution we test its statistical homogeneity using a generalized, sequential Hotelling's T^2 -test.

Future work will take into consideration of rough ground-air interface and time-delay estimation to actually localize detected mines. An assumption of this work is that the soil conditions are known. Sensitivity analysis involving mismatch in soil parameters will be valuable to apply this method to realistic scenario. Moreover, the additive white Gaussian noise mode will be lifted as we explore issues associated with the modeling of clutter, the incorporation of these models into our processing, and the development of test which are robust (or invariant) to uncertainty in these models.

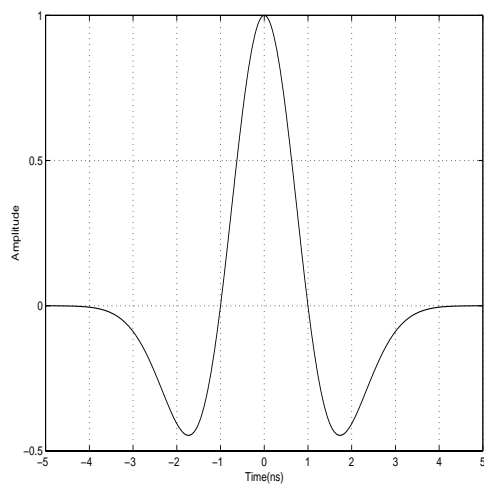
6. APPENDIX

From Mathai's⁶ work, for $\mathbf{X} \sim \mathbf{N}(\mu, \Sigma)$, the first two moments of quadratic $\mathbf{X}^T \mathbf{A} \mathbf{X}$ can be found by

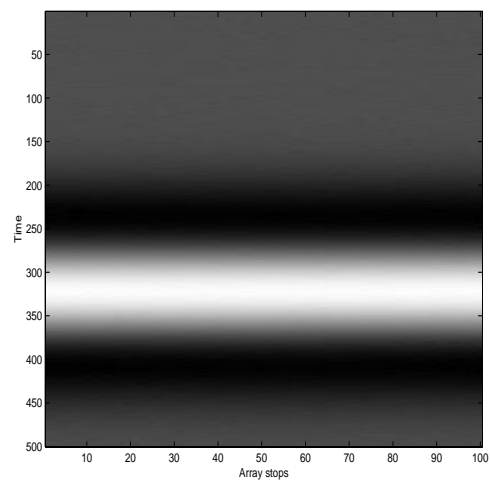
$$\begin{aligned} E[\xi] &= \text{tr}(\mathbf{A}\Sigma) + \underline{\mu}^T \mathbf{A} \underline{\mu} \\ \text{Var}[\xi] &= 2\text{tr}(\mathbf{A}\Sigma)^2 + 4\underline{\mu}^T \mathbf{A} \Sigma \mathbf{A} \underline{\mu}. \end{aligned} \quad (17)$$

For the generalized T^2 -test, we have

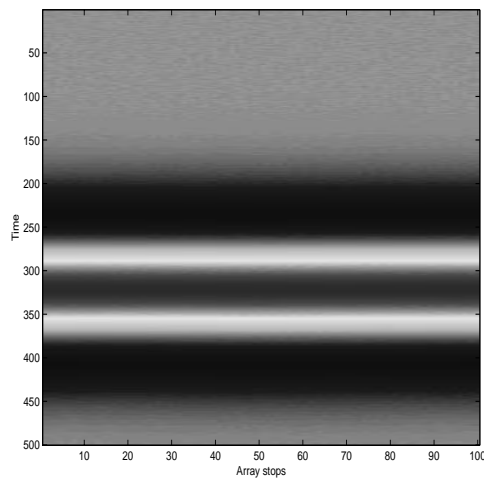
$$\underline{\mu}^T = \left[\underline{\mu}_{11}^T, \dots, \underline{\mu}_{44}^T, \underline{\mu}_{12}^T, \dots, \underline{\mu}_{34}^T, \underline{\mu}_{13}^T, \dots, \underline{\mu}_{24}^T, \underline{\mu}_{14}^T, \underline{\mu}_{41}^T \right] \quad (18)$$



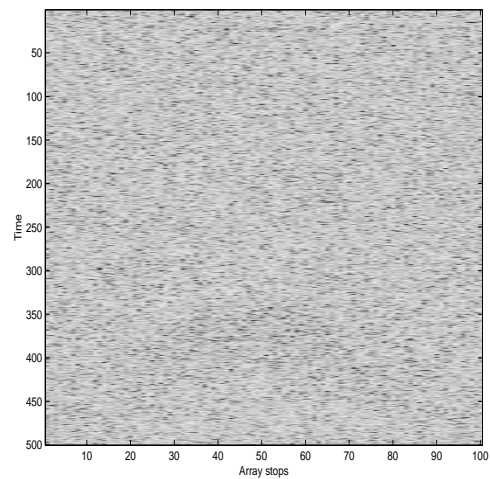
(a) Received signal, the second order derivative of a Gaussian.



(b) Time trace of ϕ_{11}



(c) Time trace of ϕ_{22}



(d) Difference of time traces, $\phi_{11} - \phi_{22}$,
SINR = -11 dB.

Figure 7. Received signal and time traces.

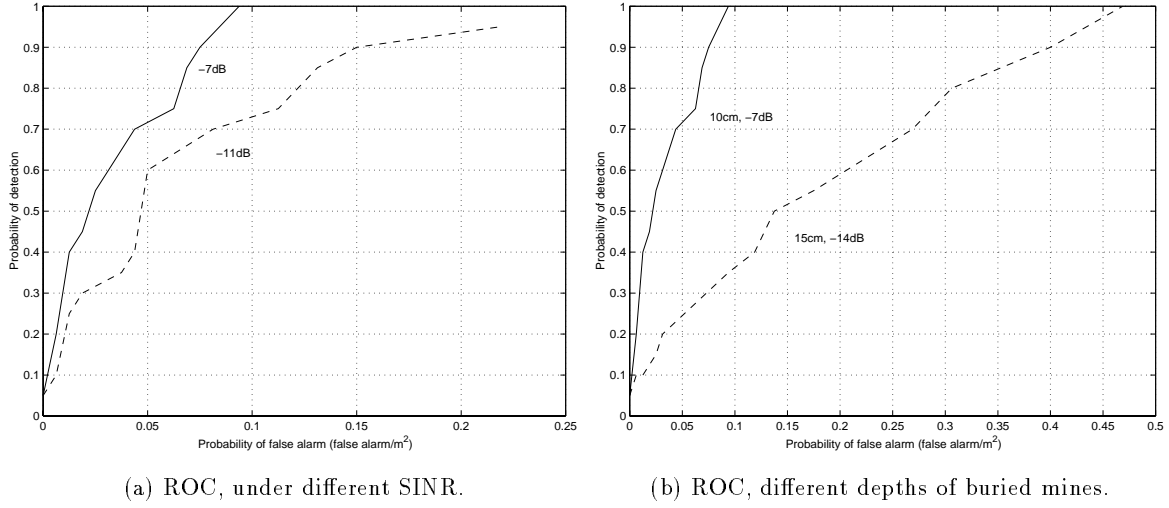


Figure 8. Receiver operating characteristics.

where

$$\begin{aligned}
 \underline{\mu}_{11} = \dots = \underline{\mu}_{44} &\equiv \underline{\mu}^{(1)} & \underline{\mu}_{12} = \dots = \underline{\mu}_{34} &\equiv \underline{\mu}^{(2)} \\
 \underline{\mu}_{13} = \dots = \underline{\mu}_{24} &\equiv \underline{\mu}^{(3)} & \underline{\mu}_{14} = \underline{\mu}_{41} &\equiv \underline{\mu}^{(3)}.
 \end{aligned} \tag{19}$$

Next, we note that $4\underline{\mu}^T \mathbf{A} \Sigma \mathbf{A} \underline{\mu} = 0$ under the condition of Eq. 19,

$$4\underline{\mu}^T \mathbf{A} \Sigma \mathbf{A} \underline{\mu} = 4 \sum_{i=1}^4 \underbrace{\begin{bmatrix} \underline{\mu}^{(i)T} & \dots & \underline{\mu}^{(i)T} \end{bmatrix}}_{N_i} \mathbf{A}_i \Sigma \mathbf{A}_i \begin{bmatrix} \underline{\mu}^{(i)} \\ \vdots \\ \underline{\mu}^{(i)} \end{bmatrix} \tag{20}$$

where

$$\begin{bmatrix} \underline{\mu}^{(i)T} & \dots & \underline{\mu}^{(i)T} \end{bmatrix} \mathbf{A}_i = 0 \tag{21}$$

because of the semidefiniteness of \mathbf{A}_i . Similarly, $\underline{\mu}^T \mathbf{A} \underline{\mu} = 0$. The mean of ξ is then given by $\frac{1}{2\sigma^2} [tr(\mathbf{A} \Sigma)]$. For $\mathbf{A} = \text{block diag} \{ \mathbf{A}_1 \eta_1, \mathbf{A}_2 \eta_2, \mathbf{A}_3 \eta_3, \mathbf{A}_4 \eta_4 \}$ we have

$$E[\xi] = \frac{1}{2\sigma^2} tr(\mathbf{A} \Sigma) = \frac{1}{2} \sum_{i=1}^4 N_i (N_i - 1) P \eta_i. \tag{22}$$

Additionally $\mathbf{A} \mathbf{A} = \text{block diag} \{ \mathbf{A}_1 \mathbf{A}_1 \eta_1^2, \mathbf{A}_2 \mathbf{A}_2 \eta_2^2, \mathbf{A}_3 \mathbf{A}_3 \eta_3^2, \mathbf{A}_4 \mathbf{A}_4 \eta_4^2 \}$ where

$$\mathbf{A}_i \mathbf{A}_i = \begin{bmatrix} (N_i^2 - N_i) \mathbf{I} & -N_i \mathbf{I} & \dots & -N_i \mathbf{I} \\ -N_i \mathbf{I} & (N_i^2 - N_i) \mathbf{I} & \dots & -N_i \mathbf{I} \\ \dots & \dots & \ddots & \dots \\ -N_i \mathbf{I} & -N_i \mathbf{I} & \dots & (N_i^2 - N_i) \mathbf{I} \end{bmatrix}. \tag{23}$$

Here \mathbf{I} is of size $P \times P$ and \mathbf{A}_i is of size $N_i P \times N_i P$. We then have

$$tr(\mathbf{A}_i)^2 = (N_i^2 - N_i)N_i P \quad tr(\mathbf{A})^2 = \sum_{i=1}^4 tr(\mathbf{A}_i)\eta_i^2 = \sum_{i=1}^4 (N_i^2 - N_i)N_i P \eta_i^2.$$

so that $Var[\xi] = \frac{1}{4\sigma^4} 2tr(\mathbf{A}\Sigma)^2 = \frac{1}{2} \sum_{i=1}^4 (N_i^2 - N_i)N_i P \eta_i^2$.

REFERENCES

1. J. Bourgeois and G. Smith, "A Fully Three-Dimensional Simulation of a Ground-Penetrating Radar: FDTD Theory Compared with Experiment," *IEEE Trans. on Geoscience and Remote Sensing* **34**, pp. 36–44, Jan. 1996.
2. C. Rappaport and D. Reidy, "Focused Array Radar for Real Time Imaging and Detection," *Proc. SPIE, Radar Sensor Technology* **2747**, pp. 202–213, 1996.
3. C. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons, Inc., New York, New York, 1989.
4. S. Kullback, *Information Theory and Statistics*, Dover, Mineola, New York, 1997.
5. E. Lehman, *Testing Statistical Hypotheses*, Springer-Verlag, New York, New York, 1986.
6. A. Mathai and S. Provost, *Quadratic Forms in Random Variables: Theory and Applications*, Marcel Dekker, Inc., New York, New York, 1992.
7. C. Helstrom, *Elements of Signal Detection and Estimation*, Prentice-Hall, Englewood Cliffs, New Jersey, 1995.
8. S. Mason, G. Birch, and M. Ito, "Improved Single-Trial Signal Extraction of Low SNR Events," *IEEE Trans. on Signal Processing* **42**, pp. 423–426, Feb. 1994.
9. B. Ghosh, *Sequential Tests of Statistical Hypotheses*, Addison-Wesley, Reading, Mass., 1970.

Localization and characterization of buried objects from multi-frequency, array inductive data

Mustafa Özdemir^a, Eric Miller^a, Stephen Norton^b

^a Center for Electromagnetics Research, Northeastern University, Boston, MA.

^bGeophex, Limited, Raleigh NC

ABSTRACT

The problem of mine localization and characterization from electromagnetic induction data is addressed. We consider processing techniques based on an inductive sensor model originally proposed by Das et. al in [2]. Given this model we examine estimation-theoretic methods for determining an object's center, its orientation, and scattering characteristics (as defined by a spectrum of low order multipole moments) from low frequency spectroscopic data obtained over a grid of spatial locations. Under this model, the data are linear functions of the multipole moment spectra and non-linearly related to object's location and rotation angles. An estimation procedure based on a low-dimensional non-linear optimization routine for the determination of the object center and rotation angles is employed with a linear least squares inversion method used to estimate the multipole moment spectra. Examples are provided for ellipsoidal objects.

1. INTRODUCTION

Electromagnetic induction (EMI) systems represent one of the more common sensing technologies for the detection and localization of buried metallic objects including landmines and unexploded ordinance. While many systems, such as the EM61, operate essentially in the time domain using pulsed induction principals, there has been significant interest recently in the use of swept frequency measurements to perform so-called electromagnetic induction spectroscopy (EMIS). Indeed, work in [1] indicates that data taken over a band from tens of hertz to tens of kilohertz convey information not only about presence of absence of an object but also about object shape, size, orientation, and material characteristics; i.e. such data can in principle be used to perform object characterization.

In this paper, we consider the development of a tractable, physical model to describe an EMIS system and associated processing methods to extract from EMIS data information regarding the location, orientation and structure of a buried object. In Section 2, we describe a generalized form of an EMIS forward model based on the work of Das *et. al.*² Assuming that the incident EMI field is uniform over the support of the object, the model approximates the scattering properties of the object in terms of multipole moment spectra (MMS) which can be used to easily determine the fields observed by the EMI receiver. From an inversion perspective, the idea motivating the work in this paper is that successful estimation of these moment spectra can form the basis for object classification and identification. Also, we note that, unlike general finite element, boundary element, or finite difference type scattering models, the one considered here is particularly well suited for the processing tasks at hand because it is parameterized directly in terms of the quantities of interest: the MMS, the co-ordinates of the object center, and the three rotation angles used to define the orientation of the scatterer relative to a global, Cartesian frame.

Given this model, in Section 3, we describe an algorithm designed to extract this information from EMIS data. Under this model, the data are linearly related to the multipole moment spectra and non-linear

Other author information: MO: mozdemir@cdsp.neu.edu, ELM: elmiller@ece.neu.edu, SJN nortonsj@ornl.gov. This work was supported by the Army Research Office Demining MURI under Grant DAAG55-97-1-0013

functions of the object location and rotation angles. This functional form is exploited in the construction of an estimation procedure based on a low-dimensional non-linear optimization routine required for the determination of the object center and rotation angles (six variables in all). Embedded within this routine is an associated regularized, linear least squares inversion procedure which determines the estimates of the MMS.

Examples demonstrating the performance of this approach are provided in Section 4. They show the simulation results for the sphere and the spheroid object. Finally, in Section 5, we provide conclusions and indicate future work in this effort.

2. PROBLEM FORMULATION AND PHYSICAL MODEL USED

In this paper we consider an extension of a physical model for EMI proposed in [2] describing the scattering of low frequency electromagnetic radiation by spherical or spheroidal objects of known conductivity and permeability. As seen in Fig. 1 the transmitters and receivers are taken to be square coils (not necessarily co-located) with sides of length $2A$. The target center is located at $r_0 = (x_0, y_0, z_0)$ in the $x - y - z$ coordinate system. For the problems of interest in this work the effects of the low conductivity ground typically can be ignored² so that the entire sensor system is taken to reside in free space.

The physical model is based on the assumption that scattering characteristics of the object of interest can be approximated using a low order multipole model. The electromagnetic force EMF, s , induced in a single turn receive coil by an object at r_0 in response to the field setup by a single loop transmitting coil is given as:

$$s = \frac{i\omega\mu_0\beta_0}{I} g^T \Lambda f \quad (1)$$

where g is a 3×1 vector holding the x , y , and z components of the magnetic field produced at r_0 by a current I flowing in the receive coil, g^T indicates the transpose of g , f is the same as g but for the transmitting coil, ω is the operating frequency, $i = \sqrt{-1}$, and μ_0 is the permeability of free space. As described in Appendix A of [2], the vectors g and f are functions only of the relative positions of the object, the receiver, and the transmitter.

The tensor $\beta_0\Lambda$, defines the multipole scattering characteristics of the object. Here Λ is the normalized polarizability tensor, and $\beta_0 = 3\frac{\mu_r-1}{\mu_r+2}$ is the sensitivity factor for a sphere of relative permeability μ_r . In this work we consider targets to be well modeled as ellipsoids. In the event that the target's axes are parallel to those of a global Cartesian co-ordinate system Λ can be represented as

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}. \quad (2)$$

The three frequency dependent λ 's (here referred to as moment spectra) each are associated with one of the principal axes of the ellipsoid. For a sphere, all three are identical and closed form expressions can be found for all orders of multipoles.¹ In [2], scattering from spheroids was considered. In such cases, two of the λ 's are the same, and closed form expressions for their dipole moment structure can only be found in the case of $\omega = 0$. More recently, the work in [3] indicates how one might employ multiple poles in the complex frequency plane to accurately model the scattering process for arbitrary objects. Generally, the problem of determining the moment spectra given the axis lengths and material of the object is an open problem and one which we are currently pursuing. Here we assume that such a correspondence can be found and concentrate instead on the estimation of Λ from a given set of data.

In the event that the ellipsoid is rotated relative to the global co-ordinate system, it is necessary to mathematically express the components of g and f in the frame of the ellipsoid as follows. The field vectors

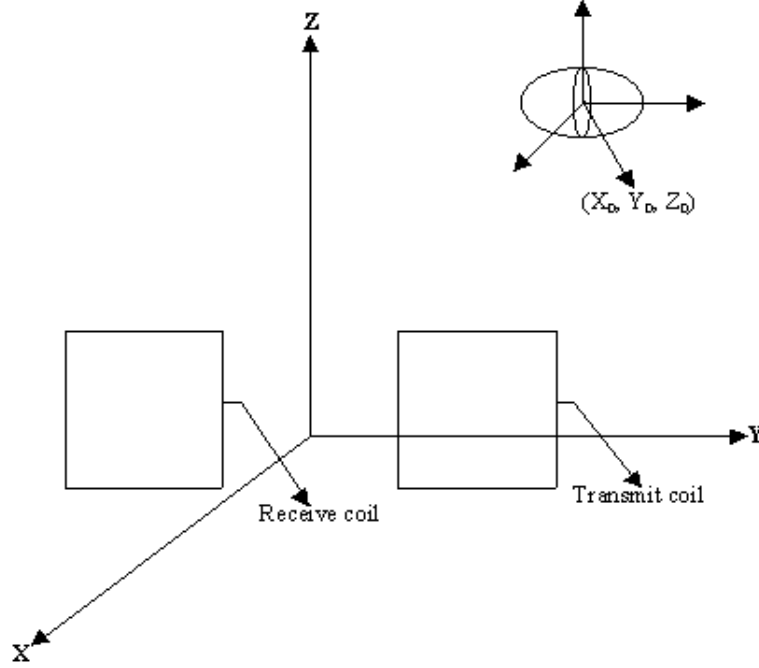


Figure 1. One sensor comprising sensor coils and target object.

f and g are assumed to be expressed in a global Cartesian co-ordinate system $x - y - z$. A second system $x''' - y''' - z'''$, whose axes coincide with the ellipsoid axes can be found by doing the following rotations. It is first rotated through the angle of ϕ about the z -axis. We obtain the new system $x' - y' - z'$. Next the new system is rotated about y' -axis through the angle of θ , resulting in another system $x'' - y'' - z''$. To complete the transformation, the axes are rotated about x'' through the angle of ψ , and we find the last system $x''' - y''' - z'''$, coinciding with the ellipsoid axes. All rotations are shown in Fig. 2. This sequence of transformations is described by the rotation matrix,

$$R = \begin{bmatrix} \cos \theta \cos \phi & \cos \theta \sin \phi & -\sin \theta \\ -\cos \psi \sin \phi + \sin \phi \sin \theta \cos \phi & \cos \psi \cos \phi + \sin \phi \sin \theta \sin \phi & \sin \phi \cos \theta \\ \sin \psi \sin \phi + \cos \psi \sin \theta \cos \phi & -\sin \psi \cos \phi + \cos \phi \sin \theta \sin \phi & \cos \phi \cos \theta \end{bmatrix}. \quad (3)$$

The matrix R is incorporated into the model as follows

$$s = \frac{i\omega\mu_0\beta_0}{I} g^T R^T \Lambda R f \quad (4)$$

where the vector Rf represents the components of the transmitted field in the co-ordinates of the rotated ellipsoid with an analogous interpretation for Rg .

In this work, we are concerned with processing methods based on multi-frequency data obtained from multiple transmitter/receiver locations. Assuming we collect M frequency samples from each of N combinations of transmitters and receivers then we can write the k th frequency sample at the n th position as

$$y_{n,k} = \frac{i\omega_k\mu_0\beta_0}{I} g_n^T R^T \Lambda_k R f_n + w_{n,k} \quad (5)$$

where $w_{n,k}$ is measurement noise. From this data set, our goal is to determine the position, the orientation, and the moment spectra of the buried object.

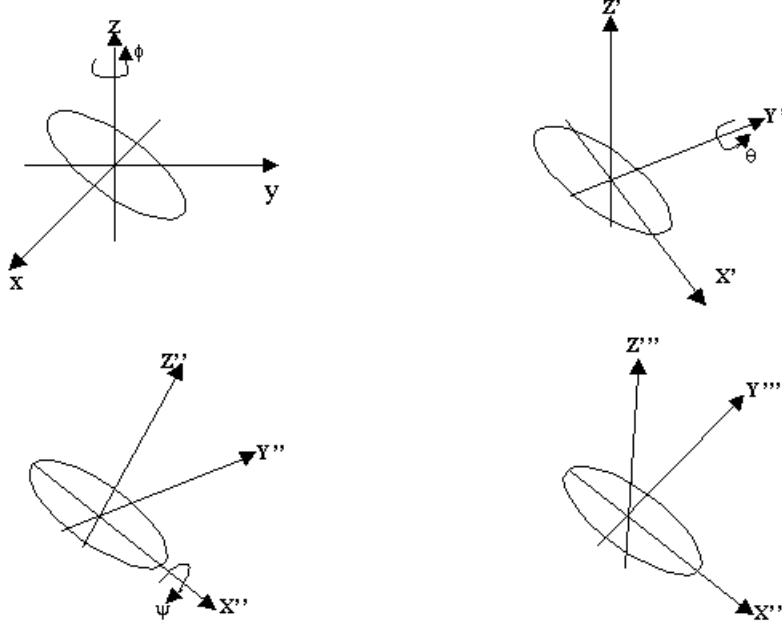


Figure 2. Angle transformations about three coordinate axes.

3. INFORMATION EXTRACTION ALGORITHM

The model developed in the previous section is particularly well suited to the processing task at hand. First, the position of the target appears only in the vectors g and f and the orientation angles are seen in the matrix R . While the data are non-linear functions of these variables (six in all) the analytical nature of the model makes determination of these quantities relatively straightforward using a non-linear optimization routine. More importantly, the shape and electrical characteristics of the object are encoded in the moment spectra λ_i which are linearly related to the data. Thus, determination of these large vectors (three complex valued unknowns per frequency) reduces to a linear least squares problem the solution of which can be obtained in closed form. In this section, we provide a more detailed description of how these observations are exploited in the design of an efficient processing scheme.

First of all, we manipulate the model to a form that is more suitable for processing. Because $\lambda_i(\omega)$ $i = 1, 2, 3$ is a complex quantity in general, so is $y_{n,k}$. Thus, after separating the data into into real and imaginary parts, we make explicit the linear dependence of the data on the multipole moments as follows

$$\begin{bmatrix} y_{n,k}^R \\ y_{n,k}^I \end{bmatrix} = \underbrace{\begin{bmatrix} a_n^R & 0 \\ 0 & a_n^I \end{bmatrix}}_{A_n} \lambda_k + \begin{bmatrix} w_{n,k}^R \\ w_{n,k}^I \end{bmatrix} \quad (6)$$

$$\lambda_k = [\lambda_{1,k}^R \ \lambda_{2,k}^R \ \lambda_{3,k}^R \ \lambda_{1,k}^I \ \lambda_{2,k}^I \ \lambda_{3,k}^I]^T \quad (7)$$

where superscript R indicates real part and superscript I indicates imaginary part. The a_n^R and a_n^I are 1×3 vectors depending on (a) r_0 , (b) the locations of the transmitter and receiver, and (c) the rotation angles, $\alpha = [\theta \ \psi \ \phi]$. These vectors can be obtained from the f 's, R , and g 's after some straightforward

algebra. Finally, the vector λ_k holds the real and imaginary parts of the samples of the three multipole moment spectra for the ellipsoid at frequency ω_k .

Stacking the data from all transmitter-receiver pairs for all frequencies gives the discretized data model:

$$y = A(r_o, \alpha)\lambda + n. \quad (8)$$

For M frequencies $A = I_M \otimes A_1$ where I_N is the $N \times N$ identity matrix, \otimes denotes the Kronecker product and A_1 is the block diagonal matrix obtained from all the A_n 's. Note that if we collect data from a total of N transmitter/receiver pairs then A_i is a $2N$ by 6 matrix and A is $2NM \times 6M$. Finally, we take the noise vector, w , to be zero mean, independent identically distributed Gaussian random variables with variance σ^2 .

Eq. (8) is used in a penalized least squares approach to determine the location of the object, r_o , the orientation angles, α and multiple moments, λ . Estimates of these quantities, denoted as \hat{r}_o , $\hat{\alpha}$, and $\hat{\lambda}$ respectively, are defined as those values which minimize the following cost function:

$$C(r_o, \alpha, \lambda) = \|y - A(r_o, \lambda)\lambda\|_{K^{-1}}^2 + \sum_{i=1}^3 \beta_i \|L_i \lambda\|_2^2. \quad (9)$$

In (9), $K = \sigma^2 I$ is the noise covariance matrix, $\|x\|_A \equiv x^T A x$, and the L_i are used to regularize the problem by enforcing smoothness in the spectra of the multipole moment estimates. Specifically, L_i is built such that

$$\|L_i \lambda\|_2^2 = \sum_{m \in \{R, I\}} \sum_{k=1}^{M-1} (\lambda_{i,k+1}^m - \lambda_{i,k}^m)^2. \quad (10)$$

The regularization parameters β_i in (9) are used to determine the tradeoff in the reconstruction between the two terms in the cost function. The first terms enforces fidelity to the data while the second ensures smooth spectra. By providing for up to three such parameters, we allow for flexibility in adapting the processing structure to the problem at hand. For example, in the case that we knew we were looking for spherical objects then all three λ_i function would be the same and we would require only one β . For spheroidal objects, where two of the axes are the same, only two β_i and λ_i are required: one for the major and one for the minor axis. Finally, we note that in general, the on-line determination of β_i is a well-studied, non-trivial issue beyond the scope of this paper.⁴⁻⁶ For simplicity, in the examples in Section 4, we assume that the β_i are known.

To minimize the cost function, we note first that because (9) is quadratic with respect to λ , $\hat{\lambda}$ can be explicitly stated in terms of α and r_o via

$$\hat{\lambda} = \left(A^T K^{-1} A + \sum_{i=1}^3 \beta_i L_i^T L_i \right)^{-1} A^T K^{-1} y \equiv Q(r_o, \alpha) y \quad (11)$$

so that we can write:

$$\hat{r}_o, \hat{\alpha} = \arg \min_{r_o, \alpha} C(r_o, \alpha, Q(r_o, \alpha) y) \quad (12)$$

$$\hat{\lambda} = Q(\hat{r}_o, \hat{\alpha}) y. \quad (13)$$

In our experiments we have found that C is generally quite well behaved with respect to the location parameters but exhibits many local minima in terms of the orientation angles. Thus, we have adopted the following strategy for first determining rough estimates of r_o and α and then refining these quantities. We

begin by imposing a coarse grid on the three dimensional space of all permitted orientation angles. For each α -value in the grid, a 3D non-linear least squares solver is used to find the optimal r_0 . We use that α values with the smallest overall cost and the associated estimate of r_0 for that cell to initialize a full 6D non-linear least squares scheme to find the final values of $\hat{\alpha}$ and \hat{r}_0 . Using these values, we construct $\hat{\lambda}$ according to (13).

4. EXAMPLES

In this section, the performance of the estimation approach is demonstrated and analyzed under two mine shapes. We simulate data taken on a 10×10 grid of 100 cm^2 pixels by a monostatic transmit/receive system comprised of square coils 5 cm on a side. Ten frequencies logarithmically spaced between 0 and 4.3 KHz are used. One corner of the grid is taken to be $(0,0) \text{ m}$ while the opposite is at $(1,1) \text{ m}$.

As a first example, we consider a spherical mine located at $(x_0, y_0, z_0) = (0.50, 0.50, .10) \text{ m}^*$, and with radius 5 cm . The medium as well as the object are taken to be non-ferrous and the conductivity of the sphere is 10^6 S/m . We assume that the sphere's response can be modeled as a dipole and we use the results of [1] to compute the dipole moment spectrum (DMS). The real and imaginary parts of this spectrum are shown as a solid line in Fig. 3. Because the sphere is rotationally invariant, there is no need to estimate the rotation angles so the problem here reduces to determining the location and the DMS. To demonstrate the performance of our approach, we perform 10 Monte Carlo simulations at a signal to noise ratio of 20 dB . In this case, the sample mean of the estimated object center is $(0.5000, 0.4998, 0.0993) \text{ m}$ with a standard deviation of $\pm(0.0003, 0.0004, 0.0030) \text{ m}$. In Fig. 3 the dotted lines show the sample mean of the estimated DMS with associated error bars. We see from these results that the approach is highly accurate both in terms of estimating the position as well as the moment spectrum.

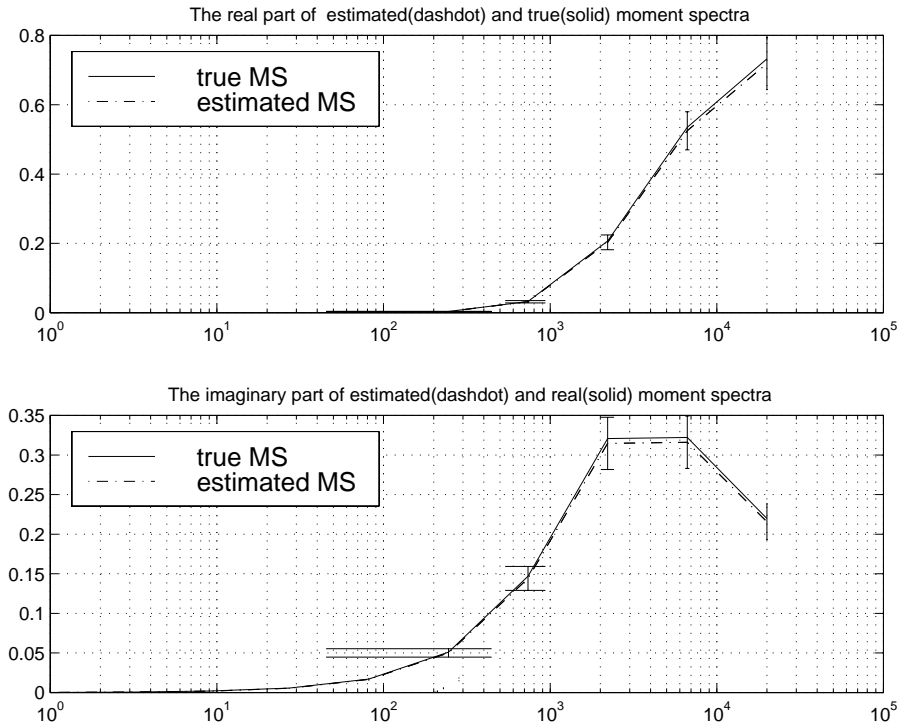


Figure 3. The real and imaginary part of estimated and real moment spectra of sphere mine for $\beta = 0.0001$.

*Increasing depth here corresponds to increasing z

As a second example, we consider a spheroidal object[†] again located at $(x_0, y_0, z_0) = (0.50, 0.50, .10)\text{m}$ and which has been rotated using $\phi = 0.75$ radians and $\psi = 2.30$ radians. In this case, we presently have no closed form expression for the frequency dependent MMS of such an object. However, under the assumption that the scattering characteristics of an eccentric object will be substantially different for the major versus minor axes, we hypothesises MMS spectra shown in Fig. 4 and Fig. 5 as solid lines and examine the performance of our approach under these conditions.

In this case, we estimate the center of the object, minor and major axis moment spectra and two rotation angles. Further, we assume that the optimum β_1 and β_2 values for the major and minor axis are known. For $\beta_1 = 0.001$ and $\beta_2 = 1$, after performing 10 Monte Carlo simulations at 20 dB SNR, the sample mean of the estimated object center is $(0.5002, 0.4995, 0.0969)\text{m}$ with a standard deviation of $\pm(0.0020, 0.0014, 0.0034)\text{m}$. The sample mean of estimated rotation angles are $\phi = 0.7439$ radians and $\psi = 2.3527$ radians with standard deviations of $\pm(0.0424)$ and $\pm(0.1271)$, respectively. The real and imaginary parts of the two estimated moment spectra for the major and minor axes are shown in Fig. 4 and Fig. 5, respectively.

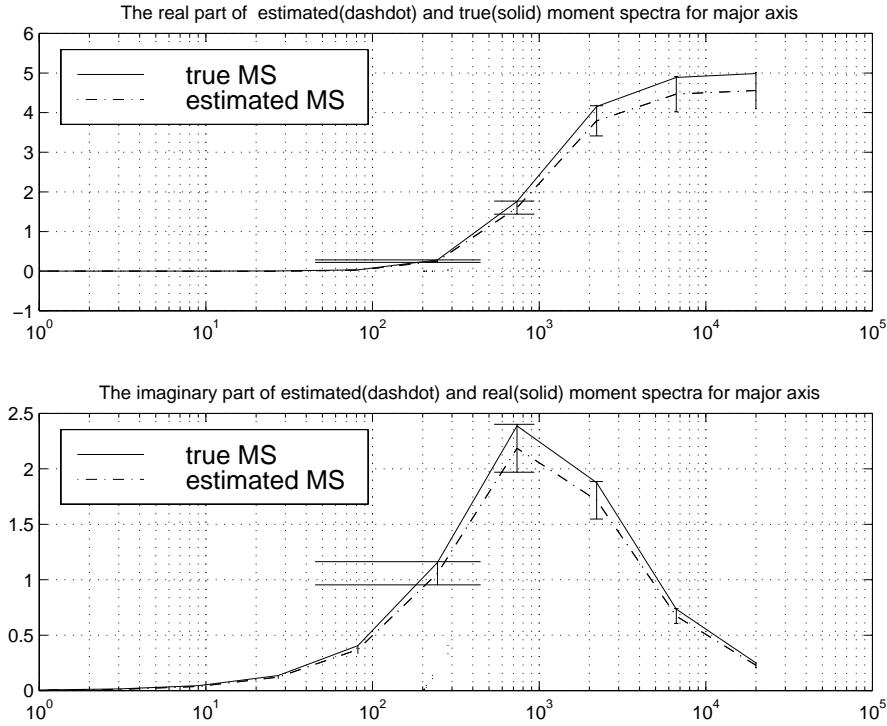


Figure 4. The real and imaginary part of estimated and real moment spectra of spheroid mine for $\beta_1 = 0.001$ and $\beta_2 = 0.001$, and for the major axis.

As a last example, we consider an ellipsoid object at the same location as the previous example. For this case we use the same data as for the second, spheroidal example; however, we lift the assumption that we know we are looking for a spheroidal object and instead estimate three MMS and three rotation angles. For $\beta_1 = 0.001$, $\beta_2 = 0.001$ and $\beta_3 = 1$, after we perform 10 Monte Carlo simulations at 20 SNR, the sample mean of the estimated object center is $(0.5003, 0.4992, 0.0911)\text{m}$ with a standard deviation of $\pm(0.0013, 0.0015, 0.0046)\text{m}$. The sample mean of the estimated rotation angles are $\phi = 0.7443$ radians,

[†]Note that because the object is taken to be spheroidal, two of the three principle axes are identical so we only need estimate a pair of rotation angles and a pair of MMS.

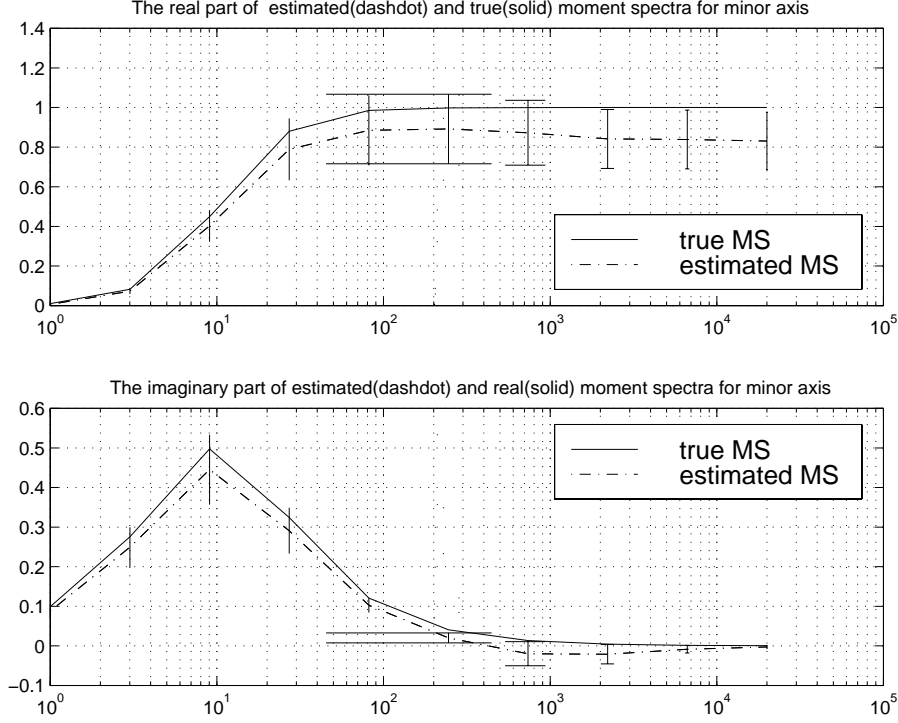


Figure 5. The real and imaginary part of estimated and real moment spectra of spheroid mine for $\beta_1 = 0.001$ and $\beta_2 = 1$, and for the minor axis.

$\psi = 2.5231$ and $\theta = 0.0619$ radians with standard deviations of $\pm(0.3015)$, $\pm(0.0841)$ and $\pm(0.2652)$ respectively. The real and imaginary parts of the estimated moment spectra for the three axes are shown in Fig. 6, Fig. 7 and Fig. 8, respectively. From Fig. 6 and Fig. 7 we see that the MMS estimates are approximately the same for the two equal axes. From this observation and the fact that the ϕ estimate was so close to zero, we conjecture that one could use these results in a classification scheme to conclude reliably that the object under investigation possessed spheroidal symmetry. Constructing such a classifier is a project we are currently pursuing.

The sample means of the estimated object center and rotation angles for three examples are illustrated in the following table with standard deviations. From all results the approach estimating the position, the rotation angle and the moment spectrum is highly precise.

	Sphere Case		Spheroid Case		Ellipsoid Case	
	mean	Standard Deviation	mean	Standard Deviation	mean	Standard Deviation
\hat{x}_0	0.5000	0.0003	0.5002	0.0020	0.5003	0.0013
\hat{y}_0	0.4998	0.0004	0.4995	0.0014	0.4992	0.0015
\hat{z}_0	0.0993	0.0030	0.0969	0.0036	0.0911	0.0046
$\hat{\phi}$	-	-	0.7439	0.0424	0.7563	0.3015
$\hat{\psi}$	-	-	2.3527	0.1271	2.5982	0.0841
$\hat{\theta}$	-	-	-	-	0.0619	0.2652

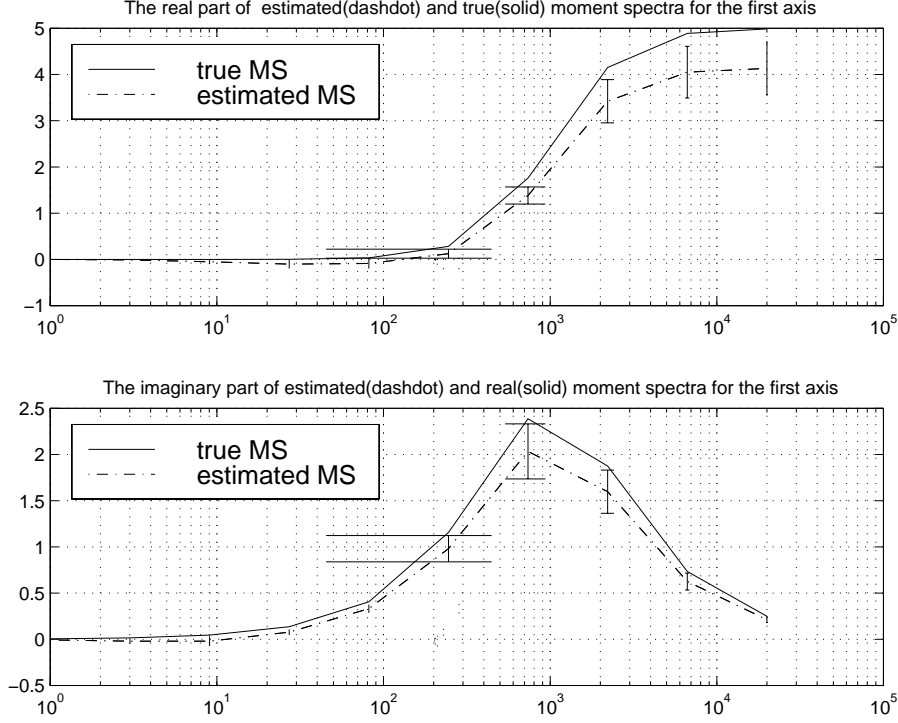


Figure 6. The real and imaginary part of estimated and real moment spectra of spheroid mine for $\beta_1 = 0.001$, $\beta_2 = 0.001$ and $\beta_3 = 1$, and for the first axis.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a technique for the estimation of the MMS, the co-ordinates of the object center, and the rotation angles from EMIS data. Under this approach, the data are linearly related to the multipole moment spectra and non-linear functions of the object location and rotation angles. We determined the object center and rotation angles by using a low-dimensional non-linear optimization method and employed a linear least square inversion procedure to determine the estimates of the MMS.

While the results in this paper are encouraging much work remains to be done in this area. First, we are currently looking to verify our work by applying these methods to real sensor data in collaboration with Geophex Inc. Second, the closed form analytical nature of the model makes it well suited to extensive performance analysis based on Cramer-Rao lower bounds on the variances of the estimates we obtain for angles, location, and MMS. Using this performance metric allows one to start looking at issues of optimizing sensor configurations for particular detection/characterization problems. Moreover, for much of this work we assumed that we knew whether the targets of interest possessed spherical or ellipsoidal symmetry. More interesting is the case where we estimate three rotation angles and three moment spectra and employ a statistical test to determine the symmetry characteristics of the underlying target. Again, performance analysis is also of interest. Finally, as has been mentioned previously in this paper, from a modeling perspective we are currently looking to techniques for mapping object characteristics (size, shape, and material parameters) into the λ functions used in this model. As any such mapping represents an approximation to the true physics, it would be interesting to explore methods for doing this which explicitly minimize the error in the approximation.

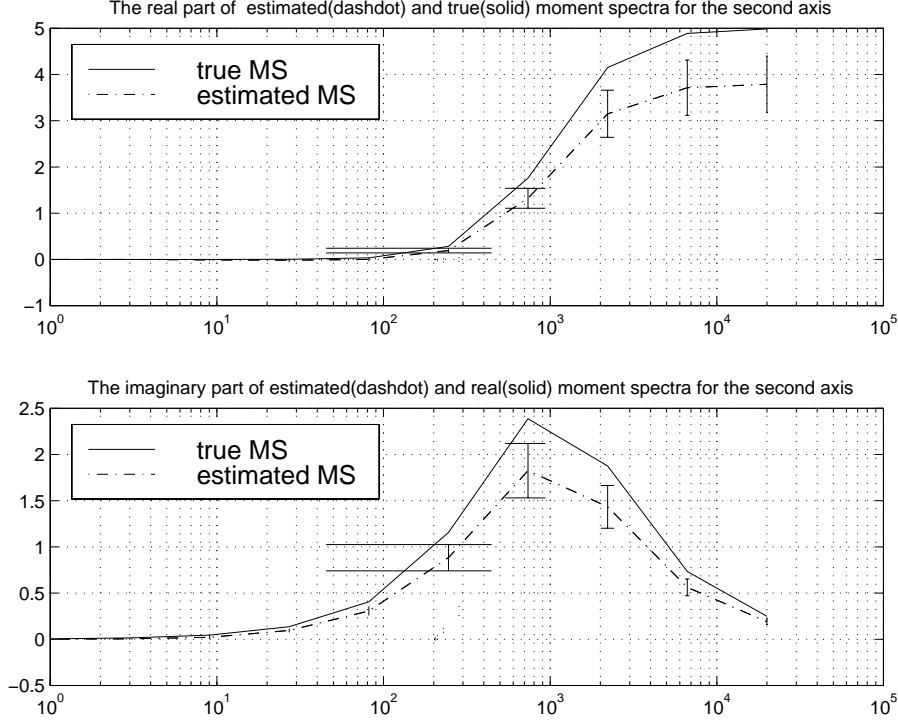


Figure 7. The real and imaginary part of estimated and real moment spectra of spheroid mine for $\beta_1 = 0.001$, $\beta_2 = 0.001$ and $\beta_3 = 1$, and for the second axis.

REFERENCES

1. I. Won, D. Keiswetter, and E. Novikova, "Electromagnetic induction spectroscopy," *JEEG* **3**, pp. 27–40, March 1998.
2. Y. Das, J. E. McFee, J. Toews, and G. C. Stuart, "Analysis of an electromagnetic induction detector for real-time localization of buried objects," *IEEE Trans. Geoscience and Remote Sensing* **28**, pp. 278–287, May 1990.
3. N. Geng, C. E. Baum, and L. Carin, "On the low-frequency natural response of conducting and permeable targets," *IEEE Trans. Geoscience and Remote Sensing* **37**, pp. 347–359, January 1999.
4. P. C. Hansen, "The use of l-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.* **14**, pp. 1487–1503, 1993.
5. P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Review* **34**, pp. 561–580, December 1992.
6. M. Belge, M. E. Kilmer, and E. L. Miller, "Simultaneous multiple regularization parameter selection by means of the l-hypersurface with applications to linear inverse problems posed in the wavelet domain," in *Proceedings of SPIE'98-Bayesian inference for inverse problems*, vol. 3459, July 1998.

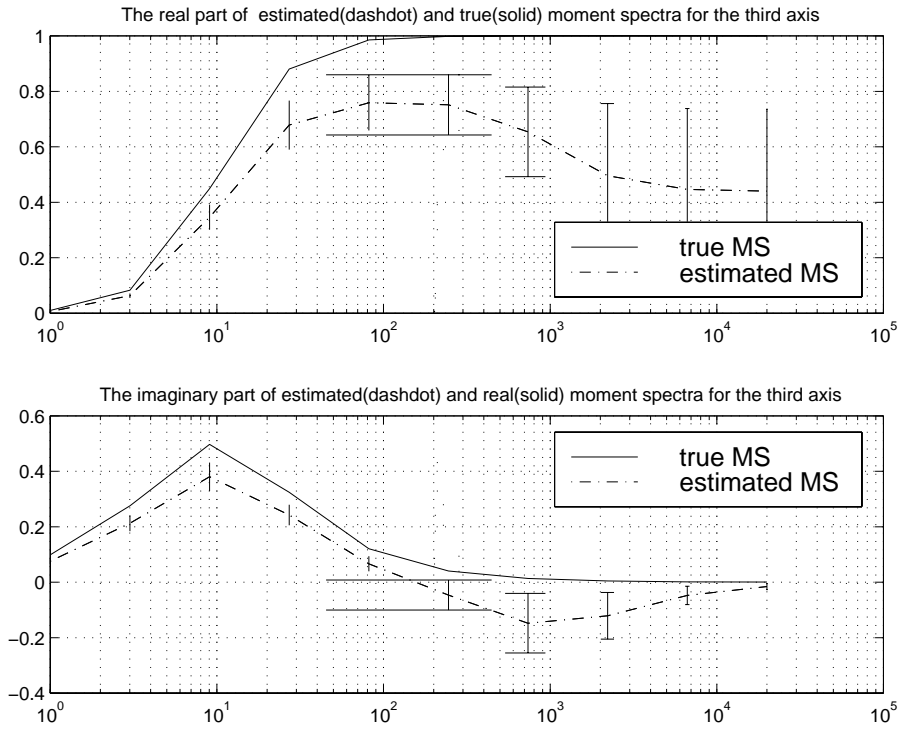


Figure 8. The real and imaginary part of estimated and real moment spectra of spheroid mine for $\beta_1 = 0.001$, $\beta_2 = 0.001$ and $\beta_3 = 1$, and for the third axis.

Recursive T-Matrix Methods for Scattering from Multiple Dielectric and Metallic Objects*

Adnan Şahin Eric L. Miller
Center for Electromagnetics Research,
235 Forsyth Building, 360 Huntington Ave.
Northeastern University, Boston, MA 02115
Telephone: (617) 373-8386
Telefax : (617) 373-8627
email: adnan@cdsp.neu.edu

February 14, 1997

Abstract

We present an efficient, stable, recursive T-matrix algorithm to calculate the scattered field from a heterogeneous collection of spatially separated objects. The algorithm is based on the use of higher order multipole expansions that are typically employed in recursive T-matrix techniques. The use of these expansions introduces instability in the recursions developed in [1, 3] specifically in the case of near field computations. By modifying the original recursive algorithm to avoid these instabilities we arrive at a flexible and efficient forward solver appropriate for a variety of scattering calculations. The algorithm can be applied when the objects are dielectric, metallic, or a mixture of both. We verify this method for cases where the scatterers are electrically small (fraction of λ) or relatively large ($1-2 \lambda$). While developed for near field calculation, this approach is applicable for far field problems as well. Finally, we demonstrate that the computational complexity of this approach compares favorably with comparable recursive algorithms.

Submitted to *IEEE Trans. on Antennas and Propagation* in February, 1997.

*This work was supported in part DOE contract DE-FC07-95ID13395, NSF Grant MIP-9623721, and by sub-contract GC123920NDG from Boston University under the AFOSR MURI Program on Reduced Signature Target Recognition

1 Introduction

Calculation of scattered electromagnetic fields is of interest in many different science areas. For example, an important component of many ground penetrating radar (GPR) problems is the efficient computation of the scattered fields produced by a collection of buried objects when illuminated by a radar source. The choice of technique for computing these fields is often driven by a variety of factors including computational complexity and the flexibility to handle easily a wide range of configurations of scatterers. In applications such as mine detection and hazardous waste removal objects can be either dielectric, metallic or mixtures. Their sizes can range from sub-wavelength to a few multiples of a wavelength. Therefore, one desires an efficient, flexible forward solver that is useful both for analysis and that can be incorporated into signal processing algorithms.

The most popular forward solver for these and related complex scattering problems, the Method of Moments (MoM) [6], is based on a fine discretization of the region of interest and requires the inversion of a large dense matrix to calculate the scattered field. As this task requires $O(N^3)$ calculations where N is the number of grid points, MoM is known to be quite computationally intensive. While fast multipole techniques [9] are useful for reducing the complexity of MoM-type linear systems, these algorithms are typically used for metallic objects and have theoretical difficulties with the spurious modes that can be circumvented only by “complexification” [12]. Finite difference techniques are also frequently used as forward solvers and like MoM rely on a full space discretization. Although the resulting matrices are sparse, one still faces the delicate task of specifying an absorbing boundary condition to terminate the computational grid.

Here we consider the use of transition matrix (T-matrix) methods [14,15] for solving the scattering problems of interest. Unlike MoM and finite differences, the T-matrix approach does not require an absorbing boundary condition and substitutes the discretization of space with harmonic

expansions of the fields thereby reducing the number of unknowns for a wide range of problems. For scattering problems involving single objects, this technique is applicable for metallic and dielectric scatterers. For problems involving multiple objects, Chew and co-workers have pioneered the development of a number of fast, recursive T-matrix algorithms for determining the scattered fields in a variety of scenarios [1–4, 8, 13]. For example, in [3], problems involving electrically large dielectric objects are considered. By tessellating the objects into many small sub-scatterers and using low order multipole expansions of the fields for each sub-scatterer the authors arrive at a highly efficient, T-matrix based algorithm for computing the scattered fields. In [4], Chew et.al. consider a scattering problem involving a group of metallic strips. Here the method of moments is used to compute the T-matrices for each, individual strip and the same recursion as in [13] is employed to solve the overall, multi-object scattering problem.

The motivation for the algorithm developed in this paper is the need to solve scattering problems for the GPR-type geometry shown in Fig. 1. Specifically, we are interested in the development and verification of a recursive algorithm capable of computing scattered fields from multiple dielectric and/or metallic objects in the near field. For simplicity, we considered an E_z polarized planewave incident on a two dimensional problem geometry in which multiple scatterers each possessing a circular cross-section (i.e. infinite circular cylinders) are located in an infinite medium of constant, complex permittivity. Because the use of a tessellated scheme is inappropriate for metallic objects, to handle this mixed-object problem we formulate a recursive algorithm based on high order harmonic expansions for the full scatterers. We demonstrate that this approach causes instabilities in the original recursive algorithm when near field computations are required. By modifying these recursions we obtain a stable algorithm which avoids these instabilities and which is capable of accurate near and far field calculations for the mixed scatterer problem of interest. Finally, we

demonstrate that this approach retains the low computational complexity of the method in [3].

The remainder of this paper is organized as follows. In Section 2 we review the T-matrix theory for single scatterers and the recursive T-matrix algorithm for multiple scatterers. In Section 3, we discuss how T-matrix techniques can be applied to GPR geometries, and we will present two alternative approaches to calculate the scattered field and the modification in the recursive algorithm. In Section 4, we will discuss the results and show examples and finally in Section 5 we will draw conclusions and suggest future work.

2 T-Matrix Background

2.1 Single Scatterer T-matrix

The total scalar electromagnetic (or acoustic) wave in a homogeneous background with a homogeneous scatterer is given by¹:

$$\psi(\underline{r}) = \psi^{inc}(\underline{r}) + \psi^{sca}(\underline{r}) \quad (1)$$

where $\psi^{inc}(\underline{r})$ is the wavefield incident on the scatterer and $\psi^{sca}(\underline{r})$ is the field scattered from the object.

Applying the Poincaré-Huygens principle and the Gauss theorem we can write the total field outside the scatterer as [11]:

$$\psi(\underline{r}) = \psi^{inc}(\underline{r}) + \int_S dS' \{ \psi_+(\underline{r}') \nabla' g(k|\underline{r} - \underline{r}'|) - [\nabla'_\perp \psi(\underline{r}')] g(k|\underline{r} - \underline{r}'|) \} \quad (2)$$

where S is a piecewise smooth surface enclosing the scatterer, ψ_+ and $\nabla'_\perp \psi$ are the total field $\psi(\underline{r})$ and its gradient on the outer surface of the scatterer, and $g(k|\underline{r} - \underline{r}'|)$ is the free space Green's function. The vectors \underline{r} and \underline{r}' are from the scattering origin to observation points and to source points on the scatterers, respectively. The Green's function can be expanded in terms of cylindrical

¹Time factor of $e^{j\omega t}$ is suppressed

basis functions (in 2-D) as follows [11]:

$$g(k|\underline{r} - \underline{r}'|) = -j\frac{\pi}{2} \sum_n \psi_n(k\underline{r}_{>}) Rg\psi_n(k\underline{r}_{<}) \quad (3)$$

where $\psi_n(k\underline{r}) = H_n^{(2)}(kr)e^{-jn\phi}$ are the basis functions representing traveling waves and $Rg\psi_n(k\underline{r}) = J_n(kr)e^{-jn\phi}$ are the basis functions representing the standing waves. Here, Rg stands for “regular part of”, $H_n^{(2)}(z)$ is the n th order Hankel function of second kind, $J_n(z)$ is the n th order Bessel function, and $\underline{r}_{>}$ ($\underline{r}_{<}$) means the larger (smaller) of \underline{r} and \underline{r}' .

Based on the same decomposition, the scattered and incident fields can be expanded as [11]:

$$\psi^{sca}(\underline{r}) = \sum_n f_n \psi_n(\underline{r}) = \underline{\psi}^T \underline{f} \quad \text{if } |\underline{r}| > |\underline{r}'| \quad (4)$$

and

$$\psi^{inc}(\underline{r}) = \sum_n a_n Rg\psi_n(\underline{r}) = Rg\underline{\psi}^T \underline{a} \quad (5)$$

where $\underline{\psi}$ and $Rg\underline{\psi}$ are column vectors filled with $\psi_n(\underline{r})$ and $Rg\psi_n(\underline{r})$, respectively and T stands for transposition.

The T-matrix now is defined as [11,14]:

$$\underline{f} = \mathbf{T}\underline{a}. \quad (6)$$

The elements of \mathbf{T} can be found by using (2), (3) and the boundary conditions. For a detailed analysis of the single object T-matrix method, the reader is referred to [14,15,11].

2.2 Recursive T-matrix Algorithm

The recursive T-matrix algorithm uses the basic principle of single scatterer T-matrix formulas in that for each object, the scattered fields from others are assumed a part of total incident field. This way for every scatterer a T-matrix can be assigned. The recursion starts with the T-matrices of individual scatterers, then one by one scatterers are incorporated into the equation and the T-matrices are updated until, for every scatterer, the final form of the T-matrix, including all multiple scattering effects, is obtained.

Formally, for L scatterers, the harmonic expansion of scattered field, similar to (4), can be written as [3]:

$$\psi^{sca}(\underline{r}) = \sum_{i=1}^L \underline{\psi}^T(\underline{r}_i) \mathbf{T}_{i(L)} \boldsymbol{\beta}_{i,0} \underline{a} \quad (7)$$

where $\mathbf{T}_{i(L)}$ is the T-matrix for i th object in the presence of L scatterers and $\boldsymbol{\beta}_{i,0}$ is the translation matrix used to translate same type basis functions between scattering coordinate center (x_s, y_s) and i th object's local coordinate center (x_i, y_i) , i.e. *standing* waves in i th local coordinate system to *standing* waves in 0th (scattering) coordinate system; or *traveling* waves in i th local coordinate system to *traveling* waves in 0th (scattering) coordinate system.² Fig. 1 pictorially shows the coordinate systems and how the translation matrices work. Expansion of the scattered field in (7) is valid if all observation points are outside the circle enclosing all scatterers. Following Chew's derivation, the recursive construction of $\mathbf{T}_{i(L)}$ can be written as [3, eq.10-11] :

$$\mathbf{T}_{n+1(n+1)} \boldsymbol{\beta}_{n+1,0} = \left[\mathbf{I} - \mathbf{T}_{n+1(1)} \sum_{i=1}^n \boldsymbol{\alpha}_{n+1,i} \mathbf{T}_{i(n)} \boldsymbol{\beta}_{i,0} \boldsymbol{\alpha}_{0,n+1} \right]^{-1} \mathbf{T}_{n+1(1)} \left[\boldsymbol{\beta}_{n+1,0} + \sum_{i=1}^n \boldsymbol{\alpha}_{n+1,i} \mathbf{T}_{i(n)} \boldsymbol{\beta}_{i,0} \right] \quad (8)$$

and

$$\mathbf{T}_{i(n+1)} \boldsymbol{\beta}_{i,0} = \mathbf{T}_{i(n)} \boldsymbol{\beta}_{i,0} + \mathbf{T}_{i(n)} \boldsymbol{\beta}_{i,0} \boldsymbol{\alpha}_{0,n+1} \mathbf{T}_{n+1(n+1)} \boldsymbol{\beta}_{n+1,0} \quad (9)$$

where $n = 1, 2, \dots, L, i = 1, 2, \dots, n$ and $\boldsymbol{\alpha}_{n,i}$ is the translation matrix used to change different basis functions between reference coordinate systems (Fig. 1), i.e. *standing* waves in n th local coordinate system to *traveling* waves in i th local coordinate system. The recursion starts with the individual T-matrices, $\mathbf{T}_{i(1)}$, of the scatterers, i.e. the T-matrix of the i th scatterer when there are no other scatterers in the medium.

Theoretically the matrices $\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{T}$ are of infinite dimension. T-matrix algorithms truncate these matrices with finite values N and M such that the residual error is below the machine precision

²The translation matrices $\boldsymbol{\beta}_{i,0}$ contain Bessel functions and complex exponentials. For details about these matrices see [1, 11].

or acceptable levels. Here N represents the number of harmonics used to expand the fields at the scattering origin and M represents the number of harmonics used to expand the fields in the objects' local coordinate systems. Thus, the T-matrix is of size $M \times M$, $\beta_{i,0}$ is of size $M \times N$ and $\alpha_{i,n+1}$ is of size $M \times M$. The parameters N and M are related to the distance of scatterers from the scattering origin and the radii of the scatterers, respectively. As the distances between scatterers and the scattering origin increase, N needs to be increased, and as the radii of scatterers increase, M needs to be increased [3].

3 A Modified Recursive T-Matrix Method

The work in this paper was motivated by the desire to obtain a fast, accurate forward modeling code for ground penetrating radar type geometries illustrated in Fig. 2. As discussed in Section 1 this application requires the computation of near field values of scattered field arising from mixtures of dielectric and metallic objects. To effectively handle these requirements, we propose a formulation of the recursive T-matrix algorithm based on the representation of the scattered field from each full object using high order expansions (i.e. large M) in the recursions in (8) and (9).

In principle, this approach supports the computation of scattered fields from arbitrary collections of dielectric and metallic objects. In fact, we demonstrate that this is true specifically for far-field calculations. Unfortunately, the use of higher order expansions results in an instability in a particular harmonic expansion formula upon which the original recursive T-matrix algorithm is based when near field computations are required. In the remainder of this section, we describe explicitly this difficulty and propose a modified recursion which by-passes this addition formula and results in a stable method for solving the problem of interest.

3.1 Determination of Scattering Origin

Unlike most radar applications, in a GPR measurement geometry the scattered field is generally observed in the near or intermediate field. Since the harmonic expansions upon which the recursive T-matrix algorithm is based have validity regions (see eqn.(4)), there are certain limitations as to where the scattering origin can be placed relative to the receiver array. In this section, we will briefly discuss how the scattering origin is determined, when the object locations and radii are given for the GPR-type configuration in Fig. 2. The triplet $(x_i, y_i; a_i)$ represents x and y coordinates and radius of the i th object relative to the global origin O_g and L is the number of objects buried under the receiver array.

Because of the requirements on the loci of observation points imposed by (4) for single objects and (7) for multiple objects, the scattering origin (x_s, y_s) relative to O_g must be selected such that there must be at least one circle, centered at (x_s, y_s) , encircling all objects with no receivers inside it. The dashed circle in Fig. 2 depicts such a circle. Assuming a linear receiver array, the condition to choose the scattering coordinate system is:

$$R_c < |y_s| \quad (10)$$

where

$$R_c = \max_{i \in \{1, 2, \dots, L\}} \left\{ \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2} + a_i \right\}. \quad (11)$$

This condition must be met by individual objects as well as by all objects collectively. Therefore, we can rewrite the condition in (10) and (11) as the intersection of regions as follows:

$$(x_s, y_s) \in \left\{ (x, y) \mid \bigcap_{i=1}^L \sqrt{(x - x_i)^2 + (y - y_i)^2} < |y| - a_i \right\}. \quad (12)$$

In fact, each term under the intersection sign in (12) defines the region under an upside-down parabola. Fig. 3 depicts the parabolic regions for three objects. Placement of objects in this figure is very typical of a mine detection problem. In this geometry, any point inside the shaded

area, representing the intersection of all three parabolic regions, can be selected as the scattering origin. Ideally, we can place the scattering origin at $y_s \approx -\infty$. This choice of (x_s, y_s) will always satisfy the condition in (10). However, the order of harmonics used in the T-matrix algorithm is proportional to the distance between scattering origin and object centers [3], i.e. $N \propto kr_i$ where N is the harmonic used for translations to and from the scattering origin, k is the wave number and $i = 1, 2, \dots, L$. Therefore, the optimum scattering origin should be within this shaded area and as close as possible to the objects in order to minimize the harmonics used for translations. As we show in Section 3.2, with this choice of (x_s, y_s) , the distances between object centers and the scattering origin can be very close, which causes convergence problems in the addition formulas of T-matrix algorithm. In Section 3.3, we describe a modification in the recursive T-matrix algorithm that lets us use the algorithm with optimum choice of scattering origin.

3.2 Problems With Higher Order Harmonic Expansions

The convergence problems alluded to earlier can be traced to the fact that equation (8) uses the identity

$$\alpha_{p,q} = \beta_{p,0} \alpha_{0,q} \quad \text{if } |\underline{r}_q| \geq |\underline{r}_p| \quad (13)$$

which in turn requires the ordering of the objects such that $|\underline{r}_1| \leq |\underline{r}_2| \leq \dots \leq |\underline{r}_L|$. By using definitions of $\alpha_{p,q}$, $\beta_{p,0}$ and $\alpha_{0,q}$ [1,11], we can write the (m, m') th entry, $[\alpha_{p,q}]_{m,m'}$, as:

$$H_{m-m'}^{(2)}(k|\underline{r}_{pq}|)e^{-j(m-m')\phi_{pq}} = \lim_{N \rightarrow \infty} \sum_{n=-N}^N J_{m-n}(k|\underline{r}_p|)e^{-j(m-n)(\phi_p+\pi)} H_{n-m'}^{(2)}(k|\underline{r}_q|)e^{-j(n-m')\phi_q} \quad (14)$$

where $\underline{r}_{pq} = |\underline{r}_{pq}|e^{-j\phi_{pq}} = \underline{r}_q - \underline{r}_p$ and $\underline{r}_i = |\underline{r}_i|e^{-j\phi_i}$, $i = p, q$. This truncated sum does not converge if $\underline{r}_q = \underline{r}_p + \underline{\delta}$ where $|\underline{\delta}|$ is small as compared to $|\underline{r}_p|$ and $|\underline{r}_q|$, and if $m - m'$ is a large number ($-M \leq m \leq M$ and $-M \leq m' \leq M$.) Fig. 4 shows the convergence of the series in (14) for the corner entries of (13) for $M = 5$, i.e. $\max\{m - m' = 10\}$. Here we have three curves, showing the

convergence for $\underline{\delta} = 0.1r_p$, $\underline{\delta} = 0.25r_p$ and $\underline{\delta} = 0.5r_p$, $M \geq 5$ and $\underline{\delta} < 0.1r_p$ are typical parameter choices for the problems of interest in this paper. It is clear from this figure that as the magnitudes of two vectors get closer, the convergence rate slows. Chew et.al. [4] suggested a windowed addition theorem (which is originally developed for H_z polarized scattering) to overcome this problem, but the implementation of this method introduces two new variables to choose in order to set the width and shape of the window. In addition, the implementation of windowed summation introduces errors in the sum for vectors for which the convergence is not a problem.

It should be noted that not all valid scattering origins for a given problem give rise to this convergence problem. Indeed, trial an error will quickly demonstrate that, for a given collection of scatterers, there exist scattering origins where the original T-matrix recursions work just fine. These points are typically far from the scatterers thereby requiring large N in the recursions and moreover there does not appear to be an easy means of *a priori* determining whether a chosen origin will or will not give rise to a coverage difficulty. Thus, in the following sections, we introduce a modified recursion which bypasses the convergence issue for all valid scattering origins thereby allowing us to use the closest valid origin (i.e. smallest N) to solve the problem.

3.3 Modified Recursive T-Matrix Algorithm

The recursion in (8) and (9) takes place over the quantities $\mathbf{T}_{i(n)}\boldsymbol{\beta}_{i,0}$, and we have determined that the convergence problem stems from (13). Therefore, to eliminate the need to use this identity, we go one step back in the derivation of recursion formulas, and write (8) as [3, eq.7-8]:

$$\mathbf{T}_{n+1(n+1)}\boldsymbol{\beta}_{n+1,0} = \left[\mathbf{I} - \mathbf{T}_{n+1(1)} \sum_{i=1}^n \boldsymbol{\alpha}_{n+1,i} \mathbf{T}_{i(n)} \boldsymbol{\alpha}_{i,n+1} \right]^{-1} \mathbf{T}_{n+1(1)} \left[\boldsymbol{\beta}_{n+1,0} + \sum_{i=1}^n \boldsymbol{\alpha}_{n+1,i} \mathbf{T}_{i(n)} \boldsymbol{\beta}_{i,0} \right] \quad (15)$$

and (9) as:

$$\mathbf{T}_{i(n+1)}\boldsymbol{\beta}_{i,0} = \mathbf{T}_{i(n)} \left[\boldsymbol{\beta}_{i,0} + \boldsymbol{\alpha}_{i,n+1} \mathbf{T}_{n+1(n+1)} \boldsymbol{\beta}_{n+1,0} \right] \quad (16)$$

without using (13). Since (13) is not used in (15) and (16) we can base a new recursion on these two equations and the identity:

$$\beta_{i,0}\beta_{0,i} = \mathbf{I} \quad (17)$$

where $\beta_{i,0}$ is $M \times N$, $\beta_{0,i}$ is $N \times M$, and (17) holds as long as $N > M$ which is always true as long as objects are not overlapping. By using (15), (16) and (17) the modified recursion equations can be written as:

$$\mathbf{T}_{n+1(n+1)}\beta_{n+1,0} = \left[\mathbf{I} - \mathbf{T}_{n+1(1)} \sum_{i=1}^n \alpha_{n+1,i} \mathbf{T}_{i(n)} \beta_{i,0} \beta_{0,i} \alpha_{i,n+1} \right]^{-1} \mathbf{T}_{n+1(1)} \left[\beta_{n+1,0} + \sum_{i=1}^n \alpha_{n+1,i} \mathbf{T}_{i(n)} \beta_{i,0} \right] \quad (18)$$

and

$$\mathbf{T}_{i(n+1)}\beta_{i,0} = \mathbf{T}_{i(n)}\beta_{i,0} + \mathbf{T}_{i(n)}\beta_{i,0}\beta_{0,i}\alpha_{i,n+1}\mathbf{T}_{n+1(n+1)}\beta_{n+1,0}. \quad (19)$$

Note that the recursion is still over the same block, $\mathbf{T}_{i(n)}\beta_{i,0}$, but since (13) is eliminated these new recursion equations do not suffer from convergence problems.

As reported in [3] the original recursive T-matrix algorithm has a complexity of $O(M^2N)$ per recursion. It is easily shown that the modified algorithm also has a complexity of $O(M^2N)$ per recursion with a slightly larger constant in front of M^2N resulting from extra multiplications to obtain $\mathbf{T}_{i(n)}$ from $\mathbf{T}_{i(n)}\beta_{i,0}$. To calculate the scattered field from L objects, $L(L-1)/2$ recursions are required. Therefore, the overall complexity of both the recursive and modified recursive algorithm is $O(L^2M^2N)$.

4 Discussion And Examples

In this section, we first verify our new scattering algorithm against published results and then provide a collection of examples that are particularly relevant for near field, GPR-type applications. As most previously published results for mixed scatterer problems involve far field computations, in verifying our approach we also demonstrate its ability to handle far zone calculations. Where

appropriate, we compare the computational complexity of our higher order, modified recursive algorithm (HO-MRA) against two alternate T-matrix approaches. First, we implement the lower order, original recursive algorithm (LO-ORA) of [3] for near and far field, dielectric-only problems. For far zone problems with mixtures of dielectric and metallic scatterers, we consider high order (i.e. large M) forms of the original recursions (labeled HO-ORA here) (8) and (9), where, because of the far field assumption, the instability problem is not an issue.

Before we proceed, we define the terms used in this section. The echo width, i.e. scattering cross-section per unit length, and normalized echo width are defined as [10]:

$$\sigma(\phi) = \lim_{r \rightarrow \infty} 2\pi r \left| \frac{\psi^{sca}(\underline{r})}{\psi^{inc}(\underline{r})} \right|^2, \quad (20)$$

and

$$\sigma_n(\phi) = \sqrt{\frac{\sigma(\phi)}{\lambda}} \quad (21)$$

where λ is the wavelength in the medium of propagation. The normalized scattering field pattern is defined as:

$$F(\phi) = 10 \log_{10} \left\{ \lim_{r \rightarrow \infty} 2\pi r \frac{|\psi^{sca}(\underline{r})|^2}{\max\{|\psi^{sca}(\underline{r})|^2\}} \right\}. \quad (22)$$

In order to ensure that the modified algorithm can indeed find the true scattered fields, we verified our calculations against published scattered field patterns. We first calculated the scattered field due to two dielectric cylinders placed in free space, each with relative dielectric constant of 2.6, and radius of 0.5λ . The distance between the cylinders is 3λ (Fig. 5(a).) An E_z polarized planewave is incident from 0° . Fig. 5(b) shows the echo width calculated using the HO-MRA of this paper (solid line), the LO-ORA of [3] (dashed line) and results in [10] (circles). Fig. 6(b) shows a similar comparison for a mixed object case depicted in Fig. 6(a), i.e. one cylinder is metallic and the other is lossy dielectric with $\epsilon_r = 4 - j5$. For this case, we did not include the echo width calculated using the original recursive algorithm, since the method in [3] is limited to dielectric

objects. As in previous examples, the echo width obtained using the modified algorithm and that reported in [10] are very close.

Next, we compared the scattering patterns for two metallic cylinders. In this case both cylinders have a radius of 1.1λ ($ka = 7$) and separated by a distance of 2.6λ ($kd = 16$.) Fig. 7(a) shows the scattering geometry for this example and (b) shows the scattered field patterns of both our solution (solid line) and the one given in [5] (circles.) As seen from these figures, the scattering pattern obtained using the modified algorithm and that given in [5] are very close.

Now, we present scattering examples that are especially useful in GPR applications. All objects are assumed to be buried in a homogeneous, lossy background ($\epsilon_b = 6\epsilon_0$, $\sigma_b = 5 \times 10^{-2} S/m$; typical properties of 5% moist San Antonio clay loam or 10% moist Puerto Rico clay loam [7]); the operating frequency is $1GHz$ and a planewave is incident from 90° , see Fig. 8(a). We first find the scattered field from 3 dielectric mine-like objects with diameters $7.5cm$ as shown in Fig. 8(a). All objects have a relative permittivity of 2.5. The scattering origin has to be placed far away from the receiver array ($x_s = 0.5m$, $y_s = -1.25m$), because the objects are close to the receivers, which in turn requires a large value, 120, for N . For this case, we calculated the scattered field using both the LO-ORA and HO-MRA defined in Section 2.2 and Section 3.3, respectively. Fig. 8(b) shows the scattered fields observed along the receiver array using the HO-MRA (solid line) and the LO-ORA (circles). It is clear from this figure that both approaches yield very similar fields but the computational complexity (flop count) of our method is 18.94×10^6 flops while that of the tessellated scheme is 1671×10^6 flops.

The second GPR example depicts a mixed mine-like object case since the objects at the sides are metallic and the object at the center is dielectric with a relative dielectric constant of 2.5, Fig. 9(a). The mine-like geometry is unchanged and the scattering origin is still at ($x_s = 0.5m$, $y_s = -1.25m$).

As a result $N = 120$, and since the object radii are relatively small $M = 12$. The scattered field observed along the receiver array for mixed mine-like object case is shown in Fig.9(b). The last example demonstrates the calculation of scattered field from buried waste drums. For this case we have 2 metallic drums of radius $0.3m$ buried in the same lossy background before, as shown in Fig. 10(a). The scattering origin is placed at $(x_s = 1m, y_s = -1.37m)$ to minimize the harmonics used in the expansions ($M = 25, N = 110$). For drum case M is quite large since the radii of the objects are considerably large. Fig. 10(b) shows the scattered field observed along the receiver array placed directly above the cylinders.

Having verified the scattering field patterns of new recursion with the ones in the literature and presented the GPR examples, we compared the complexities of the HO-MRA, LO-ORA, and HO-ORA. To ensure a fair comparison, whenever a tessellation is required, we set the density of sub-scatterers to be close to that used in [3]. Performance of the each approach is measured by the floating point operations (flops) required to calculate the scattered field. Table 1 shows the flop count of all three recursive T-matrix algorithms that can be used to find the scattered fields from multiple, spatially separated cylinders. Table 2 shows the number of scatterers L , harmonics M , N and the location of the scattering origin (x_s, y_s) used in these examples.

The first three rows of Tables 1 and 2 correspond to examples from the two dimensional scattering literature. For these cases, all observation points are in the far field so that the convergence problem alluded to earlier is not an issue. As a result, with dielectric objects HO-MRA, LO-ORA and HO-ORA can be used to calculate the scattered field. LO-ORA is used only with dielectric objects [3] and as seen from Table 1 its computational complexity is quite large as compared to HO-MRA and HO-ORA. The reason behind this large complexity is that numerous sub-scatterers are required for each cylinder. The computational complexity of HO-ORA is at most 15% less than that of

HO-MRA since the latter needs extra multiplications to obtain $\mathbf{T}_{i(n)}$ from $\mathbf{T}_{i(n)}\boldsymbol{\beta}_{i,0}$.

The last three rows of Table 1 show the flops needed to find the scattered field for GPR-specific examples and Table 2 shows the number of scatterers, harmonics and the locus of the scattering origin used in these examples. The geometries of GPR cases of interest are depicted in Figures 8(a), 9(a) and 10(a). Unlike previous examples, GPR problems require measuring the scattered field in the near field, which restricts the regions where the scattering origin can be placed. As we have shown in Section 3, the choice of optimum scattering origin results in convergence problems in HO-ORA, making it inaccessible for GPR geometries. In addition LO-ORA is not used with metallic objects leaving only HO-MRA for all GPR geometries and all material types. Even when LO-ORA is used for dielectric-only objects, one has to spend approx. 88 times more flops than it is needed for HO-MRA (Table 1).

5 Conclusions

In this paper, we present a new recursive T-matrix algorithm specifically designed for the efficient solution of near field scattering problems involving heterogeneous collections of metallic and dielectric objects. We have verified this algorithm against previously published results thereby demonstrating its utility for far field computations and indicated its use for GPR-type scattering problems. For near and far-field dielectric scattering problems, this algorithm is significantly more efficient than the sub-scatterer method in [3]. For far-field computations, the technique in this paper is slightly more costly than the use of higher order expansions in the original recursive formulae.

The work in this paper suggests a variety of additional research directions. First, we are quite interested in extending the modified recursive algorithm into three dimensions while simultaneously considering scattering problems involving irregularly shaped objects where the T-matrices would be

computed using the method of moments as in [4]. Such work would allow for the easy examination as to how object shape and orientation impacts the scattered fields and ultimately, the ability to detect and localize objects. In terms of the GPR application which originally motivated this effort, we are looking to T-matrix type methods which might allow for some level of modeling the air-earth interface relevant in these scenarios without destroying the computational efficiency of the scattering model. Finally, applying the near field computational abilities of this approach to other application areas would be quite interesting.

References

- [1] W.C. Chew. *Waves and Fields in Inhomogeneous Media*. Van Nostrand Reinhold, 1990.
- [2] Weng Cho Chew, James A. Friedrich, and Robert Geiger. A multiple scattering solution for effective permittivity of a sphere mixture. *IEEE Trans. Geoscience and Remote Sensing*, 28(2):207–214, March 1990.
- [3] Y.M. Wang W.C. Chew. An efficient algorithm for solution of a scattering problem. *Microwave and Optical Technology Letters*, 3(3):102–106, March 1990.
- [4] W.C. Chew Y.M. Wang L. Gürel. Recursive algorithm for wave-scattering solutions using windowed addition theorem. *Journal of Electromagnetic Waves and Applications*, 6(11):1537–1560, 1992.
- [5] A.Z. Elsherbeni M. Hamid. Scattering by parallel conducting circular cylinders. *IEEE Trans. on Antennas and Propagation*, 35(3):355–358, March 1987.
- [6] R.F. Harrington. *Field Computation by Moment Methods*. IEEE Press, (Originally published: Malabar, Fla. :R.E. Krieger, 1968), 1993.
- [7] J.E. Hipp. Soil electromagnetic parameters as functions of frequency, soil density, and soil moisture. *Proceedings of the IEEE*, 62(1):98–103, January 1974.
- [8] W.C. Chew L. Gürel Y.M. Wang G. Otto R.L. Wagner Q.H. Liu. A generalized recursive algorithm for wave-scattering solutions in two dimensions. *IEEE Trans. on Microwave Theory and Techniques*, 40(4):716–722, April 1992.
- [9] V. Rokhlin. Rapid solution of integral equations of scattering theory in two dimensions. *Journal of Computational Physics*, 86(2):414–439, 1990.
- [10] M. Ouda M. Hussein A. Sebak. Multiple scattering by dielectric cylinders using a multi-filament current model. *Journal of Electromagnetic Waves and Applications*, 7(2):215–234, 1993.
- [11] B. Peterson S. Ström. Matrix formulation of acoustic scattering from an arbitrary number of scatterers. *J. Acoust. Soc. Am.*, 56(3):771–780, September 1974.
- [12] N. Engheta W.D. Murphy V. Rokhlin M.S. Vassiliou. The fast multipole method (FMM) for electromagnetic scattering problems. *IEEE Trans. on Antennas and Propagation*, 40(6):634–641, June 1992.
- [13] W.C. Chew Y.M. Wang. A fast algorithm for solution of a scattering problem using a recursive aggregate τ matrix method. *Microwave and Optical Technology Letters*, 3(5):164–169, May 1990.
- [14] P.C. Waterman. New formulation of acoustic scattering. *J. Acoust. Soc.*, 45(6):1417–1429, 1969.
- [15] P.C. Waterman. Symmetry, unitarity and geometry in electromagnetic scattering. *Physical Review D*, 3(4):825–839, February 1971.

	HO-MRA	LO-ORA	HO-ORA
Fig. 5	0.6	1150	0.55
Fig. 6	1.9	n/a	1.76
Fig. 7	8.84	n/a	7.73
Fig. 8	18.94	1671	c/p
Fig. 9	18.95	n/a	c/p
Fig. 10	29.5	n/a	c/p

Table 1: Complexity Comparison for Recursive Algorithms. All numbers in FLOPS/ 10^6 , n/a means not applicable and c/p means convergence problems

	HO-MRA	LO-ORA	HO-ORA
Fig. 5	L=2,M=7,N=23 (x_s, y_s)=(0,0)	L=398,M=1,N=23 (x_s, y_s)=(0,0)	L=2,M=7,N=23 (x_s, y_s)=(0.5 λ , 0.5 λ)
Fig. 6	L=2,M=10,N=40 (x_s, y_s)=(0,0)	n/a	L=2,M=10,N=40 (x_s, y_s)=(0.5 λ , 0.5 λ)
Fig. 7	L=2,M=20,N=40 (x_s, y_s)=(0,0)	n/a	L=2,M=20,N=40 (x_s, y_s)=(λ , λ)
Fig. 8	L=3,M=12,N=120 (x_s, y_s)=(0.5,-1.25)	L=93,M=3,N=120 (x_s, y_s)=(0.5,-1.25)	c/p
Fig. 9	L=3,M=12,N=120 (x_s, y_s)=(0.5,-1.25)	n/a	c/p
Fig. 10	L=2,M=25,N=110 (x_s, y_s)=(1,-1.37)	n/a	c/p

Table 2: Parameter list for Table 1

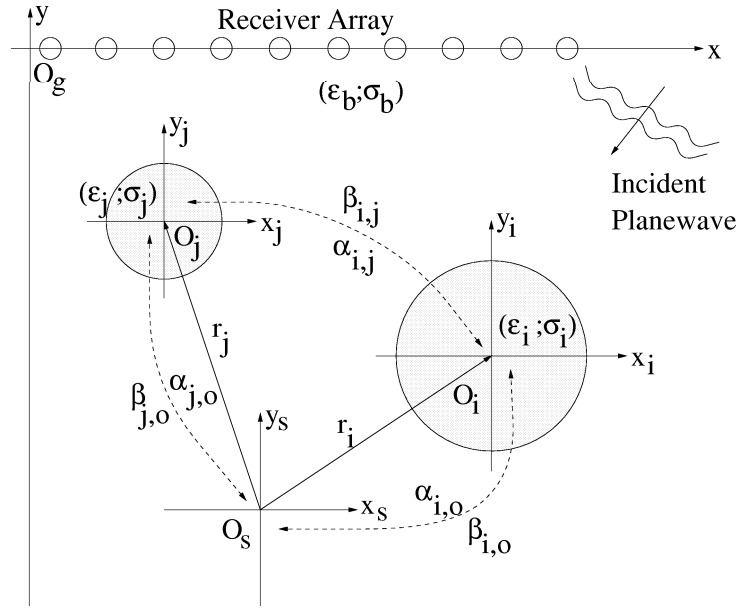


Figure 1: GPR geometry and translation matrices

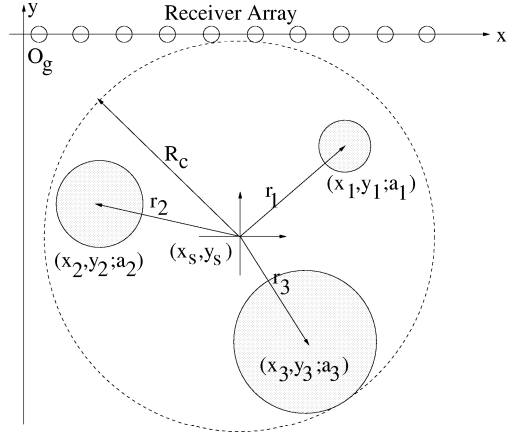


Figure 2: Scattering origin relative to the scatterers, the global origin O_g and the receivers

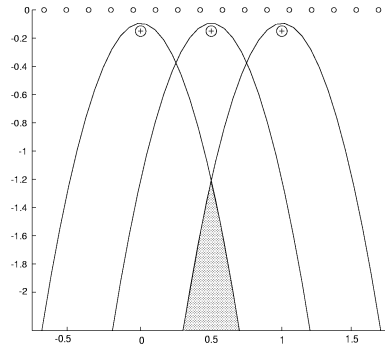


Figure 3: Scattering Origin Regions

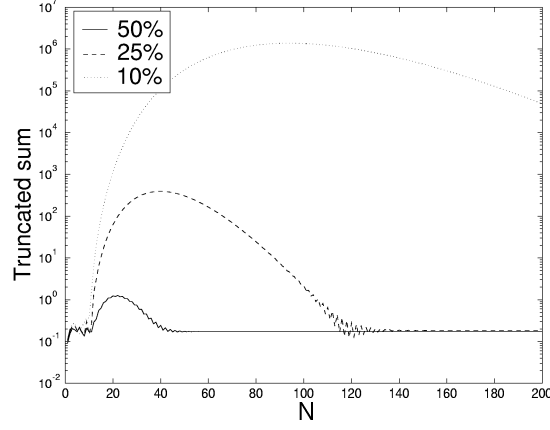
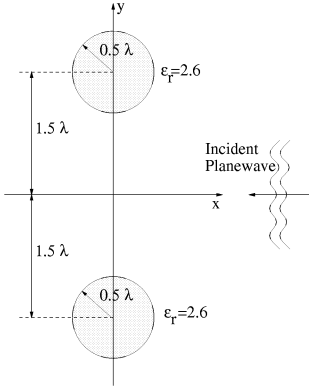
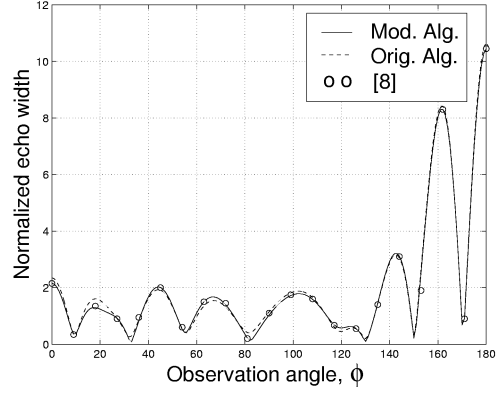


Figure 4: Convergence pattern of the truncated sum in (14) for $M=5$. Curves show the convergence for $\underline{\delta} = 0.1\underline{r}_p$, $\underline{\delta} = 0.25\underline{r}_p$ and $\underline{\delta} = 0.5\underline{r}_p$.

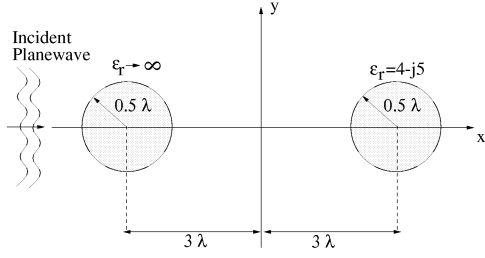


(a) Scattering geometry

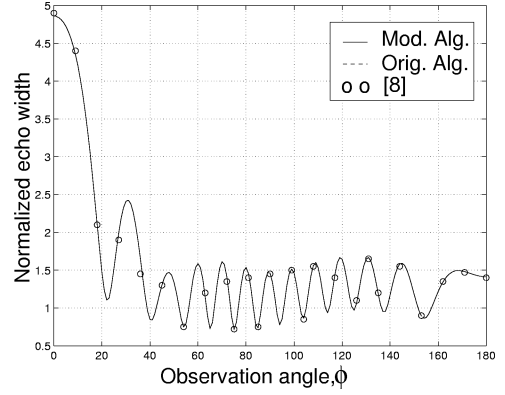


(b) Normalized echo width for geometry of (a)

Figure 5: Comparison of echo width with [10] for two equal dielectric cylinders

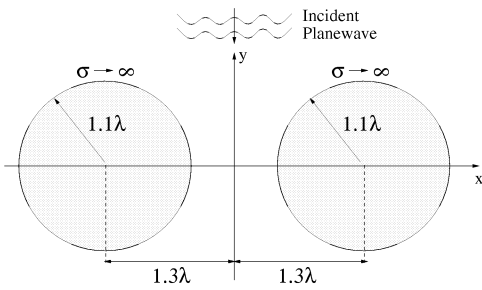


(a) Scattering geometry

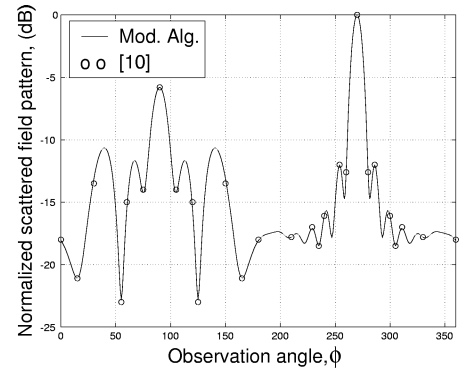


(b) Normalized echo width for geometry of (a)

Figure 6: Comparison of echo width with [10] for two cylinders, one lossy dielectric and one metallic

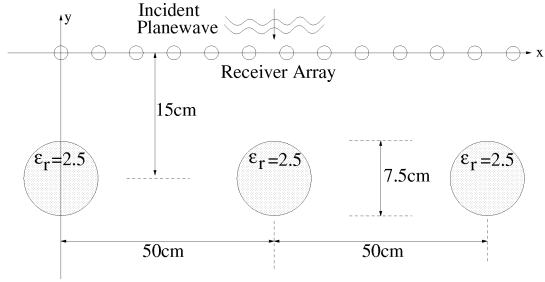


(a) Scattering geometry

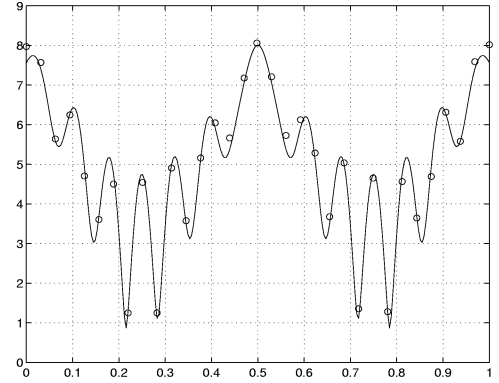


(b) Scattered field pattern for geometry of (a)

Figure 7: Comparison of scattered field pattern with [5] for two metallic cylinders

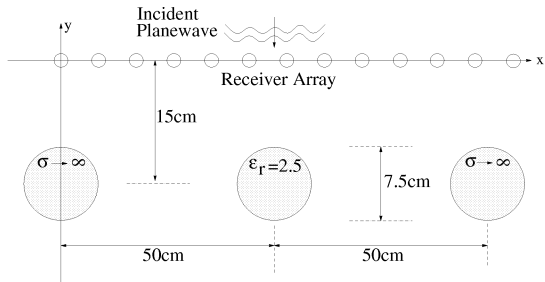


(a) GPR-type mine detection geometry

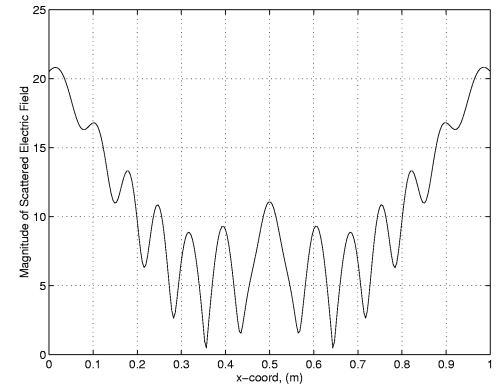


(b) Scattered field observed along the receivers for geometry of (a)

Figure 8: Implementation of T-matrix method to calculate the scattered field from 3 dielectric mine-like objects

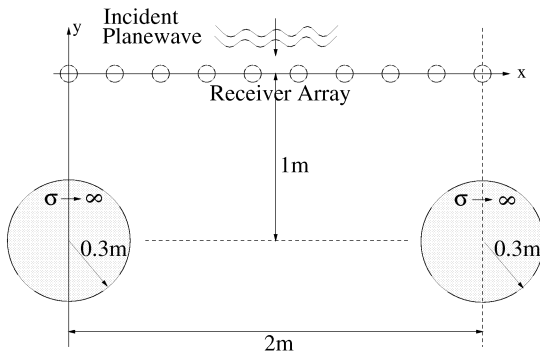


(a) GPR-type mine detection geometry

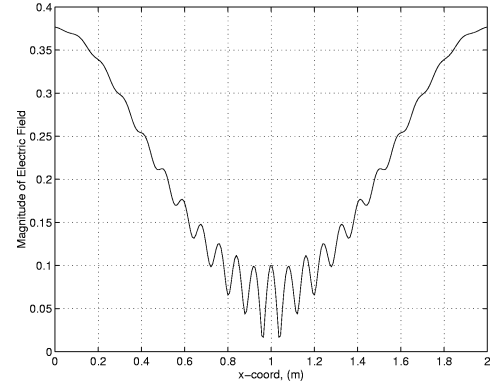


(b) Scattered field observed along the receivers for geometry of (a)

Figure 9: Implementation of T-matrix method to calculate the scattered field from 2 metallic and a dielectric mine-like objects



(a) GPR-type drum detection and localization geometry



(b) Scattered field observed along the receivers for geometry of (a)

Figure 10: Implementation of T-matrix method to calculate the scattered field from drum-like objects

Response of Realistic Soil for GPR Applications with Two Dimensional FDTD

Bo Yang and Carey Rappaport
Center for Subsurface Sensing and Imaging Systems
235 Forsyth Building
Northeastern University
Boston, MA 02115
Tel: 617 373-2043, Fax: 617 237-8627
rappaport@neu.edu

Abstract -- The three-dimensional, wideband, bistatic ground penetrating radar (GPR) scatter response of rough, realistic ground is efficiently and accurately simulated using a hybrid high resolution 3-D and large area 2-D FDTD model. The 3-D computation carefully models the transmitting and receiving antennas, while the 2-D FDTD models wave propagation between the antennas and the scattering by the soil below them. The FDTD soil model considers realistic frequency dependent (dispersive) soil with Gaussian height variations. The modeling results are compared to experiments performed with the Geo-Centers, Inc. commercially available GPR system used for mine detection. Despite the simplicity of the 2-D model, the results of the simulation and the experiment agree quite well.

Introduction

Ground penetrating radar is widely used to detect buried objects including land mines and buried waste containers, in excavation planning, and in roadway and bridge structural quality assessment [1]. Computationally modeling the performance of GPR is important in determining its feasibility in various situations and for designing the best sensing system for the given application. Appropriate computational models, such as the finite difference time domain (FDTD) method [2, 3], can be used for transmitter/receiver design and in parameter studies that include soil moisture and density variations [4]. The FDTD method is particularly well suited both for modeling soil surface roughness and volume inhomogeneities and the complex structures of wideband GPR antennas. Computational modeling is an essential complement to laboratory and field GPR experiments, which are particularly hard because of the difficulty in burying objects so they appear to be in long undisturbed ground.

The most accurate models for GPR are three dimensional, but 3-D FDTD modeling is computationally complex and expensive. Two-dimensional FDTD models can simulate much of the characteristic performance of GPR with orders of magnitude less computational expense. To model the entire link between the signal fed to the transmitting antenna and the signal measured at the output port of the receiving antenna, a reasonable compromise is to combine 3-D and 2-D models.

The present approach models the complex antenna volume with 3-D FDTD with a mesh fine enough to capture the variations in material geometry, and combines it with lower resolution free-space and soil scattering models using multiple 2-D FDTD meshes. As long as the small-scale wave variations (such as in the vicinity of the antenna or a localized three-dimensional scatterer) are partitioned in the 3-D model, the remaining piecewise-plane-wave-like propagation can be modelled with 2-D computation without introducing too much inaccuracy. Three-dimensional reflection from, and transmission through planar boundaries are also adequately approximated with 2-D FDTD, as long as the boundary is perpendicular to the computation plane. In addition, the distortion effects of random rough ground can be included in this 2-D modeling scheme. Although rough ground has randomness in two surface dimensions, as long as the surface height variation is small compared to the excitation wavelength, the depolarization and out-of-plane scattering effects are small. Fast computation is particularly important when the average scattering response of many surface realizations of random rough ground must be calculated.

In a subsequent study, the scattering of buried targets will be reported. In this present report, only the rough ground surface scattering with no target in the soil will be discussed. Indeed, to understand the scattering mechanisms of buried targets, it is first necessary to test the transmission through the ground surface and quantify scattering from the ground surface, which is the greatest source of clutter. We model the scatter response of realistic soil using the geometry and parameters employed by the Geo-Centers, Inc. Energy Focused Ground Penetrating Radar (EFGPR).

Geo-Centers EFGPR

The EFGPR is a time domain focused multistatic linear array radar used for buried mine detection. Each element of the array is a wideband transverse electromagnetic rhombus (TEMR) antenna, which resembles a bow-tie antenna that is rolled back and joined to itself using a pair of resistors. The measured performance of the TEMR indicates a very uniform radiation pattern in both planes throughout a broad frequency range [5]. This TEMR antenna is used for both transmitting and receiving elements. A Picosecond Labs, Inc. pulser feeds the elements with a single nanosecond pulse with the frequency response between 700 MHz and 1.3 GHz.

In order to enhance the performance of TEMR antennas in detecting buried mines, the transmitting and receiving antennas elements in the array (shown in Fig. 1) operate simultaneously in a fully multistatic radar mode. The system can also be used in studying the scattering response of the ground surface in order to understand the scattering mechanism of buried mines.

As shown in Fig. 1, the distance between the centers of adjacent transmitters, and between adjacent receivers, is about 0.244m, as is the distance between centers of transmitter T_i and corresponding receiver R_i ($i = 1,2,3,4$). The transmitter and receiver linear arrays lie about 0.39m above the ground surface. The transmitters and receivers are arranged in the x - z plane, with z -polarization so that waves incident in planes parallel to

the y - z plane are transverse magnetic (TM), while those incident in the x - y plane are transverse electric waves (TE).

The multistatic nature of the EFGPR array configuration allows for focusing on sample volumes of the subsurface and enhances the target signal to noise ratio. Examining a target from multiple views not only increases the response of the desired signal, but also reduces ground surface clutter, since different patches of ground are illuminated by the different sources and viewed by different receivers, and these variations tend to cancel. To characterize the ground clutter response for this array configuration, each possible transmitter element/receiver element combination must be analyzed separately. This is adequately modeled by considering the received signals at the four receivers R_1 , R_2 , R_3 and R_4 due to just transmitter T_1 . One aspect of these various links is that although the transmitting and receiving antenna elements are both z -polarized, the wave interactions with the ground have both TE and TM components.

FDTD Simulation

The link between the pulser input signal and received output voltage involves three modeling steps: one for the transmitting TEMR antenna, one for the receiving TEMR antenna, and one for the propagation through air and scattering from ground. The TEMR antennas have been successfully modelled with fine resolution ($\Delta = 0.00152\text{m}$) in three dimensions in both transmit and receive modes, as reported in Sahin [6]. The measured

transmitter excitation pulse is shown in Fig. 2. The waves radiated from the downward pointing transmitting antenna are first computed for all electric and magnetic field components throughout the 161 X 161 X 108 point grid, and the entire time record of electric field on the six bounding faces of the surrounding rectangular box are stored. Next, these calculated field values are coupled to a coarser 2-D grid by down-sampling by a factor of eight at grid points along the line segments where the transmission (2-D incidence) plane intersects the faces of the bounding box.

In addition, to avoid corner issues, the fields on the vertical (x - y and y - z) sides of the bounding box are projected onto the line crossing the bottom face (referred to as the aperture line), as shown in Figure 3. This aperture line will serve as one of the edges of the 2-D computational grid along which the excitation will be specified. It is tilted at an angle θ relative to the z -axis to account for the varying angle between transmitter and receiver. Since most of the transmitted power from the TEMR antenna propagates through the bottom face of the bounding box, the projected fields propagating through the side faces primarily ensures a smooth fall-off of power past the edge of the bottom box face. The projection of the fields from the side faces is accomplished by: 1) determining the approximate antenna phase center, 2) tracing rays from this center to the aperture line, 3) determining the attenuation spreading factor and time delay based on the length of this ray, 4) finding the electric field values at the point where the ray intersects the side face, and 5) converting these values to appropriate values on the aperture line.

The first step of finding the phase center uses the time domain fields on the bottom face of the bounding box. It is assumed that all outward propagating waves follow rays originating at the same point at the same time, as in the case of a pyramidal horn antenna [7]. By symmetry, the phase center lies along the centerline of the bounding box perpendicular to—and at a height y_c above—the aperture line. Two observation points along the aperture line are selected: one in the center, and one near the edge of the bounding box. The peak of the time pulse at each of these points is found, and its time and amplitude are recorded. Next, the distance y_c along the centerline is found by equating the path length difference between each of the two observation points and the phase center ($r_1 = y_c$ and $r_2 = \sqrt{y_c^2 + 11.6^2}$) to the time difference between the observed pulse peaks multiplied by the speed of light. The observed 210 ps time difference between arrival times of the peaks leads to $y_c = 7.5$ cm, or about 49Δ .

The second and third steps follow from similar triangles. Given the distance along the aperture line from the box centerline to the edge of the bottom face d , a point on the box side face y' projects out a distance $z' = d y_c / (y_c - y')$ along the aperture line (when $f = 0$). The distance from the phase center to this projected point is $r_p = \sqrt{y_c^2 + (d + z')^2}$, while the distance to the point on the ray at the box face is $r_f = \sqrt{(y_c - y')^2 + d^2}$. The relative time delay for the projected point relative to the point on the ray at the box face is $(r_p - r_f)/c$, and the relative attenuation is approximately r_f/r_p . A similar analysis holds for $f \neq 0$.

The fourth and fifth steps involve resolving the electric field components on the box edge face $\bar{E}(y', z = d)$ into tangential and normal components to the aperture line at the point where it intersects the projecting ray $\bar{E}_a = \hat{z}\bar{E} \bullet \hat{z} + \hat{y}\bar{E} \bullet \hat{y}$

The resulting fields \bar{E}_a are finally down-sampled in time by about a factor of seven, to yield a Courant Number of 0.5. These transverse and normal electric field components are then passed in an array of 247 time steps by 80 positions, as input to the 2-D FDTD soil scattering model along the top boundary. Fig. 4 shows one half of this array of E_{az} field values for the TE excitation. A similar array would be used for the other 2-D incidence planes.

One difficulty with dividing the model into separately calculated pieces is that the reverse coupling of waves scattered from the ground back into the transmitting antenna is ignored. This is only important from the receiver signal point of view if this backscattered wave then rescatters from the transmitting antenna. However, this effect is small, since all elements in the physical array are isolated with absorber to minimize both this rescattering and the direct-coupled signal from transmitter to receiver.

To use the 2-D FDTD to model the scattering response of the ground surface, we transform coordinates such that the incident plane includes the T_l/R_i ($i = 1,2,3,4$) antenna pair and is perpendicular to the ground. Designating the incidence plane as the X_i-Y_i plane as shown in Fig. 5, the 2-D FDTD now approximates the transmitter and receiver as infinitely long in the Z_i -direction. Since the ground surface is infinite and

perpendicular to the incidence plane, it is well modelled with 2-D FDTD. For the angle between the x -axis and X_i axis given by β_i , the electric field components in xyz coordinates, the correspond to field components in the $X_iY_iZ_i$ frame as:

$$E_{Xi} = E_x \cos \mathbf{b}_i + E_z \sin \mathbf{b}_i, \quad (1)$$

$$E_{Zi} = E_x \sin \mathbf{b}_i - E_z \cos \mathbf{b}_i, \quad (2)$$

$$E_{Yi} = -E_y. \quad (3)$$

Since the incident electric field is neither perpendicular nor parallel polarized to the incidence plane (X_i - Y_i plane) for some transmitter/receiver pairs, it is impossible to use the 2-D FDTD TE or TM codes directly. Instead, the perpendicular (E_{Xi} , E_{Yi}) and parallel (E_{Zi}) polarization components are first calculated by vector transform, then a TE FDTD code is used for the parallel polarized field while a TM code is used for the perpendicular field. The TE and TM components of the receiver field can be transformed similarly.

The space grid chosen for the 2-D FDTD air/ground interaction code is a 361×200 with space increment $\Delta = 0.0122$ m. The time increment $\Delta t = 20.5 \times 10^{-12}$ s, was chosen to maintain the Courant stability condition $c\Delta t / \Delta < 1/\sqrt{2}$. The distance between the centers of transmitter and receiver is 20 grid points (0.244m) and the bottom plane of the transmitter/receiver pair is 32 grid points, 0.39 meter above the ground surface. The soil is modelled as Puerto Rican clay loam with 5% moisture and 1.4 g/cm^3 density with dielectric constant and conductivity varying across the frequency band from 4.5 to 4.35 and 0.012 to 0.045 S/m respectively [8]. We model the frequency dependence of the soil

medium using the Z-transform representation of conductivity with constant dielectric constant [9], where $Z^{-1} = \exp(-j\omega \mathbf{D}t)$ is the Z-transform variable corresponding to a unit delay in the time domain:

$$\mathbf{s}(Z) = \frac{b_0 + b_1 \cdot Z^{-1} + b_2 \cdot Z^{-2}}{1 + a_0 \cdot Z^{-1}} \quad (4)$$

The parameters are chosen to match measured values of soil over the 30 MHz to 4 GHz frequency range: $b_0 = 0.5843$, $b_1 = -1.0649$, $b_2 = 0.4811$, $a_0 = 0.88$, and the average value of dielectric constant is $\epsilon_{Av} = 3.136$. With this conductivity model, a finite time difference equation relating conduction current to E-field is straightforward. The Perfect Matched Layer (PML) absorbing boundary condition for both free space and dispersive soil is used in the model [10]. Both flat and rough ground surfaces are used in our simulation. The roughness of the measured ground surface is accurately modeled with a Gaussian height distribution with zero mean and standard deviation σ_h , given by the probability density function relating two heights z_1 and z_2 separated by a distance x_d [11]:

$$p(z_1, z_2) = \frac{1}{2\sigma_h^2 \sqrt{1 - R^2(x_d)}} \exp\left(-\frac{z_1^2 - 2R(x_d)z_1z_2 + z_2^2}{2\sigma_h^2(1 - R^2(x_d))}\right) \quad (5a)$$

where the auto-correlation function of lateral distance is given by:

$$R(x_d) = \exp\left(-\frac{x_d^2}{l_c^2}\right) \quad (5b)$$

The standard deviation of surface height σ_h is about 1.8 cm and correlation length l_c is about 18 cm. Although with this height variation 95% of the excursion from flat is within ± 3.6 cm (or about $\pm 3\Delta$), the FDTD calculation shows noticeable differences compared to those for flat ground.

The results of the 2-D FDTD computations for waves radiated by T_I , which impinge on each of the four receivers, are shown in Fig. 6. Only one realization of rough ground is used for the responses displayed in this figure; other realizations yielded similar responses. For R_I , only the TE calculation is used, while for the three other receiver positions, the incident wave is divided into TE and TM components, each component is propagated separately, and then the waves impinging on each receiver is recombined. Note that eliminating roughness increases wave coherence, resulting in greater scattered amplitudes, but also leads to a slight delay relative to the rough ground calculation.

Antenna Transfer Function

The signal measured at each receiver output port in the experiment is an output voltage, while the signal obtained by the 2-D FDTD simulation is the electric field incident on the receiver location. Rather than perform another 3-D FDTD calculation on the receiving TEMR antenna, the receiving TEMR antenna transfer relation is used to convert the incident quasi-plane wave into the received voltage. The transfer function in the frequency domain can be obtained by dividing Discrete Fourier Transforms of the

receiver output voltage by that of its corresponding incident plane wave E-field [12]. The required receiver output voltage for the ground scattered case is thus obtained by Fourier transforming the field at the receiver, multiplying by the transfer function, and Inverse Fourier Transforming the result. In addition, before inverse transforming, we use a Wiener filter given by:

$$WF(k) = \frac{X^*(k)}{|X(k)|^2 + \epsilon} \quad (6)$$

where $X(k)$ is the DFT wave signal at the given receiver position, and $\epsilon = 10^{-5}$. The Wiener filter reduces the numerical noise in the TEMR antenna transfer function. The magnitude of the transfer function is shown in Fig. 7.

Results and Analysis

Fig. 8 shows the received voltages at the output terminals of R_1 , R_2 , R_3 , R_4 when T_1 is used as the transmitter, calculated by the hybrid 2-D/3-D FDTD model for both rough and flat ground and compares them to corresponding signals measured in the field. These signals are obtained by applying the antenna transfer function of Fig. 7 to the incident wave signals of Fig. 5. The experimental results were obtained at the Northeastern University Dedham test track. TEMR antenna elements were used as both transmitters and receivers in the configuration of Fig. 1 [13]. For each transmitter/receiver pair, the measured signals vary for different positions along track. To obtain a representative set of

measured responses, each sample signal from each position is shifted in time to align its first peak. Then the shifted sample signals for each of the four transmitter/receiver pairs are added and divided by the number of samples. These averages are taken to be the measured signals to be compared to calculated signals. This procedure takes account of the differences caused by ground surface variations along the track, but has the disadvantage that some of the small signal features are averaged-out.

To compare the amplitudes of the measured and calculated signals, each of the average measured signals are divided by the maximum average measured T1/R1 signal, and each of the calculated flat ground signals are divided by the maximum flat ground T1/R1 signal. The resulting maximum average measured signal values for T1/R2, T1/R3 and T1/R pairs are 0.9217, 0.4875 and 0.2490 respectively. The maximum calculated flat ground signal values for T1/R2, T1/R3 and T1/R pairs are 0.8145, 0.4621 and 0.2434 respectively. The close correspondence between measured and calculated maximum signal values indicate that the procedure for modeling source and receiver polarization variation is valid.

The main peaks of the FDTD calculated and measured signals have similar shape and width, and the leading behavior of each the signals is similar. The correspondence between the trailing (late time) behavior of the measured and calculated signals is worst for the closest transmitter/receiver pair (T1/R1), probably due to the multiple scattering from the ground back to the transmitter, which is ignored in the hybrid computational model. Although the largest negative excursion of the FDTD signals are slightly

different from those of the experiment, the results of the FDTD calculation are otherwise close to those of the experiment. The difference between the FDTD calculated voltage signals for rough ground and flat ground are smaller than those of the corresponding incident waves.

Considering the various approximations: 1) the actual 3-D propagation and scattering are modelled with 2-D calculations; 2) the ground surface is modelled as a rough curve in 2-D FDTD rather than a true rough surface with variations in two directions; 3) the ground scattered wave is modelled as a plane wave incident on the receiving antenna; 4) the approximations made when extrapolating the transmitted electric fields from the sides of the 3-D computational domain; 5) the possible inhomogeneity and uncertain soil type, moisture, and density inherent in the actual experiment; 6) the averaging of the experimental data for varying ground surface; and 7) the variations in the electronically generated wideband signal, electronic noise, temperature conditions, and general measurement error; the results show that the 2-D FDTD does an impressive job of modeling the electric field while maintaining the characteristics of the real electric field. Although the differences between the measured and calculated signals may be greater than the expected variations due to buried targets, the FDTD calculation captures the effects of rough dispersive ground as well as relative changes due to buried targets.

Conclusions

The approach of simulating the clutter response of realistic soil for an existing high-performance impulse GPR using relatively simple 2-D FDTD calculations has been demonstrated. The radiated fields generated using a high resolution 3-D FDTD within a small box surrounding the antenna element are sampled along the line segment of the incidence plane, decomposed into TE and TM components, and then propagated with a coarse 2-D FDTD dispersive soil model. The waves incident on the receiver are then convolved with the transfer function of the receiving antenna to give the received signal.

Computed results compared well to experimentally measured signals using the Geo-Centers TEMR antenna transmitter/receiver pairs. The simulated output voltages are close to those obtained from experiment, with similar pulse shape. Despite slight quantitative differences between the modelled and experimental data, 2-D FDTD has been shown to simply and successfully simulate scattering by rough dispersive ground.

To address scattering from finite 3-D buried targets, it is anticipated that another 3-D simulation of the sub-volume surrounding and including the target would be required. For shallow buried targets, the target volume could include the ground surface as well. Thus, the simulation could be divided into 3-D models of the transmitting antenna, the target and the receiving antenna, with 2-D simulation of the intervening space. Multiple

interactions between the antenna and the target tend to be small, especially for buried targets, so this division will not introduce much modeling error.

Acknowledgments

The authors thank Arnold Dean of Geo-Centers, Inc. for providing the experiment data and the numerical information to obtain the antenna system transfer function, and Dongping Jin for providing dispersion parameters of Puerto Rican clay loam soil. This work is supported by the Army Research Office grant No. DAAG55-97-0013.

References

1. L. Peters and J. Young, "Applications of subsurface transient radars," *Time Domain Measurements in Electromagnetics*, Miller, E. Ed., New York: Van Nostrand Reinhold, 1986.
2. K. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Transaction on Antennas and Propagation* vol. 14, no. 3, 1966, pp. 302-307.
3. A. Taflove and M. Brodwin, "Numerical solution of steady state electromagnetic scattering problems using the time-dependent Maxwell's equations," *IEEE Transactions on Microwave Theory Technology*, vol. 23, August 1975, pp. 623-630.
4. W. Weedon and C. Rappaport, "A general method for FDTD modeling of wave propagation in arbitrary frequency dispersive media", *IEEE Trans. Ant. Prop.*, vol. 45, no. 3, March 1997, pp. 401-410.
5. C. Rappaport and D. Reidy, "Focused array radar for real time imaging and detection," *Proceedings of SPIE*, vol. 2747, Orlando, FL, April 1996, pp. 202-213.
6. A. Sahin, C. Rappaport and A. Dean, Jr., "Design Considerations for Short-Time Pulse TEMR Antennas Using Finite Difference Time Domain Algorithm," *SPIE Aerosense Conference*, Orlando, FL, April 1998, pp. 784-793.
7. Collin, R., *Antennas and Radiowave Propagation*, McGraw Hill, (New York), 1985.
8. J. Hipp, "Soil Electromagnetic Parameters as Functions of Frequency, Soil Density, and Soil Moisture," *Proceedings of the IEEE*, vol. 62, January 1974, pp. 98-103.
9. C. Rappaport, Shuang Wu and S. Winton, "FDTD Wave Propagation in Dispersive Soil Using a Single Pole Conductivity Model", *IEEE Transactions on Magnetics*, vol. 35, May 1999, pp. 1542—1545.
10. C. Rappaport, and S. Winton, "Using the PML ABC Air/Soil Wave Interaction Modeling in the Time and Frequency Domains", to appear in *International Journal of Subsurface Sensing and Application*, July 2000.
11. El-Shenawee, M. and Rappaport, C., "Quantifying the Effects of Different Rough Surface Statistics for Mine Detection Using the FDTD Technique," *SPIE Aerosense Conference*, Orlando, FL, April 2000, pp. 966--975.
12. J. G. Proakis and D. G. Manolakis, *Digital Signal Processing Principles, Algorithms, and Application*, third edition, 1996.
13. Arnold Dean, Geo-Centers/Northeastern University MURI Tech. Report, Aug. 1999.

List of Figures

Figure 1: Geo-Centers' EFGPR four transmitter/four receiver array mine detection system.

Figure 2: Measured excitation voltage fed to the transmitting antenna.

Figure 3: Geometric relationship between 3-D computational volume and 2-D computational plane, with plane orientation chosen to include both transmitter and receiver (not shown) volumes. Projection lines from the transmitting antenna phase center map the electric field on the side faces onto the aperture line.

Figure 4: Electric field generated by the 3-D fine mesh FDTD antenna model, down-sampled for use as the input excitation for the 2-D TE FDTD air/soil propagation model. For clarity, only the positive half of the excitation is shown.

Figure 5: Coordinate transformation for each T_l - R_i pair.

Figure 6: Electric fields at the four receiver positions for perfect flat ground and slightly rough ground.

Figure 7: Magnitude response of the TEMR antenna transfer function.

Figure 8: Comparison of the receiver output voltage simulation results with those of experiments for the four receivers.

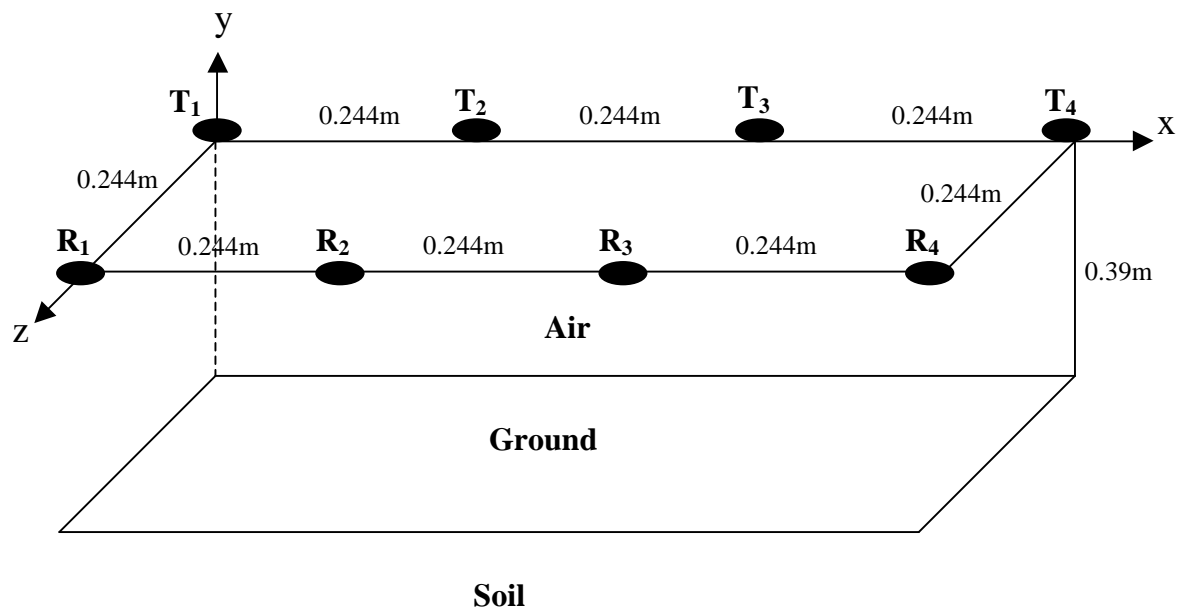


Figure 1: Geo-Centers' EFGPR four transmitter/four receiver array mine detection system.

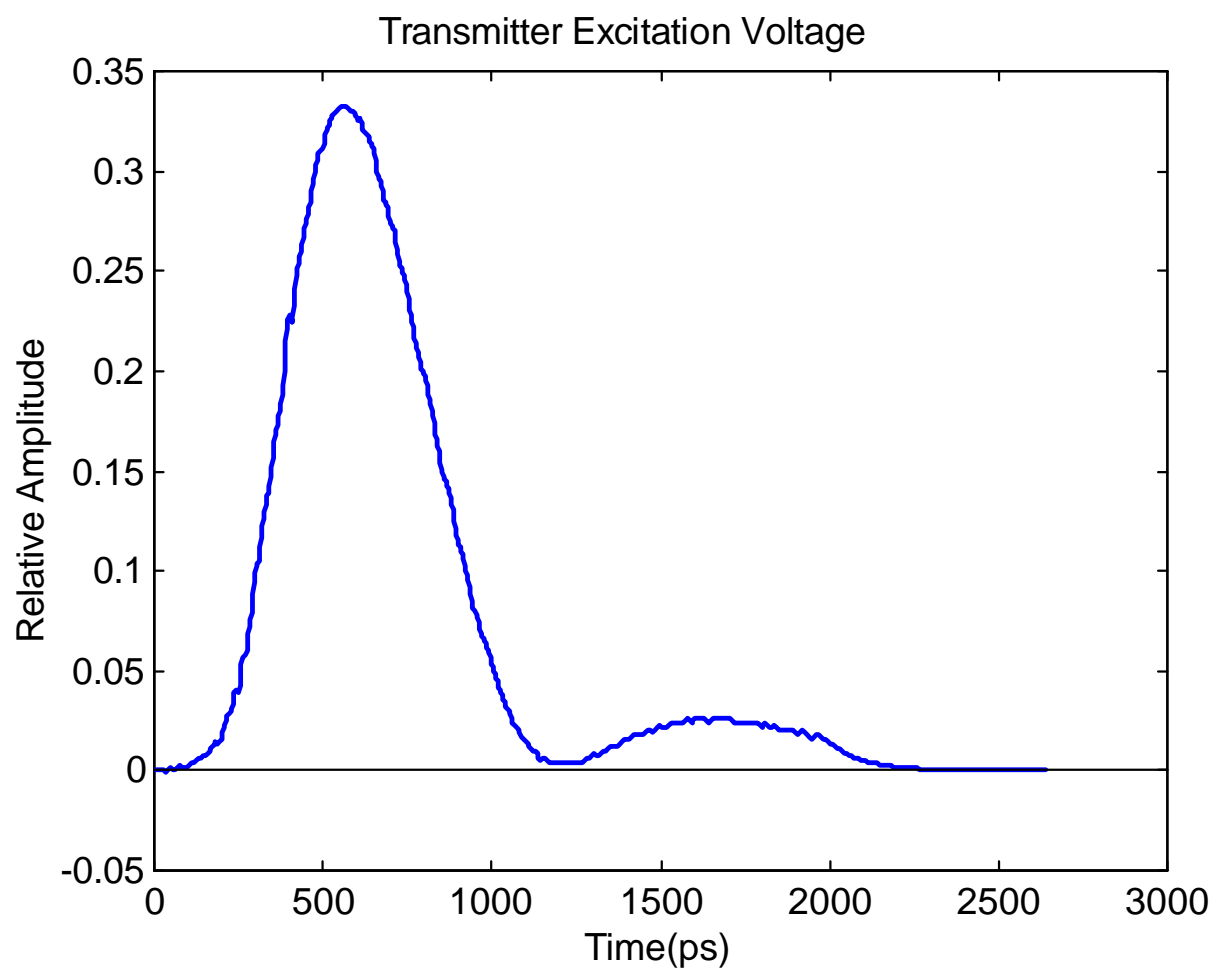
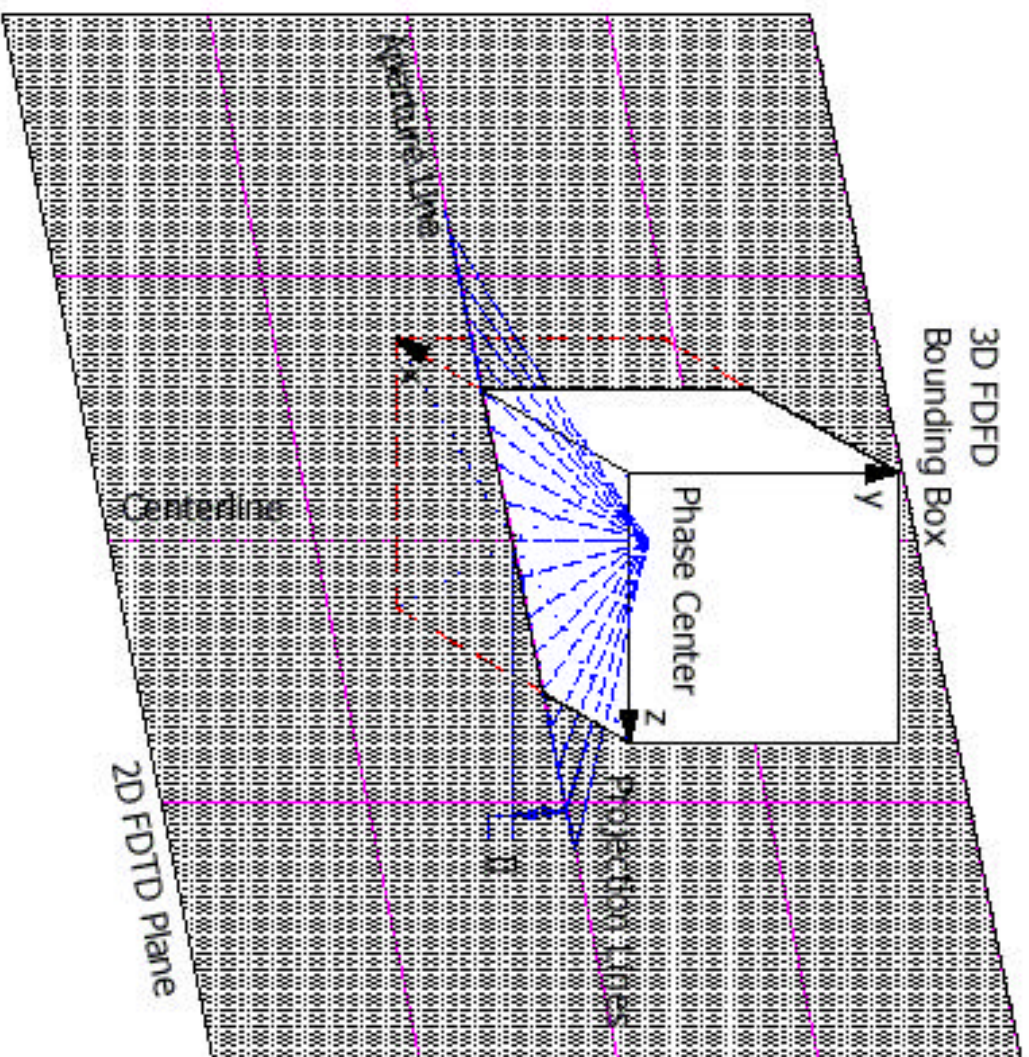


Figure 2: Measured excitation voltage fed to the transmitting antenna.

3D FDFD
Bounding Box

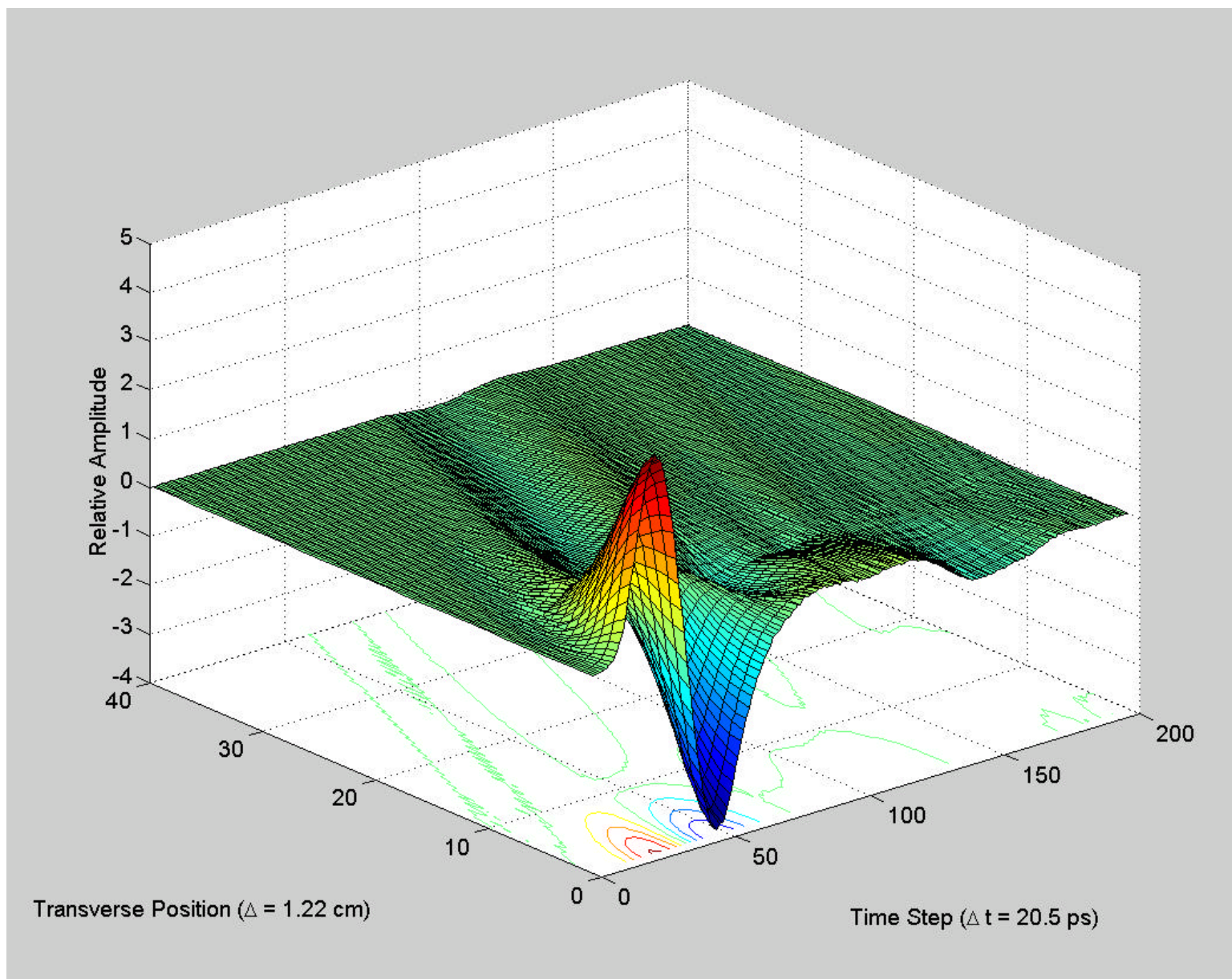


Figure 4: Electric field generated by the 3-D fine mesh FDTD antenna model, down-sampled for use as the input excitation for the 2-D TE FDTD air/soil propagation model. For clarity, only the positive half of the excitation is shown.

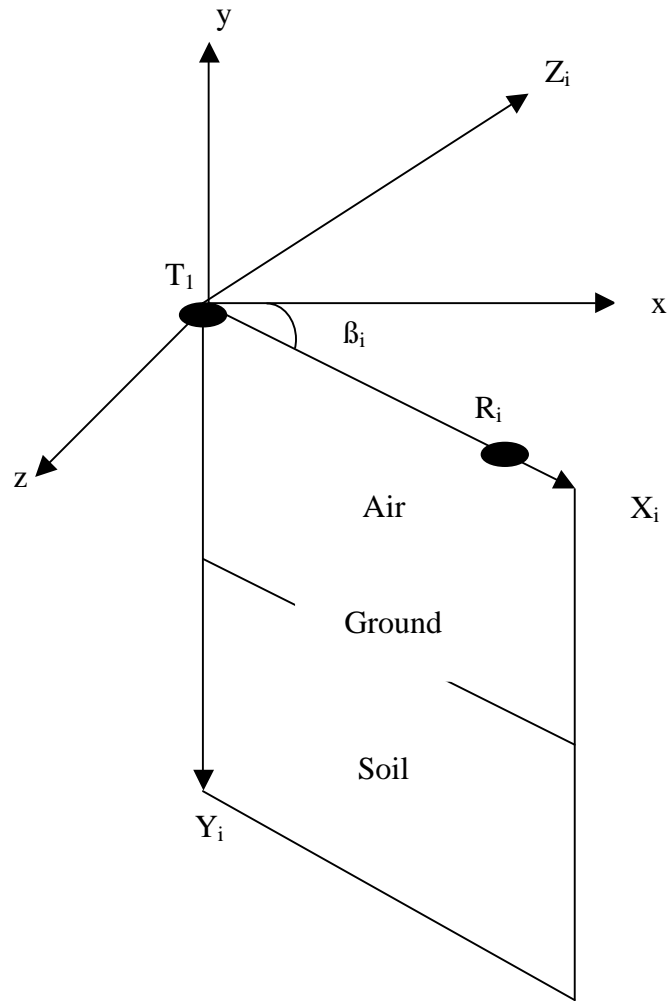


Figure 5: Coordinate transformation for each T_l - R_i pair.

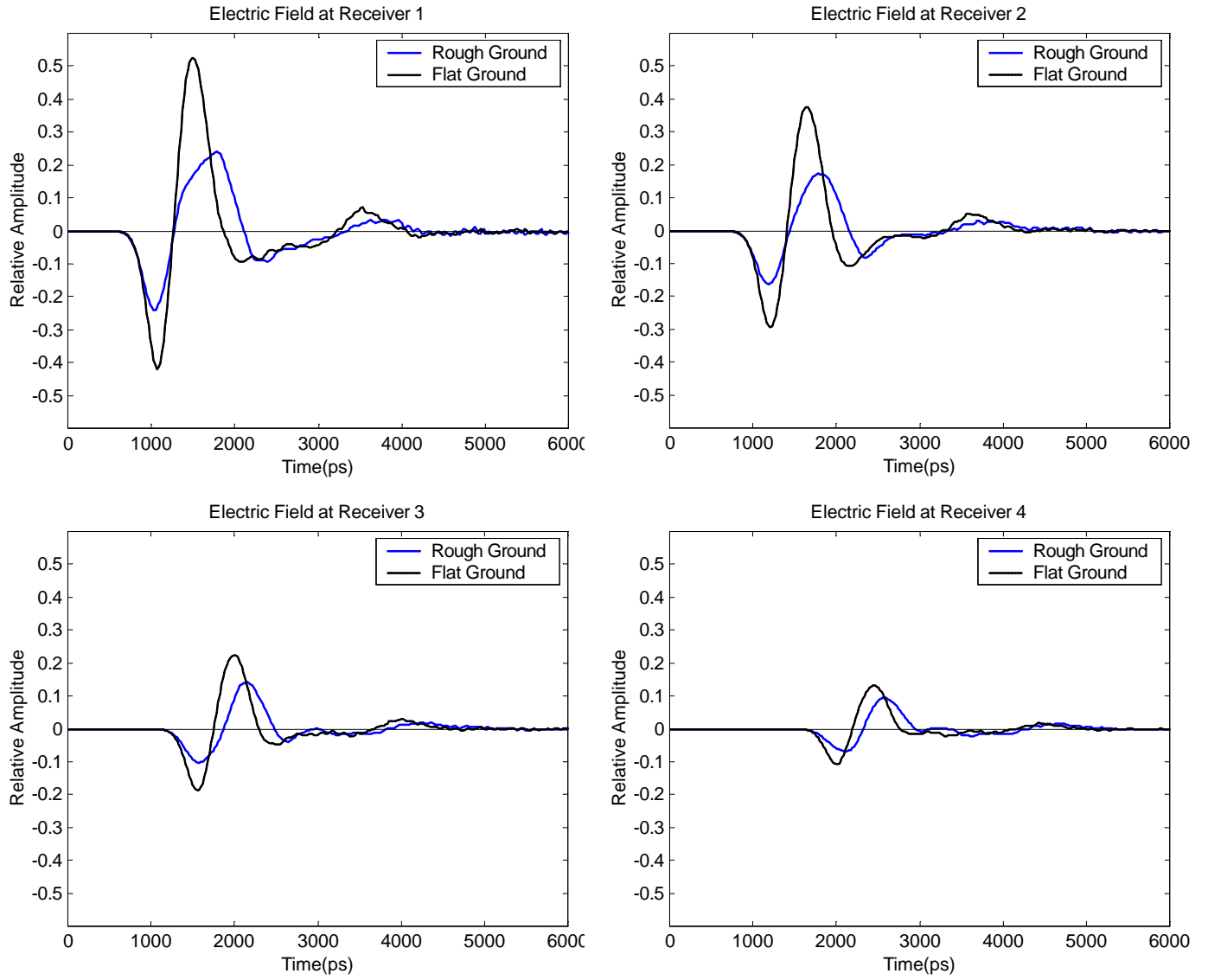


Figure 6: Electric fields at the four receiver positions for perfect flat ground and slightly rough ground.

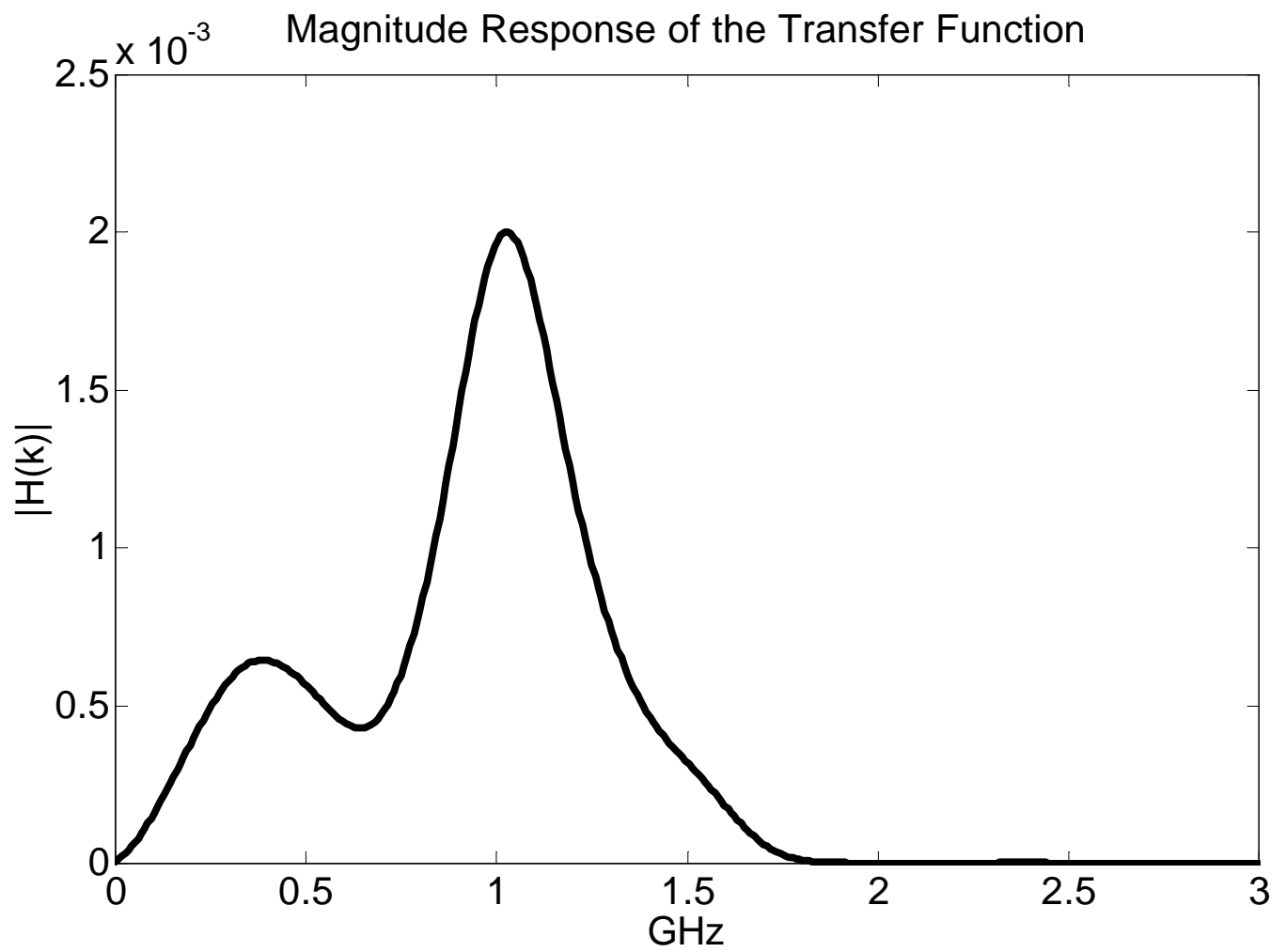


Figure 7: Magnitude response of the TEMR antenna transfer function.

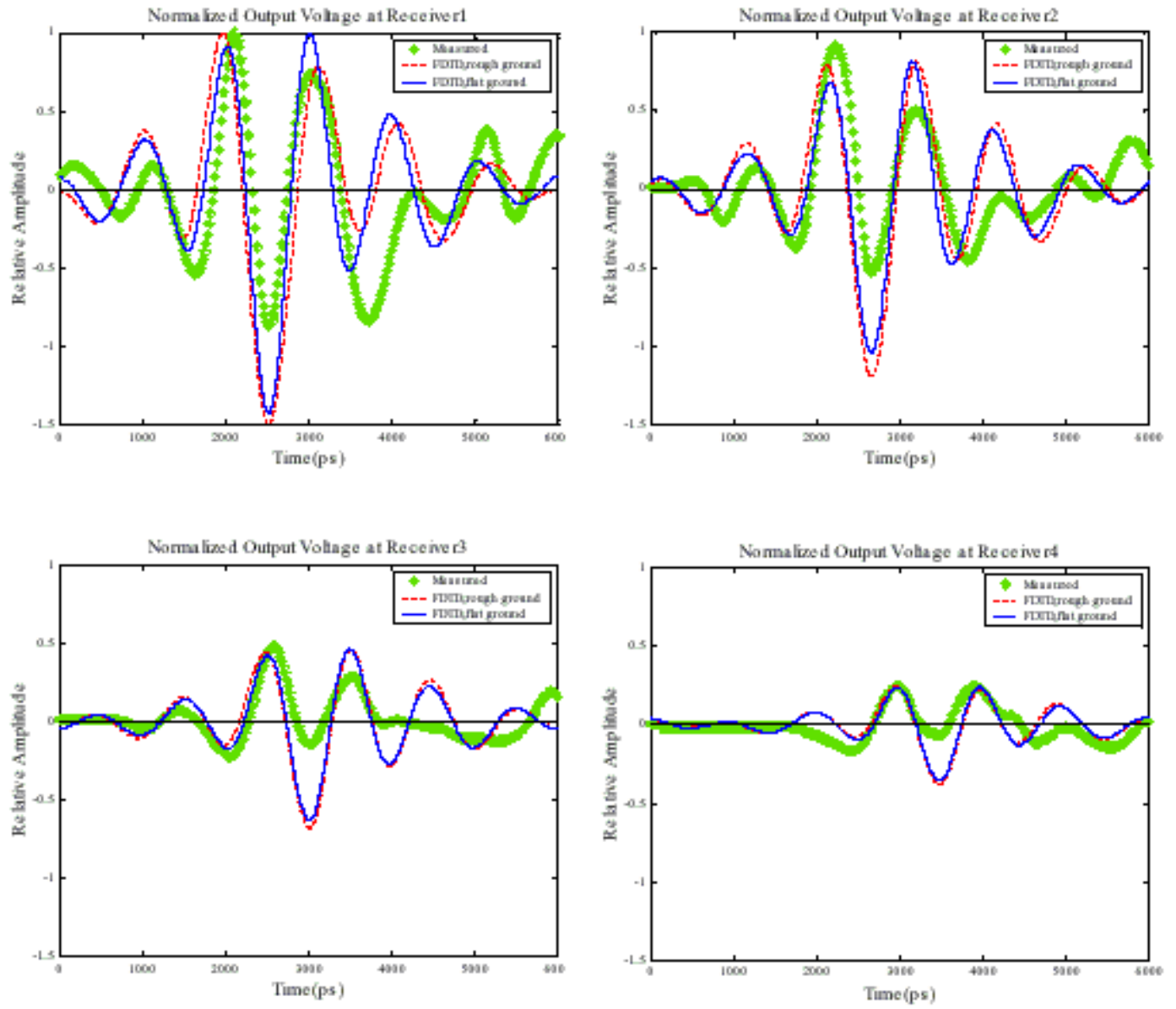


Figure 8: Comparison of the receiver output voltage simulation results with those of experiments for the four receivers

Object Detection Using High-Resolution Near-Field Array Processing*

Adnan Şahin and Eric L. Miller
Center for Electromagnetics Research,
235 Forsyth Building, 360 Huntington Ave.
Northeastern University, Boston, MA 02115
Telephone: (617) 373-8386
Telefax : (617) 373-8627
email: adnan@cdsp.neu.edu

April 24, 1998

Abstract

In this paper we present an algorithm for the detection and localization of an unknown number of objects present in the near field of a linear receiver array. To overcome the nonplanar nature of the wavefield over the array, the full array is divided into a collection of sub-arrays, such that the scattered fields from objects are locally planar at each sub-array. Using the MUSIC algorithm, directions of arrival (DOA) of locally planar waves at each sub-array are found. By triangulating these DOAs, a set of crossings, condensed around expected object locations, are obtained. To process this spatial crossing pattern, we develop a statistical model for the distribution of these crossings and employ hypothesis testing techniques to identify a collection of small windows likely to contain targets. Finally, the results of the hypothesis tests are used to estimate the number and locations of the targets. Using simulated data, we demonstrate usefulness and performance of this approach for typical background electrical properties and signal to noise ratios.

*This work was supported by the Department of Energy under contract DE-FC07-95ID13395, the Army Research Office Demining MURI under Grant DAAG55-97-1-0013, an ODDR&E MURI under Air Force Office of Scientific Research contract F49620-96-1-0028, and a CAREER Award from the National Science Foundation MIP-9623721

1 Introduction

The problem of detection and localization of objects in the near field of an antenna array has arisen in a number of application areas in recent years. For example, in the area of landmine remediation, the goal is to find relatively small, metallic and plastic objects located in a lossy medium (the soil) but a few centimeters from the transmitters and receivers. Alternatively, for purposes of environmental remediation, the targets of interest tend to be larger (eg. steel metal drums filled with hazardous waste), and located on the order of meters from the array. Although several near-field array processing methods have been reported in the literature, their use has been restricted to the localization of independent sources radiating spherical waves [1–3] and are thus not suited to the problem of interest in this work: the detection and localization of extended targets illuminated by an incident plane wave and positioned such that multiple scattering effects cannot be ignored.

The problem of target detection and localization for these and related applications often is addressed by using the data to produce a pixel-by-pixel map of the region near the array and then post-processing the image to localize the objects [4–8]. Since the initial image generation represents an ill-posed inverse problem, a stable solution requires the use of a regularization method [9]. Unfortunately, typical regularizers result in smooth images thereby making the detection all the more difficult. In this paper, we develop an alternate approach to target detection and localization which bypasses this difficult step of image generation and is aimed at extracting the number of objects and their locations more directly from the data.

Of particular interest is the measurement geometry shown in Fig. 1. A plane wave illuminates the region of interest assumed to be a homogeneous, possibly lossy medium containing one or more targets located in the near field of an array of receivers. The inherent array structure of

the measurement geometry suggests that the high resolution array processing techniques [10, 11] quite popular in the signal processing community would be well suited for the near-field detection problem. Adapting such methods to the problem of interest here presents a collection of interesting challenges. First, these array processing techniques typically assume that the sources are infinitely far away so that the waveform received on the array is planar. For our problem, since the objects are located relatively close to the receiver array, this key assumption is not valid. Second, for such near-field objects both range as well as the direction of arrival has to be determined in order to localize the object. Finally, a problem common to both the near-field and far-field array processing algorithms is that the number of incident waveforms/targets is not known *a priori*.

To deal with the nonplanar nature of the wavefronts over the array, we partition the receiver array into sub-arrays, such that the scattered field is locally planar at each sub-array. Then, using high resolution array processing techniques, each sub-array identifies a single direction of arrival (DOA) corresponding to the most dominant scatterer in the vicinity of that sub-array. The localization of the objects in terms of their ranges and bearings is achieved by triangulating the directions of arrival from all subarrays which in turn results in a crossing pattern of DOA intersections. Examination of typical crossing structures reveals that there are two distinct patterns where the crossings are either dense or sparse. Dense crossing regions clearly indicate object locations and are distinguished from “background” regions where the crossings are sparse. The problem of object detection and localization then is reduced to the processing of the crossings obtained from our triangulation procedure.

For this purpose, we introduce a simple yet accurate stochastic model describing the spatial distribution of DOA crossings. Such modeling is warranted for two reasons. First, due to the noise in the data, the DOA intersection points are inherently randomly distributed in the plane. Second,

such modeling forms a solid basis for algorithm development and quantitative performance analysis in the form of detection and false alarm rates.

In this work, we model the two classes of crossings (dense vs. sparse) using a pair of spatial Poisson distributions [12]. The Poisson model in the target region has a large rate parameter while that of the background region is considerably smaller. Based on these target and background models, we develop a hypothesis testing technique for the joint estimation of the rate parameters and the localization of dense crossing regions which indicate the existence of targets. Simple post-processing of the hypothesis testing results provides both the number of targets and estimates of their locations. Finally, we verify that the a Poisson model is in fact a rather accurate description of the spatial distribution of crossings.

We demonstrate the performance of this approach for the detection and localization of multiple mine-like and drum-like targets located in the near field of the receiver array. For mine-like targets relative positions of the objects are changed to see the effect of object geometry on detectability. We show that the detectability improves, and false alarm rate decreases as the objects are located far apart. For drum-like targets, we demonstrate the effect of relative depth as well as relative distance between objects on detectability.

The remainder of this paper is organized as follows. In Section 2 we describe the models and notation used in the paper, in Section 3 we introduce the detection algorithm and hypothesis testing. Examples depicting performance of the algorithm are given in Section 4 and in Section 5 we will draw conclusions and suggest future work.

2 Background

The multi, bi-static measurement scheme depicted in Fig. 1 is considered in this paper. A perpendicularly polarized plane wave, $E_i(\mathbf{r})$, impinges on a collection of objects in a known background, inducing surface and volume currents which in turn radiate a scattered field, $E_s(\mathbf{r})$ ¹. The scattered electric field from the targets is spatially sampled by a uniformly spaced, N -element linear array with isotropic receiver characteristics. The measured data at the sensor outputs are:

$$\mathbf{x} = \mathbf{E}_s + \mathbf{n}, \quad (1)$$

where $\mathbf{E}_s = [E_s(\mathbf{r}_1) \ E_s(\mathbf{r}_2) \ \cdots \ E_s(\mathbf{r}_N)]^T$, \mathbf{r}_i is the vector from the origin to the i th receiver location and \mathbf{n} is zero mean, white Gaussian noise.

The structure of the receivers in Fig. 1 coupled with the underlying problem of target detection suggests the use of array processing methods for localizing buried targets. In this paper we consider the MUSIC (Multiple Signal Classification) [13] algorithm. Traditionally, MUSIC and other direction finding techniques are used to determine directions of arrival (DOA) of plane waves to a receiver array. Here, we adapt MUSIC to the near-field detection problem.

To make use of the MUSIC algorithm, the experiment as represented by (1) is repeated many times to determine the statistics of \mathbf{x} . In particular, if L scattering experiments are performed, then the maximum likelihood estimate of the spatial autocovariance matrix \mathbf{R} is given by [10]:

$$\hat{\mathbf{R}} = \frac{1}{L} \sum_{l=1}^L \mathbf{x}_l \mathbf{x}_l^H \quad (2)$$

where \mathbf{x}_l is the data measured at the l th experiment, and superscript H denotes conjugate transpose.

Then, the eigenspace decomposition of $\hat{\mathbf{R}}$ yields [10]:

$$\hat{\mathbf{R}} = \hat{\mathbf{U}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{U}}_s^H + \hat{\mathbf{U}}_n [\hat{\sigma}^2 \mathbf{I}] \hat{\mathbf{U}}_n^H \quad (3)$$

where $\hat{\mathbf{U}}_s$ is the estimated signal subspace matrix and contains the M signal eigenvectors, and $\hat{\mathbf{U}}_n$

¹All analysis is in frequency domain, thus the $e^{j\omega t}$ dependence is suppressed

is the estimated noise subspace matrix and contains $N - M$ noise eigenvectors of multiple noise eigenvalue $\hat{\sigma}^2$. The projection operator onto the noise subspace is defined as [10]:

$$\hat{\Pi}_{\mathbf{n}} = \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H. \quad (4)$$

Assuming plane wave incidence on the array, the idea behind MUSIC is that the reciprocal of the “distance” between the estimated noise subspace and the true noise subspace has sharp peaks around the DOAs. Thus, if one plots this quantity versus all possible angles, estimates of DOAs can be determined by the maxima of the resulting angular spectrum which is given by [10]:

$$P_{MUSIC}(\theta) = \frac{\mathbf{a}(\theta)^H \mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \hat{\Pi}_{\mathbf{n}} \mathbf{a}(\theta)} \quad (5)$$

where $\mathbf{a}(\theta) = [1 \quad e^{j\beta d \cos \theta} \quad e^{j2\beta d \cos \theta} \quad \dots \quad e^{j(N-1)\beta d \cos \theta}]^T$ is the direction vector, β is the wave number in the medium of propagation, and d is the distance between two receivers.

As stated previously this formulation of the array processing problem assumes that the radiator is infinitely distant so that the scattered field has planar wavefronts and the elements of the direction vector $\mathbf{a}(\theta)$ are complex exponentials. However for the problems of interest here, the receivers are in the near-field region of the radiating sources, resulting in non-planar wavefronts. Additionally, the target localization problem not only requires the DOA relative to the array but also the range of the target from a point on the array (eg. the leftmost element.)

3 Algorithm

A key element of the work in this paper is the development of a sub-array processing method for detection of multiple objects in the near field of an array. In a previous work [14], we have examined such a technique for detection and localization of single metallic and dielectric objects. As illustrated in Fig. 2 (and as is generally the case), the localization problem in [14] is straightforward since typically all crossings are densely packed within the radius of the object. Therefore, the location of the object can be inferred quite easily. For multiple objects, however, the crossing pattern is

quite complicated, since DOAs of different objects create unwanted crossings as shown in Fig. 3. The clusters of object crossings are embedded in this unwanted background crossings, and have to be extracted carefully. Thus, in this work, we concentrate exclusively on the problem of multiple object detection.

Before introducing the details of our approach, we want to briefly describe the algorithm with the help of flow chart in Fig. 4. The algorithm proceeds as follows:

1. **Sub-array processing:** At this stage of the algorithm, we partition the receiver array so that the observed scattered field is locally planar at each sub-array. The directions of arrival (DOAs) are found using MUSIC as if planar waves are impinging on the sub-arrays. The DOAs are then triangulated to obtain the crossing pattern. This stage of the algorithm is repeated several times for plane waves at different temporal frequencies to improve performance and resolution. The crossing patterns obtained at different frequencies are overlaid to yield an aggregate crossing pattern which is, then, passed onto the second stage of the algorithm.
2. **Crossing analysis:** In the second stage, the crossing pattern is modeled with two Poisson counting processes, corresponding to target and background regions. After estimating the required rate parameters using the crossing data, a hypothesis testing procedure is employed to determine a set of “window” regions corresponding to areas containing targets.
3. **Target extraction:** At the final stage of the algorithm, the individual detection windows are aggregated into a number of spatially disjoint groups. The total number of groups indicates the estimated number of targets, and average coordinates of all windows in a group indicate the estimated center of the corresponding target.

3.1 Sub-array Processing

The direction finding algorithms traditionally assume plane wave incidences and determine the DOA associated with each plane wave. For near-field problems, however, both DOA and the range of the source (in our case scatterer) should be acquired. Here, we describe a sub-array processing (SAP) scheme which only requires one-dimensional search in DOA space of each sub-array. The idea behind the sub-array processing is that if the aperture of the sub-array is small enough, the scattered field impinging upon it can be assumed locally planar. Thus, the plane wave MUSIC can be used to find DOAs at each sub-array, and by triangulation, it is possible to localize the scatterers.

When there are $M > 1$ objects in the vicinity of the array, we have two options in terms of how MUSIC is employed:

1. Each sub-array finds M DOAs for all locally planar waves scattered from M objects, or
2. Each sub-array finds one DOA for the locally planar wave dominant in the total scattered field (Fig. 3 shows $M = 2$ case).

Given M objects and S sub-arrays, for each operating frequency the first and second options result in $0.5SM(SM - 1)$ and $0.5S(S - 1)$ crossings, respectively. The first option creates many unwanted crossings when DOAs belonging to different objects intersect. In addition, we have to know the number of objects under the array to use this option. On the other hand, the second option does not require the knowledge of number of objects, and the scattered fields from targets closer to the sub-arrays, particularly in a lossy medium such as soil, dominate the total scattered field at the sub-arrays. Therefore, the latter option seems more practical especially when one wants to avoid estimating the number of scatterers first, and is used in the remainder of this paper.

Once one DOA at each sub-array is determined, all DOAs are triangulated to estimate the target locations. Fig. 3(a) shows the triangulation of DOAs, and (b) shows the crossings. In contrast to the single object case, for multiple objects, the crossing pattern may get quite complicated since DOAs belonging to different objects also intersect each other to create unwanted crossings. Thus, a second level of processing is required to extract the clusters indicating the estimated object centers.

3.2 Crossing Analysis

In this section we present an approach that models the DOA crossings with Poisson point processes. Inspecting Fig. 3(b), we see two distinct regions where the density of the crossings are quite different: in the first region (*background region*) the crossings are sparse, and in the second region (*target region*) the crossings are dense. By exploiting this difference, it is possible to isolate target locations. Hence, we introduce a Poisson model for DOA crossings which has a large rate parameter (intensity) in target regions and a small rate parameter in the background region.

Formally, for a given crossing pattern, we count the number of crossing Y_j , $j = 1, 2, \dots, N_y$, in a window of size $w_x \times w_y$, where N_y is the total number of *non-overlapping* windows, w_x and w_y are the width of the windows in x and y directions, respectively. The windows must be non-overlapping to guarantee the independence of random variables Y_j .

In order to ensure that Y_j is Poisson distributed, we tested for fitness to Poisson distribution by using the graphical technique presented in [15]. The technique proposes that for each count k observed in Y_j , we plot k versus $(\ln k! + \ln F_k)$ where $F_k = \sum_{j=1}^{N_y} [Y_j = k]$ is the number of data values Y_j equal to k . If the fit to the Poisson model is satisfactory, then the plot should form a straight line with slope approximately $\ln \lambda$, where λ is the rate parameter of the distribution. When we apply this test to a typical crossing pattern, instead of a straight line, we observed the curve in Fig. 5. By examining this curve, we notice that it can be decomposed into two parts, each roughly

corresponding to a straight line. The first part is when the crossing count k is small (between 0 and 2), and the second part is when k is large (greater than 3.) It is clear that these two regions correspond to the background process which is expected to have a small count of crossings, and the target process which is expected to have a large count of crossings. Furthermore, using these two approximately linear regions, we can decouple background and target processes by identifying k_b and F_{kb} for the background, and k_t and F_{kt} for the target regions where $k_b = 0, 1, 2$, $k_t = 3, 4, \dots$, $F_{kb} = \sum_{j=1}^{N_y} [Y_j = k_b]$ and $F_{kt} = \sum_{j=1}^{N_y} [Y_j = k_t]$. Then, the rate parameters for the background and the target regions are given by their maximum likelihood estimators [15]:

$$\hat{\lambda}_b = \frac{1}{N_b} \sum_{k_b=0}^2 k_b F_{kb}, \quad (6)$$

and

$$\hat{\lambda}_t = \frac{1}{N_t} \sum_{k_t=3}^{\infty} k_t F_{kt}, \quad (7)$$

where $N_b = \sum_{k_b} F_{kb}$ and $N_t = \sum_{k_t} F_{kt}$. Having estimated $\hat{\lambda}_b$ and $\hat{\lambda}_t$, the probability mass functions in the background and target regions can be expressed as:

$$f_X(k|\text{Background}) = P\{X = k|\text{Background}\} = \frac{1}{k!} e^{-\hat{\lambda}_b} \hat{\lambda}_b^k \quad (8)$$

and

$$f_X(k|\text{Target}) = P\{X = k|\text{Target}\} = \frac{1}{k!} e^{-\hat{\lambda}_t} \hat{\lambda}_t^k. \quad (9)$$

To extract crossing clusters, we sweep the region of interest with a test window of size $w_x \times w_y$. It is important that the area of the test window is equal to the area of the non-overlapping windows used in estimating the rate parameters. At each location of the test window, we count the number of crossings T_j , $j = 1, 2, \dots, N_{test}$, where N_{test} is the total number of *overlapping* sweep windows in the region of interest. The number of overlapping test windows N_{test} defines the resolution of detection, and it is greater than N_y . Since we are going to test each T_j against the hypothesis one by one, the use of overlapping windows is allowed. Hypothesis test permits us determine whether

the test window is over a background region or over a target region. The hypothesis test is then formally written as:

- H_0 : T_j is Poisson distributed with a small rate parameter $\hat{\lambda}_b$,
- H_1 : T_j is Poisson distributed with a large rate parameter $\hat{\lambda}_t$.

Based on this hypothesis test, if H_0 is true, we decide that the window belongs to a background process with a small intensity. However, if H_1 is true, we declare that the window belongs to a target process with a large intensity and call it a detection.

The likelihood ratio for the hypothesis test is formed in terms of the probability mass functions of (8) and (9) as:

$$\Lambda(T_j) = \frac{f_X(T_j|H_1)}{f_X(T_j|H_0)}$$

The decision is, then, made based on the test:

$$\ln\{\Lambda(T_j)\} = T_j \stackrel{H_1}{>} K,$$

where the decision threshold, K , is found from a specified false alarm rate P_{fa} using (10). This means that all windows which has K or more crossings in them will be declared as target locations.

Probability of false alarm P_{fa} can be written in terms of the decision threshold K and probability mass function of background process in (8) as :

$$P_{fa} = \sum_{k=K}^{\infty} f_X(k|H_0). \quad (10)$$

Given the decision threshold K , the probability of detection for the Poisson model developed in this section is given by:

$$P_d = \sum_{k=K}^{\infty} f_X(k|H_1). \quad (11)$$

It might be argued that since target windows are obtained via thresholding, there would be no need for a Poisson-based model as described in this section. A plain thresholding scheme on DOA

crossings would also locate the targets successfully. However, the Poisson model provides a solid groundwork for a detailed statistical analysis. With the model, it is possible to define probabilities of false alarm and detection. Based on these statistical analyses, it is possible to make educated predictions about the performance of the system under different conditions.

3.3 Target Extraction

Hypothesis testing with the Poisson model results in detection windows as shown in Fig. 6. By looking at this figure, a human operator may conclude the target locations and their numbers. However, we want the detection algorithm to do these decisions and calculations for us automatically. In effect, we want the algorithm to yield the number of targets in the region of interest and their estimated locations, rather than the intertwined pattern of detection windows.

The pattern of detection windows suggests that the detection windows belonging to the same targets overlap. Therefore, we classify the detection windows so that all overlapping windows form a different group. The number of targets is, then, equal to the number of groups and the estimated object centers are obtained by averaging the coordinates of the windows in each group. The grouping algorithm we use, therefore, proceeds as follows. Start with the first window on the list of detection windows and place it in the first group. For each of the other windows, test if they overlap with any window in the k th group for $k = 1, 2, \dots, G_c$, where G_c is the number of currently available groups. If the window overlaps with only one group, add it to that group. If the window overlaps with more than one groups, merge those groups, and reduce the number of current groups G_c accordingly. If the window does not overlap with any windows among G_c groups, then form $(G_c + 1)$ th group with that window. When all detection windows are classified, G_c gives the number of objects, and averaged coordinates of all windows in each group give the estimates of center of the objects they represent.

3.4 Frequency Diversity

Frequency diversity is often used in detection applications for two important reasons: to increase the resolution (high frequencies) and to allow radar signals to penetrate deeper into the medium (low frequencies). Therefore, with a wide frequency range, one can ideally get more resolution in the vicinity of the radar, and more penetration to probe deeper objects. To take advantage of these benefits, we use the sub-array processing in a multi-frequency scheme. For each frequency the sub-array processing described in Section 3.1 is carried out to obtain the DOAs and the crossing pattern. Then, these multiple crossing patterns are overlaid to give an aggregate crossing pattern which is modeled as Poisson counting processes in Section 3.2.

4 Examples

In this section, we present applications of sub-array processing to the detection of multiple mine-like and multiple drum-like objects. In order to simplify the scattering phenomenon associated with the detection problem, both mine-like and drum-like objects are modeled with simple, circular objects. The system parameters for both applications are kept constant to provide a better comparison of the method between applications. In order to introduce frequency diversity, the objects are illuminated with plane waves at three different frequencies: 1.2, 1.0 and 0.8 GHz. The frequency range used is typical of that used in practical subsurface sensing systems. The scattered field is observed along a 33-element, uniform, linear receiver array which spans an aperture of 1.5 m. The sensors are assumed to be ideal, isotropic receivers, and the inter-element spacing of the receivers are chosen such that it is less than half a wavelength for the soil characteristics [16] at the frequencies used. The receiver array is divided into 11 three-element sub-arrays for the processing. The objects are placed in a lossy, homogeneous background which has the same electrical

characteristics of 5% moist San Antonio clay loam or 10% moist Puerto Rico clay loam ($\epsilon_b = 6\epsilon_0$, $\sigma_b = 5 \times 10^{-2} \text{ S/m}$) at around 1.0 GHz [16].

For the simulations, the definition of signal to noise ratio (SNR) is not obvious. In practical problems, SNR is imposed by the nature of the system noise. However, in computer simulations we want to reference the noise power to a fixed quantity that does not change as the positions of the objects change. For this purpose SNR is referenced to the scattered field strength of a single, cylindrical, metallic object placed at the same depth as the objects, in the same lossy medium. The radius of the reference scatterer is the same as the radii of the targets. With this definition, the noise power is always proportional to the power of reference scattered field, not the power of field scattered from targets which changes as the positions of objects change.

In all examples the exact scattered field due to multiple objects embedded in a homogeneous, lossy background is calculated using the recursive T-matrix algorithm [17–20] to keep the computational requirements at reasonably low levels.

4.1 Multiple Mine-like Objects

In these examples, we placed two mine-like objects, each with 7.5 cm. diameter, 15 cm. under the receiver array. Even though the algorithm is capable of detecting more than two objects, it seems that for practical purposes no more than two mines will be placed in the array’s aperture of 1.5 meters. We have not explored the performance of the processing with respect to depth assuming that mine-like objects will be placed at uniform depths under the array.

The first example demonstrates the utility of the sub-array processing in detecting and localizing both a metallic and a dielectric object in the same medium. For this purpose, a metallic object and a dielectric object with dielectric constant of 2.5 are placed at $(20, -15)$ cm and $(80, -15)$ cm, respectively, in the homogeneous background described before, Fig 6. Signal to noise ratio is fixed

at 10 decibels. Fig 6 shows the directions of arrivals, and detection windows when probability of false alarm is 10^{-8} , which corresponds to a crossing threshold of $K = 7$. The two circles in this figure denote the objects, and the intertwined squares, due to overlaps in test windows, depict the estimated target windows. The detection windows are then used in the grouping algorithm described in Section 3.3. As expected two distinct groups of overlapping windows indicated that there are two objects beneath the array. The estimated center for the objects are found to be $(19.52, -14.02)$ cm and $(80.10, -14.82)$ cm. Consequently, both metallic and dielectric objects are detected within acceptable estimation error margins.

Next, we consider an example that demonstrates the detectability of objects and the performance of SAP as relative positions of the objects change. In this example, both objects are assumed to be metallic in order to see the influence of relative distance between same type of objects on detectability and estimation error. For this purpose, we fixed the location of the first object at $(x, y) = (-40, -15)$ cm. The other one is moved from $x = 0$ cm to $x = 125$ cm in the lateral direction while its depth is kept at the same level as the fixed object, Fig. 7. The signal to noise ratio is assumed to be 30 decibels.

With this geometry, the moving object is always located under the array, and thus detected for all combinations of relative positions. Detection of the fixed object, on the other hand, is challenging since it lies outside the span of the array. Fig 8 shows the average simulated probability of detection of the fixed object over 100 Monte-Carlo simulations as the other object is moved under the array when probability of false alarm is set to 10^{-3} . As it is clear from this figure, the fixed object can be detected only if the other object is well away from it. The fixed object may also be detected with less than 10% probability when the moving object is located between $x = 0$ and $x = 20$ cm. This is due to the fact that for these relative locations, both objects are close enough so

that DOAs belonging to the moving object create crossings around the fixed object as well as the moving object. In fact, because of this effect, the region between the two objects is incorrectly detected as targets, and thus false alarm rate is large when relative distance between two objects is small, Fig. 9. Simulated $P_{fa, sim}$ as depicted in Fig. 9 approaches to the false alarm rate set at the beginning of the simulation (dashed line), as the relative distance between the objects increase.

We have also investigated the effect of relative distance on estimated object centers. For this purpose, we plotted the averaged estimation error in x -direction ($\Delta x = x_{true} - x_{estimated}$) versus the averaged estimation error in y -direction ($\Delta y = y_{true} - y_{estimated}$) for both fixed and moving objects in Fig. 10(a-b). The estimated object coordinates are obtained by averaging 100 Monte-Carlo simulations. The dashed circles indicate the boundaries of the objects. The closer the symbols (o's or \diamond 's) are to the center, the smaller is the estimation error. Since the moving object is always detected, each small circle in Fig. 10(a) corresponds to a different position of the moving object. As seen from this plot, the estimation error of the moving object is only a small fraction of the radius. Small circles outside the object boundary (dashed circle) correspond to locations where the moving object is close to the fixed object. Each small circle and diamond in Fig. 10(b) corresponds to a relative position when the fixed object is detected. The small circles denote the error in estimated centers when the moving object's x -coordinate is greater than 95 cm, and small diamonds correspond to other locations of the moving object for which the fixed object is detected. The loci of small circles and diamonds clearly imply that as the relative distance between the objects gets larger, the estimation error in the position of the fixed object gets smaller.

4.2 Multiple Drum-like Objects

In this section, two drum-like objects, each with 50 cm. diameter, are placed at various depths from the receiver array. Since drums are made up of metals, the objects are assumed to have

infinitely large conductivity. The signal to noise ratio is set at 10 decibels. The detection windows for a typical case is shown in Fig. 11. In this example one object is at $(-40, -125)$ cm and the other is at $(140, -125)$ cm, while the lossy, homogeneous background is the same as previous example. The threshold of detection corresponding to a false alarm rate of 10^{-8} is found to be $K = 6$. Fig. 11(b) depicts the detection windows obtained after hypothesis testing. These windows are then processed by the grouping algorithm of Section 3.3. As expected, there found to be two overlapping window groups which imply that there are two objects in the region of interest with centers located at $(-40.02, -111.70)$ cm and $(139.93, -108.46)$ cm.

As the second example of this section, we considered keeping one of the objects at a fixed location, and moving the other object around below the array. The first object is fixed at $(-40, -125)$ cm, and the other is moved from $x = 40$ cm to $x = 240$ cm in the lateral direction at two different depths, -125 cm and -137.5 cm, Fig. 12. The SNR is set to 30 dB. Contrary to mine-like object example, in this case the fixed object is at an advantageous location and detected with a probability of 1.0 regardless of the position of the moving object in the defined region. The moving object is hard to detect, since it is placed either outside the span of the array most of the time or deeper than the fixed object. Fig. 13 depicts the average probability of detection of the moving object over 100 Monte-Carlo simulations for two depths when probability of false alarm is set to 10^{-3} . It is clear that as the moving object is placed far from the array, it is less likely to be detected. Relative to the depth of the fixed object, if the moving object is placed deeper, it has to be nearer to the array to be detected. Simulated probability of false alarm ($P_{fa,sim}$) for this example is zero for all positions of the moving object, since both objects are relatively far apart to cause a false detection window.

We have also investigated the effect of relative distance and depth on estimated object centers.

For this purpose, we plotted the estimation error in x -direction ($\Delta x = x_{true} - x_{estimated}$) versus the estimation error in y -direction ($\Delta y = y_{true} - y_{estimated}$) for both fixed and moving objects for two depths in Fig. 14(a-b). The estimated object coordinates are obtained by averaging 100 Monte-Carlo simulations. The dashed circles indicate the boundaries of the objects. Circles shows the estimation errors when both objects are at the same depth, and diamonds show those when the moving object is a half radius deeper than the fixed object. Symbols (circle/diamond) inside the dashed circle indicate the estimation errors less than the radius of the object. In Fig. 14(a-b) it is observed that as the moving object gets far from the array, the estimation error in position of moving object increases, and that in position of fixed object becomes smaller. Placing the moving object at a deeper location increases the estimation error in its position, and slightly improves that in fixed object's position.

5 Conclusions

In this paper we present an algorithm that can detect and localize an unknown number of objects in the near field of a linear sensor array. The issues related to near-field scattering are overcome by partitioning the full array into sub-arrays so that the non-planar scattered field becomes locally planar at each sub-array. DOAs corresponding to these locally planar waves are then determined using array processing techniques. Triangulation of such DOAs results in dense and sparse regions of crossings which are modeled with a pair of spatial Poisson distributions. Estimated object locations, and the number of objects are finally obtained by applying a hypothesis test to Poisson models and then extracting groups of spatially disjoint detection windows.

We demonstrate the performance of the algorithm using simulated data. The usefulness of this algorithm is exhibited for both mine-like and drum-like objects. For mine-like targets, we show that

the algorithm can detect and localize multiple targets with different electrical properties. Then, we demonstrate the detectability of such objects when relative distance between them changes. We conclude that the detectability improves as the objects are located farther from each other while staying within a reasonable distance from the array. In addition, it is shown that theoretical and simulated probabilities of false alarm and detection are in agreement. For drum-like targets, we demonstrate the usefulness of the algorithm for detection and localization of multiple objects. Furthermore, the effect of relative distance and relative depth on detectability is treated. Results of this analysis supports the results obtained for mine-like objects.

As the future work, we want to combine the algorithm of this paper with the matched field processing (MFP) adapted for near-field object detection in [14]. The MFP is known for its accuracy and when used in tandem with the SAP, where SAP supplies rough estimates of object positions to MFP, it would be possible to localize closely located objects. The combined algorithm would inherit the best merits of both algorithms, viz. speed and efficiency of SAP and accuracy of MFP.

References

- [1] J.W. Choi and Y.H. Kim, “Spherical Beam-Forming and MUSIC Methods for the Estimation of Location and Strength of Spherical Sound Sources”, *Mechanical Systems and Signal Processing*, vol. 9, no. 5, pp. 569–588, 1995.
- [2] J.W. Choi and Y.H. Kim, “Estimation of Locations and Strengths of Broadband Planar and Spherical Noise Sources Using Coherent Signal Subspace”, *Journal of Acoustical Society of America*, vol. 98, no. 4, pp. 2082–2093, October 1995.
- [3] Y.D. Huang and M. Barkat, “Near-Field Multiple Source Localization by Passive Sensor Array”, *IEEE Transactions on Antennas and Propagation*, vol. 39, no. 7, pp. 968–974, July 1991.
- [4] E.L. Miller and A.S. Willsky, “A Multiscale, Statistically Based Inversion Scheme for Linearized Inverse Scattering Problems”, *IEEE Trans. on Geoscience and Remote Sensing*, vol. 34, no. 2, pp. 346–357, March 1996.

- [5] E.L. Miller and A.S. Willsky, "Multiscale, Statistical Anomaly Detection Analysis and Algorithms for Linearized Inverse Scattering Problems", *Multidimensional Systems and Signal Processing*, vol. 8, pp. 151–184, 1996.
- [6] A.J. Witten and J.E. Molyneux, "Ground Penetrating Radar Tomography: Algorithms and Case Studies", *IEEE Trans. Geoscience and Remote Sensing*, vol. 32, pp. 461–467, March 1994.
- [7] W.C. Chew and G.P. Otto, "Microwave Imaging of Multiple Conducting Cylinders Using Local Shape Functions", *IEEE Microwave and Guided Wave Letters*, vol. 2, no. 7, pp. 284–286, July 1992.
- [8] W.C. Chew and Y.M. Wang, "Reconstruction of Two-Dimensional Permittivity Distribution Using the Distorted Born Iterative Method", *IEEE Trans. on Medical Imaging*, vol. 9, no. 2, pp. 218–225, June 1990.
- [9] A.N. Tikhonov and V.Y. Arsenin, *Solutions to Ill-Posed Problems*, V.H. Winston & Sons, 1977.
- [10] H.Krim and M.Viberg, "Sensor Array Signal Processing: Two Decades Later", Tech. Rep. CICS-P-448, Center for Intelligent Control Systems, January 1995.
- [11] M. Bouvet and G. Bienvenu, Eds., *High-Resolution Methods in Underwater Acoustics*, Springer-Verlag, Heidelberg, Germany, 1991.
- [12] D.L. Snyder and M.I. Miller, *Random Point Processes in Time and Space*, Springer-Verlag New York Inc., 2nd edition, 1991.
- [13] R.O.Schmidt, "Multiple Emitter Location and Signal Parameter Estimation", *IEEE Trans. Antennas and Propagation*, vol. AP-34, no. 3, pp. 276–280, March 1986.
- [14] A. Şahin and E.L. Miller, "Object-Based Localization of Buried Objects Using High Resolution Array Processing Techniques", in *Proc. of SPIE– AeroSense: Detection and Remediation Technologies for Mines and Minelike Targets*, May 1996, vol. 2765, pp. 409–419.
- [15] D. C. Hoaglin, "A Poisson Plot", *The American Statistician*, vol. 34, no. 3, pp. 146–149, August 1980.
- [16] J.E. Hipp, "Soil Electromagnetic Parameters as Functions of Frequency, Soil Density, and Soil Moisture", *Proceedings of the IEEE*, vol. 62, no. 1, pp. 98–103, January 1974.
- [17] W.C. Chew, *Waves and Fields in Inhomogeneous Media*, Van Nostrand Reinhold, 1990.
- [18] L. Gürel and W.C. Chew, "A Recursive T-matrix Algorithm for Strips and Patches", *Radio Science*, vol. 27, pp. 387–401, May-June 1992.
- [19] A. Şahin and E.L. Miller, "Recursive T-matrix Methods for Scattering from Multiple Dielectric and Metallic Objects", *IEEE Trans. on Antennas and Propagation*, to appear in May 1998.
- [20] A. Şahin and E.L. Miller, "Recursive T-matrix Algorithm for Multiple Metallic Cylinders", *Microwave Opt. Tech. Lett.*, vol. 15, no. 6, pp. 360–363, August 1997.

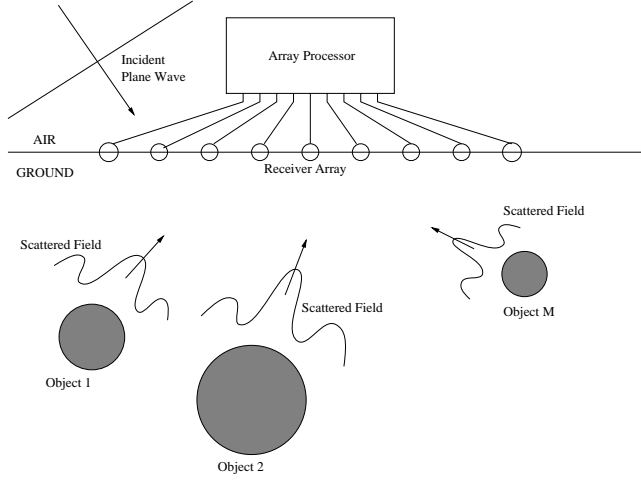
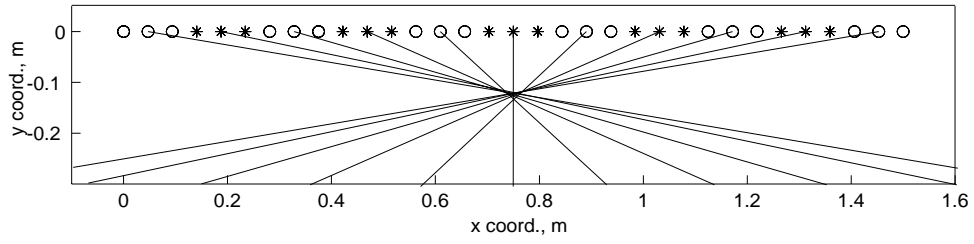
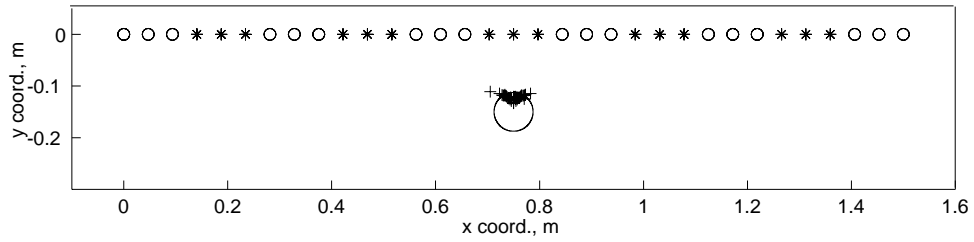


Figure 1: Problem geometry



(a) DOA triangulation



(b) Crossing pattern (+) and the object (dashed)

Figure 2: Single object localization with SAP: metallic mine-like object in lossy background

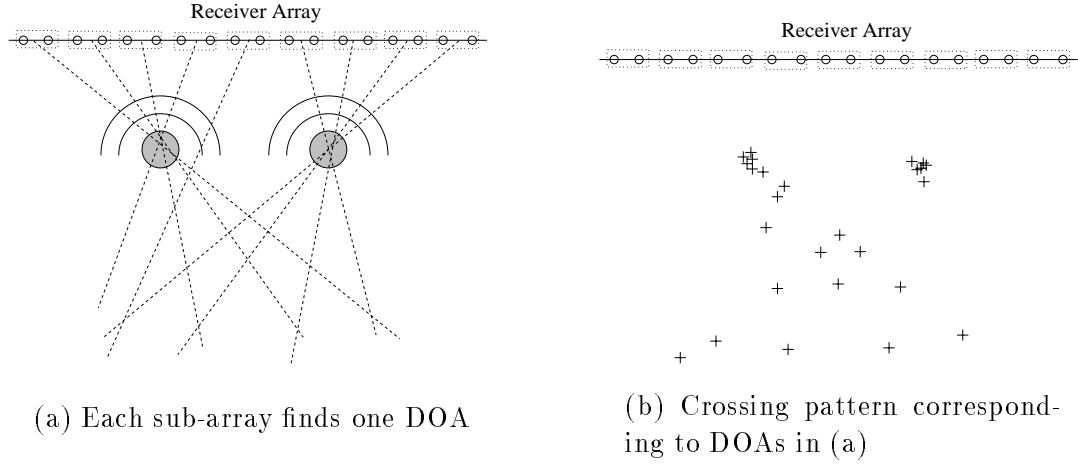


Figure 3: Multiple object subarray processing concept: directions of arrivals and crossing pattern

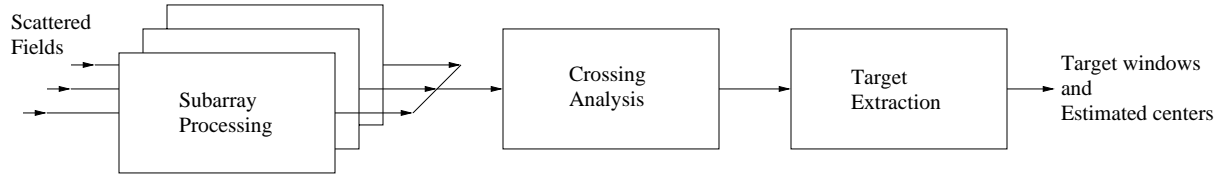


Figure 4: Flow chart of the algorithm

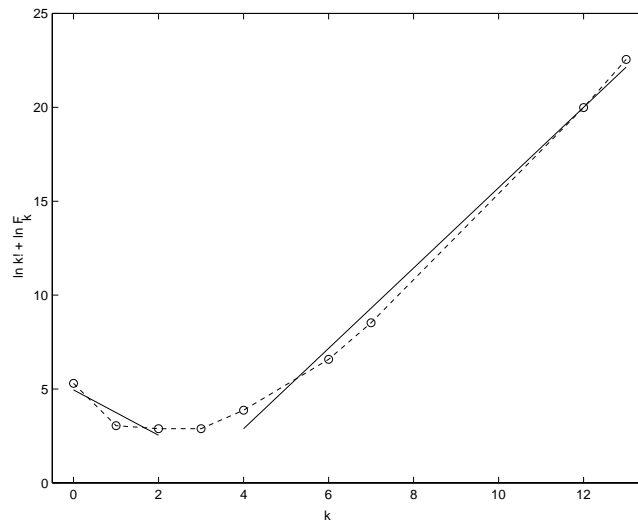
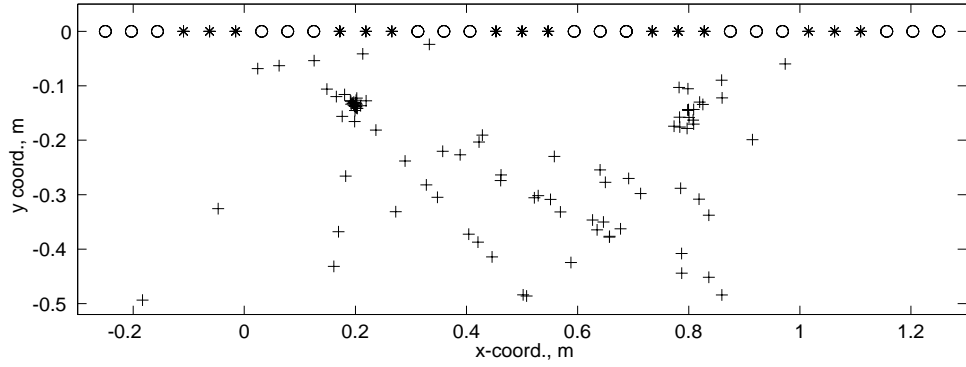
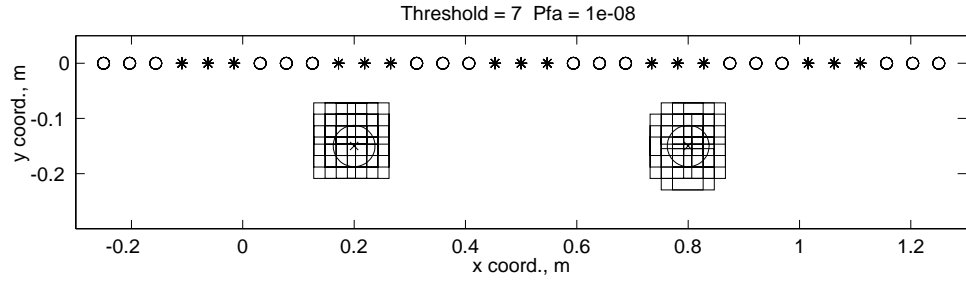


Figure 5: Fitness to Poisson model for the example of Fig. 6. Solid lines are fitted to data using linear least squares



(a) DOA crossings



(b) Detection windows overlaid on true object positions

Figure 6: Multiple mine-like object detection with SAP: object on the left is metallic and object on the right is dielectric with $\epsilon_r = 2.5$. o's and *'s denote subarrays

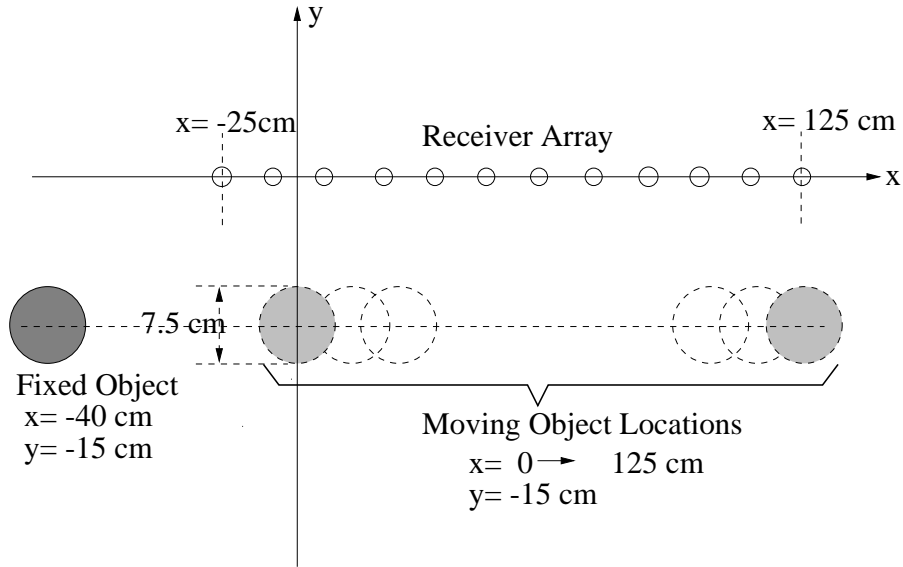


Figure 7: Multiple mine-like object detection with SAP: geometry for the example about the effect of relative distance between objects. Both objects are perfect conductors

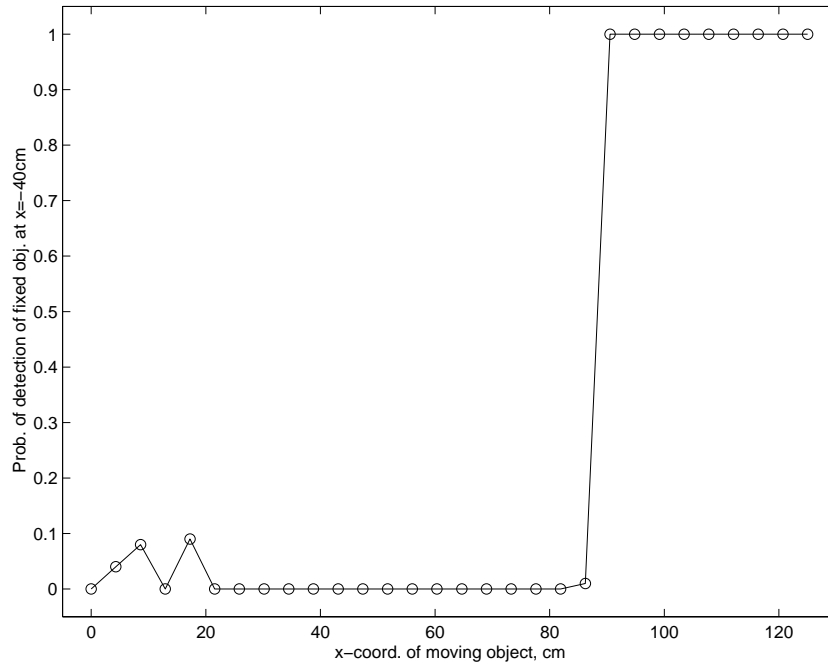


Figure 8: Probability of detection of the fixed object when the other object moves under the receiver array

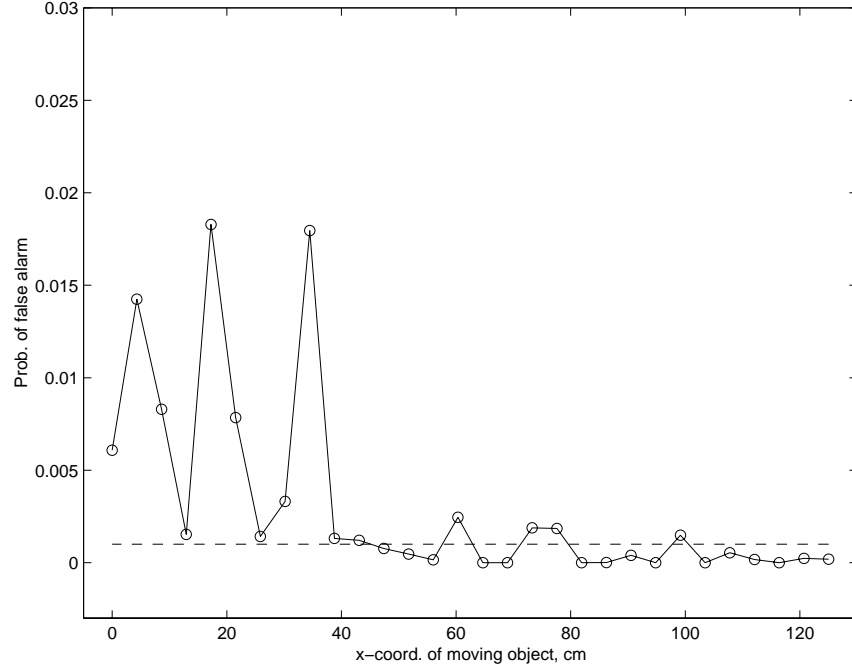


Figure 9: Probability of false alarm, when the moving object is located under the receiver array at various positions, dashed line shows the desired false alarm rate of 10^{-3}

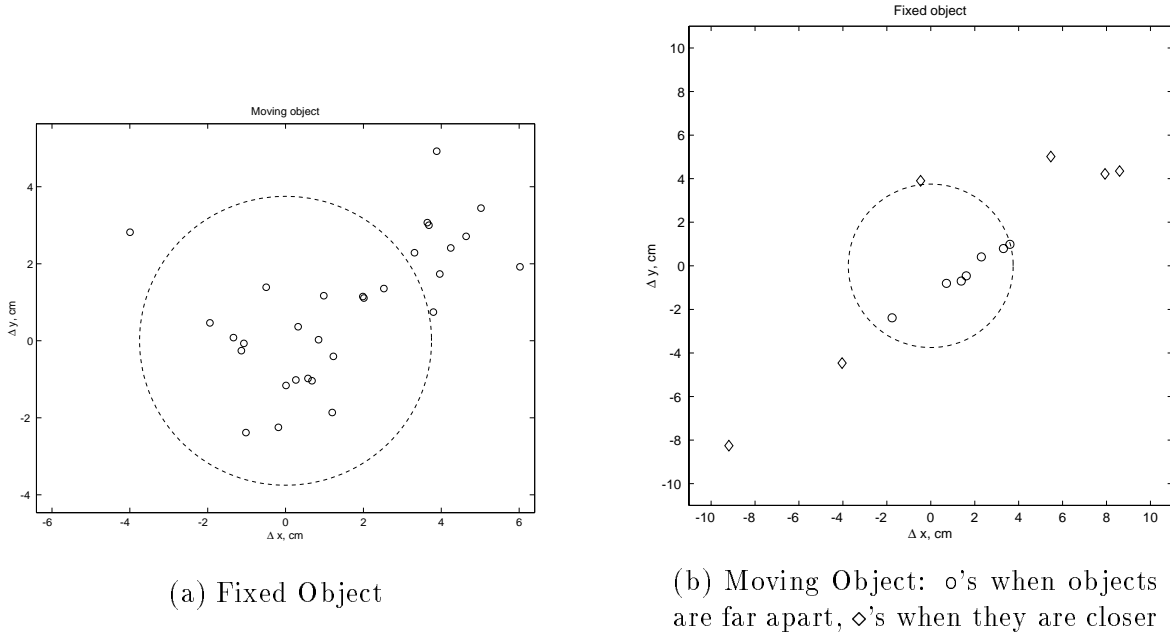
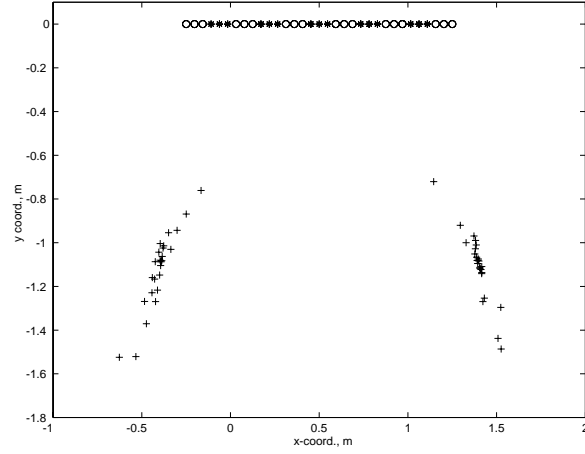
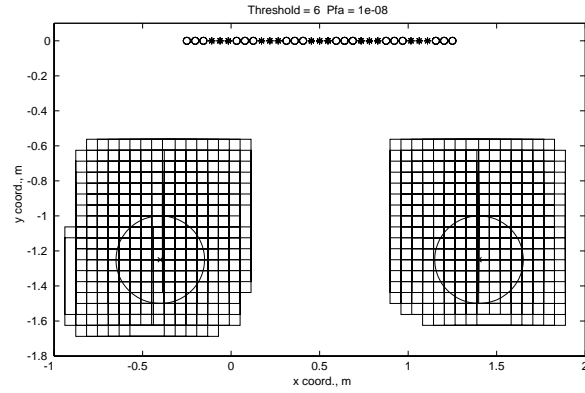


Figure 10: Estimation errors in object centers, each symbol corresponds to a different position of the moving object. Number of symbols in (b) is considerably less, since the fixed object is not detected for all positions of the moving object, see Fig. 8.



(a) DOA crossings



(b) Detection windows overlaid on true object positions

Figure 11: Localization of two metallic drum-like objects with SAP, o's and *'s denote subarrays

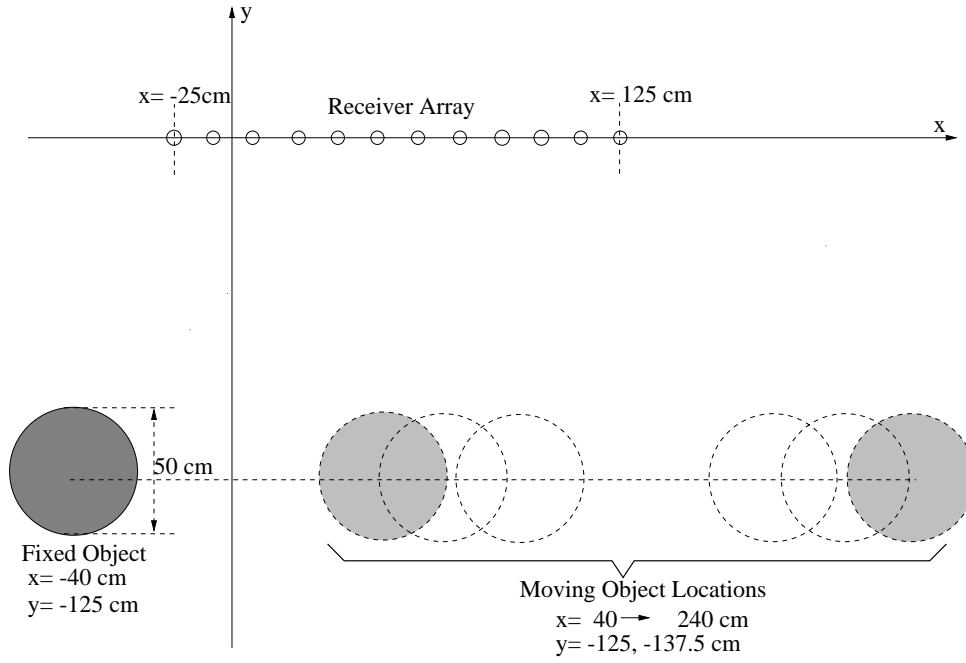


Figure 12: Multiple drum-like object detection with SAP: geometry for the example about the effect of relative distance and depth between objects. Both objects are perfect conductors

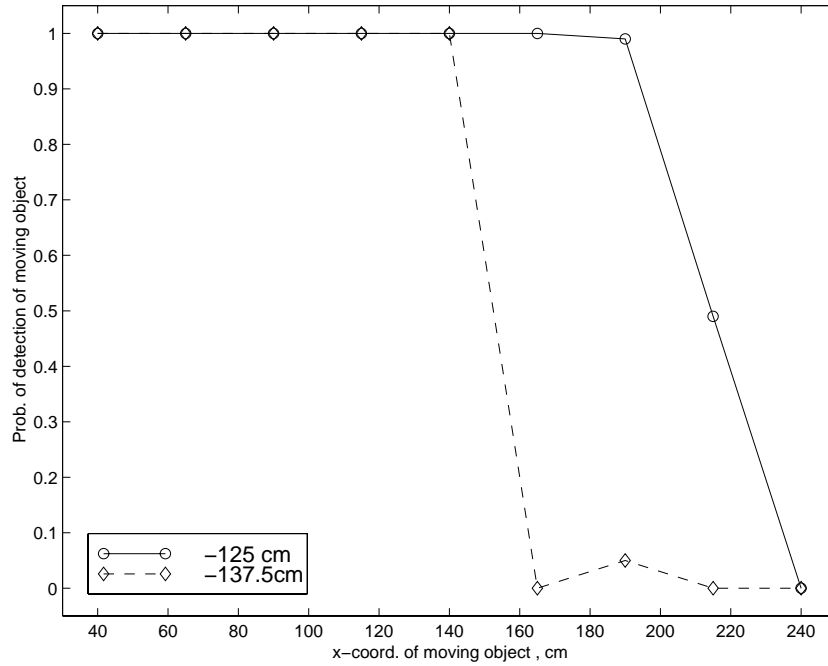
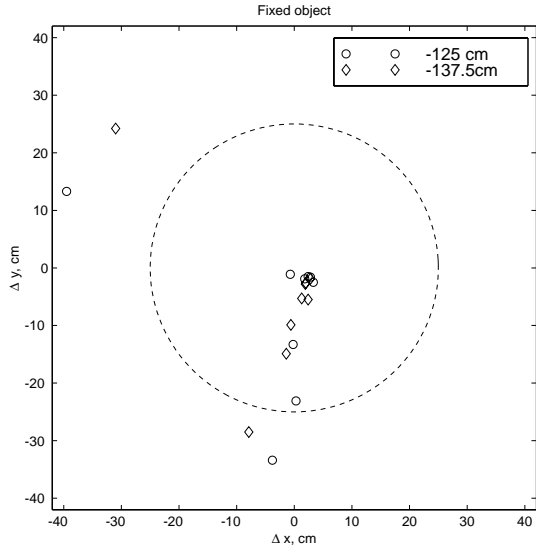
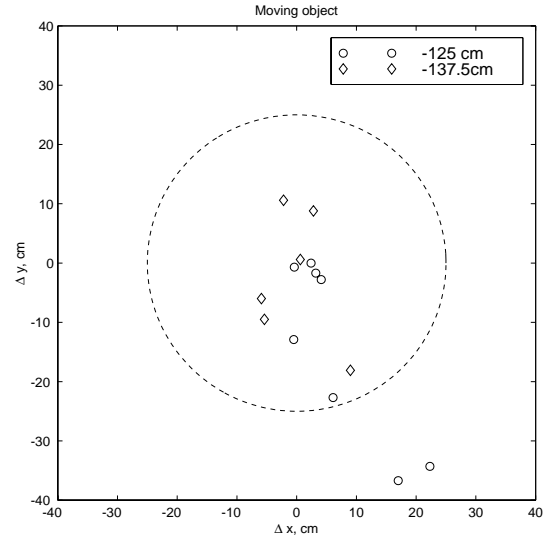


Figure 13: Probability of detection of the moving object, when the fixed object is located at $(-40, -125)$ cm. The solid line and dashed lines show the probability of detection when the moving object is at depth -125 cm and -137.5 cm, respectively.



(a) Fixed Object. Depth of the moving object: o=-125 cm and ◊=-137.5 cm.



(b) Moving Object. Depth of the moving object: o=-125 cm and ◊=-137.5 cm.

Figure 14: Estimation errors in object centers, each symbol corresponds to a different position of the moving object. Number of symbols in (b) is less than that in (a), since it is not detected for all positions of the moving object, see Fig. 13.

Performance Bounds for Matched Field Processing in Subsurface Object Detection Applications*

Adnan Şahin and Eric L. Miller

Center for Electromagnetics Research,
235 Forsyth Building, 360 Huntington Ave.
Northeastern University, Boston, MA 02115
Telephone: (617) 373-8386
Telefax : (617) 373-8627
E-mail:adnan@cdsp.neu.edu

ABSTRACT

In recent years there has been considerable interest in the use of ground penetrating radar (GPR) for the non-invasive detection and localization of buried objects. In a previous work, we have considered the use of high resolution array processing methods for solving these problems for measurement geometries in which an array of electromagnetic receivers observes the fields scattered by the subsurface targets in response to a plane wave illumination. Our approach uses the MUSIC algorithm in a matched field processing (MFP) scheme to determine both the range and the bearing of the objects.

In this paper we derive the Cramer-Rao bounds (CRB) for this MUSIC-based approach analytically. Analysis of the theoretical CRB has shown that there exists an optimum inter-element spacing of array elements for which the CRB is minimum. Furthermore, the optimum inter-element spacing minimizing CRB is smaller than the conventional half wavelength criterion. The theoretical bounds are then verified for two estimators using Monte-Carlo simulations. The first estimator is the MUSIC-based MFP and the second one is the maximum likelihood based MFP. The two approaches differ in the cost functions they optimize. We observe that Monte-Carlo simulated error variances always lie above the values established by CRB. Finally, we evaluate the performance of our MUSIC-based algorithm in the presence of model mismatches. Since the detection algorithm strongly depends on the model used, we have tested the performance of the algorithm when the object radius used in the model is different from the true radius. This analysis reveals that the algorithm is still capable of localizing the objects with a bias depending on the degree of mismatch.

Keywords: Matched Field Processing, MFP, Subsurface Detection, MUSIC, Cramer-Rao Bound, Performance Bound

* This work was supported in part DOE contract DE-FC07-95ID13395 and Army Research Office Demining MURI under Grant DAAG55-97-1-0013

1. INTRODUCTION

Non-invasive detection and localization of buried objects using ground penetrating radar (GPR) type geometries have been of interest to many researchers in recent years. Some of the most promising application areas for this technology include mine detection and hazardous waste remediation.

In subsurface object detection most often a pixel-by-pixel map of the region of interest is produced and the detection is then made by post-processing the image [1–3]. However, since this inverse problem is inherently ill-posed, the image-then-detect approach is quite sensitive to noise. While the ill-posedness can be off-set through regularization [4], typical regularization methods result in smooth images thereby making the detection all the more difficult. On the other hand, making use of the fact that the ultimate goal of the problem is the localization of objects with known structures, we can reparameterize the problem within an object-based detection framework [5–7]. Rather than looking for a large number of pixels, we extract the information about a small number of unknowns from the data: specifically the coordinates of the buried objects. By constraining the degrees of freedom in this manner, ill-posedness is substantially reduced and the resulting algorithms prove to be robust to noise.

The inherent array structure of GPR measurement geometry in Fig. 1 suggests that the high resolution array processing techniques [8], in particular, multiple signal classification (MUSIC) algorithm [9], quite popular in the signal processing community would be well suited for the subsurface detection problem. However, direction finding array processing techniques assume that the sources are infinitely far away so that the waveform received on the array is planar. For subsurface object detection case, on the other hand, the objects are located relatively closer to the receiver array, thus the scattered field from these objects across the array is not planar. In addition, for such near field objects, range as well as the direction of arrival has to be determined in order to localize the object. In ocean acoustics, the matched field processing (MFP) [10] has been successfully used for localization of sources modeled as point radiators. MFP is an array processing technique that uses the spatial complexities of the fields to localize sources, and thus allows for estimation of both the range and the bearing of the objects. Recently, MFP has been adapted to other application areas such as stratospherical electromagnetics [11], acoustics [12], and subsurface object detection [13]. In [13], we have demonstrated that MUSIC-based MFP can be used to detect and localize simple shaped metallic and dielectric objects over a wide range of soil characteristics and signal to noise ratios quite accurately.

In this paper, we extend the results of [13] to include the performance analysis for the matched field processing. Since we use a model based approach, it is imperative to know the performance of the algorithm against changes in problem parameters. For this purpose, we derive the Cramer-Rao performance bound for an unbiased estimator of the object position similar to derivations in [14]. Analytical bounds of range and bearing variables are then validated by running Monte-Carlo experiments for two estimators. The first one is MUSIC-based MFP described in [13], and the second one is the maximum likelihood (ML) based MFP. We observe that asymptotically MUSIC and ML estimators perform very similarly as reported in [14]. In addition, plotting the analytical CRB against the array length with fixed number of receivers reveals that there are optimal array lengths for estimating the range and bearing of an object, which minimizes the lower bound. Therefore, if one aims to determine the target's range or bearing with minimum variance, there exists an optimal array length, corresponding to an optimal inter-element spacing. Finally, we show the sensitivity of model mismatch on object radius by using a different radius in MFP model than the true radius of the object. This analysis demonstrates that under model mismatches, MFP can still localize the objects with a bias dependent on the degree of mismatch.

The remainder of the paper is organized as follows. In Section 2, we introduce the problem geometry, and brief overview of plane wave MUSIC. In Section 3, we review the matched field processing for subsurface

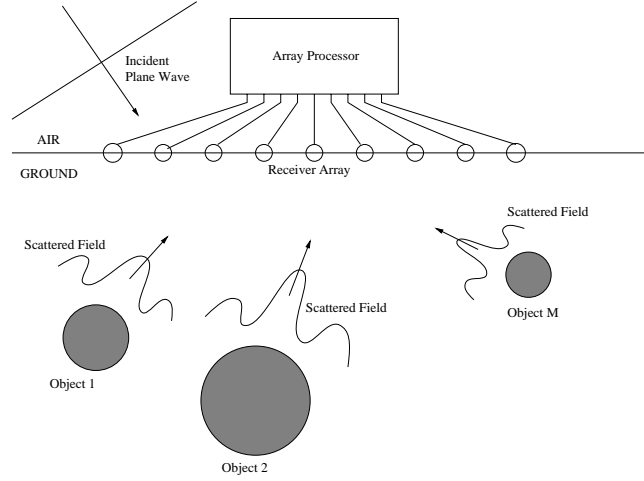


Figure 1. Problem geometry

detection problems. The Cramer-Rao lower bound for MFP is derived in Section 4. Numerical examples are presented in Section 5, and in Section 6, conclusions will be drawn and future directions will be presented.

2. PROBLEM GEOMETRY AND BACKGROUND

The ground penetrating radar (GPR) type measurement scheme depicted in Fig. 1, is considered in this paper. The objects are buried in a background for which the constant electrical characteristics (relative permittivity and conductivity) are assumed known. A transverse magnetic (TM) polarized plane wave, $E_i(\mathbf{r})$, impinges on the objects, inducing surface and volume currents which in turn radiate a scattered field, $E_s(\mathbf{r})^\dagger$. An important assumption in our approach is that the material properties and the shapes of the objects are known a priori. The scattered electric field from these objects is spatially sampled by a uniformly spaced, linear array with isotropic receiver characteristics. The measured data at the sensor outputs are:

$$\mathbf{y} = \mathbf{E}_s \mathbf{x} + \mathbf{n}, \quad (1)$$

where $\mathbf{E}_s = [E_s(\mathbf{r}_1) \ E_s(\mathbf{r}_2) \ \cdots E_s(\mathbf{r}_N)]^T$, \mathbf{r}_i , $i = 1, 2, \dots, N$, is the vector from the origin to the i th receiver location and \mathbf{n} is spatially and temporally white Gaussian noise. If M objects scatter the incident field, then \mathbf{E}_s becomes an $N \times M$ matrix with each column formed by the scattered field due to a different object. For time domain applications, the vector \mathbf{x} contains the time variations. Since we do our analysis in the frequency domain, and suppress $e^{j\omega t}$, $\mathbf{x} = 1$ for multiple scatterers, and $x = 1$ for single object. For the sake of simplicity, this paper deals with the case where there is only one scatterer in the region of interest. Thus, the data model can simply be written as:

$$\mathbf{y} = \mathbf{E}_s + \mathbf{n}. \quad (2)$$

The structure of the receivers in Fig. 1 coupled with the underlying problem of target detection suggests the use of array processing methods for localizing buried targets. In this paper we consider the MUSIC (Multiple Signal Classification) [9] algorithm. Traditionally, MUSIC and other direction finding techniques

[†]All analysis is in frequency domain, thus the $e^{j\omega t}$ time dependence will be dropped.

are used to determine the directions of arrival (DOA) of plane waves to a receiver array. In the rest of this section, we will briefly review the plane wave MUSIC.

The experiment as represented by the data model in (1) is repeated many times in order to determine the statistics of \mathbf{y} . In particular, if L experiments are performed, then the maximum likelihood estimate of the spatial autocovariance matrix $\mathbf{R} = E\{\mathbf{y}\mathbf{y}^H\}$ is given by:

$$\hat{\mathbf{R}} = \frac{1}{L} \sum_{l=1}^L \mathbf{y}_l \mathbf{y}_l^H \quad (3)$$

where \mathbf{y}_l is the data measured at the l th experiment, and superscript H denotes conjugate transpose. The sample covariance matrix $\hat{\mathbf{R}}$ is used in MUSIC algorithm to separate signal and noise subspaces through the eigenspace decomposition [8]:

$$\hat{\mathbf{R}} = \hat{\mathbf{U}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{U}}_s^H + \hat{\mathbf{U}}_n [\hat{\sigma}^2 \mathbf{I}] \hat{\mathbf{U}}_n^H \quad (4)$$

where $\hat{\mathbf{U}}_s$ is the estimated signal subspace matrix and contains the M signal eigenvectors, and $\hat{\mathbf{U}}_n$ is the estimated noise subspace matrix and contains $N - M$ noise eigenvectors of multiple noise eigenvalue $\hat{\sigma}^2$. The projection operator onto the noise subspace is defined as [8]:

$$\hat{\mathbf{\Pi}}_n = \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H. \quad (5)$$

The basic idea behind the planewave MUSIC algorithm is that the reciprocal of the “distance” between the estimated noise subspace and the true noise subspace has sharp peaks around the DOAs. Thus, if one plots this quantity versus all possible angles, estimates of DOAs can be determined by the maxima of the angular spectrum. The spatial spectrum of the MUSIC algorithm is given by [8]:

$$P_{MUSIC}(\theta) = \frac{\mathbf{a}(\theta)^H \mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \hat{\mathbf{\Pi}}_n \mathbf{a}(\theta)} \quad (6)$$

where $\mathbf{a}(\theta) = [1 \ e^{-j\beta d \cos \theta} \ e^{-j2\beta d \cos \theta} \ \dots \ e^{-j(N-1)\beta d \cos \theta}]^T$ is the direction vector that accounts for a plane wave impinging on the array, β is the wave number in the medium of propagation, and d is the distance between two receivers.

It is important to realize that the formulation of the array processing problem presented in this section implicitly assumes that the radiator is infinitely distant so that the scattered field has planar wavefronts at the sensor array. Thus, the elements of the direction vector $\mathbf{a}(\theta)$ are complex exponentials indicative of plane wave signals. However in many applications, including GPR, the receivers are in the near-field region of the radiating sources, resulting in non-planar wavefronts. The target localization problem, therefore, not only requires the DOA relative to the array but also the range of the target from a point on the array. The MUSIC-based MFP described in the next section supplements the plane wave MUSIC to address the near field localization.

3. MATCHED FIELD PROCESSING

As mentioned in the introduction, the matched field processing [10] uses the spatial complexities of the fields to localize sources in underwater acoustics. In a similar manner, plane wave MUSIC outlined in Section 2 can be modified so that the direction vector is filled with the type of the wavefront impinging on the array. As a result, the problem of non-planar wavefronts caused by near field scattering is resolved, and both the range and angle of the scatterer can be estimated. Therefore, MUSIC-based MFP forms the following spectrum:

$$P_{MUSIC}(r, \theta) = \frac{\mathbf{a}(r, \theta)^H \mathbf{a}(r, \theta)}{\mathbf{a}(r, \theta)^H \hat{\mathbf{\Pi}}_n \mathbf{a}(r, \theta)} \quad (7)$$

where the direction vector $\mathbf{a}(r, \theta)$ is now filled with the type of wavefront expected to impinge on the receiver array. For example, if one is concerned with the localization of independent point sources in the near field of the array, spherical MUSIC [12] can be employed to find both the DOAs and the distance of the object from the array. Spherical MUSIC uses $e^{-j\beta r_{mn}}/r_{mn}$ to fill in the direction vector, as opposed to plane wave MUSIC which uses $e^{-j\beta r_n \cos \theta_m}$, where r_{mn} is the distance between the n th array receiver and the m th source (object), r_n is the distance between the origin and the n th receiver, θ_m is the DOA of m th source, and β is the wavenumber in the medium of propagation. Since the detection problem is parameterized by an angle and a distance variable, a two dimensional search in this parameter space is necessary.

For the inverse scattering problem, however, the direction vector $\mathbf{a}(r, \theta)$ is filled with the scattered fields from possible object locations, and the location $(\hat{r}, \hat{\theta})$ maximizing the MUSIC spectrum in (7) is selected as the estimated object center. Because a two dimensional search requires that the exact scattered field be calculated at each point of the parameter mesh, MFP is in general computationally intensive. In the event that the object to be detected is modeled as a simple shape, however, computing the exact scattered field can be relatively simple. In this paper, we simplify the shapes of the objects by replacing them with infinitely long circular cylinders. With simple shapes, the forward scattering problem can easily be solved using harmonic expansions of the fields, see Appendix A for details. In these field calculations, the air-earth interface is ignored in order to simplify the formulation.

To summarize, the MUSIC-based MFP algorithm proceeds as follows:

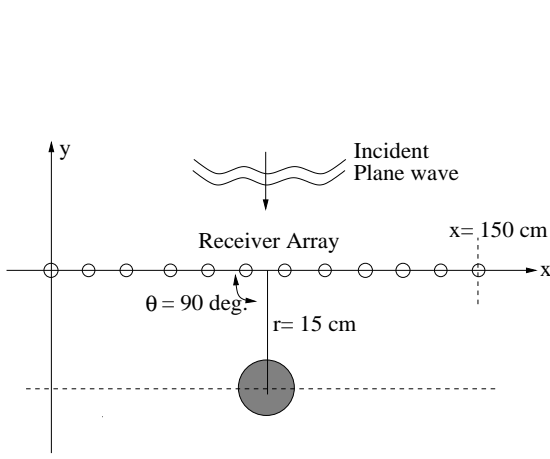
1. Using the data model in (1), perform L , single-frequency scattering experiments each producing a snapshot vector, \mathbf{y}_l , composed of the observed scattered fields over the receiver array
2. Estimate the autocovariance matrix $\hat{\mathbf{R}}$ using (3)
3. Perform an eigenanalysis on $\hat{\mathbf{R}}$ to find the noise-subspace projector $\hat{\mathbf{\Pi}}_n$ (4) and (5)
4. For each point (r_k, θ_k) on a predefined range and angle grid, fill in the direction vector $\mathbf{a}(r_k, \theta_k)$ with the scattered field due to an object placed in that location, and calculate the associated value of $P_{MUSIC}(r_k, \theta_k)$ in (7)
5. Choose $(\hat{r}, \hat{\theta})$ as the estimate of the target location that grid point with the largest P_{MUSIC} .

Fig. 2(b) shows the MUSIC spectrum when a metallic mine-like object in 7.5 cm diameter is placed 15 cm below the array as depicted in Fig. 2(a). The 33-element linear, uniform receiver array spans an aperture of 1.5 meters. All sensors are assumed to be ideal, isotropic receivers. The operating frequency is 1.0 GHz and the plane wave is incident with 90 degrees. The lossy, homogeneous background has the same electrical characteristics of 5% moist San Antonio clay loam or 10% moist Puerto Rico clay loam ($\epsilon_b = 6\epsilon_0$, $\sigma_b = 5 \times 10^{-2} \text{ S/m}$) at around 1.0 GHz [15]. The signal to noise ratio (SNR) is fixed at 0 decibels. As Fig. 2(b) depicts, the location of the object ($r = 15\text{cm}$, $\theta = 90^\circ$) is indicated with a very sharp peak. For relatively high SNR values, the peak remains prominent, and the estimation error is usually on the order of grid size. For lower SNR levels, the peak gets flatter, and as a result estimation errors increase.

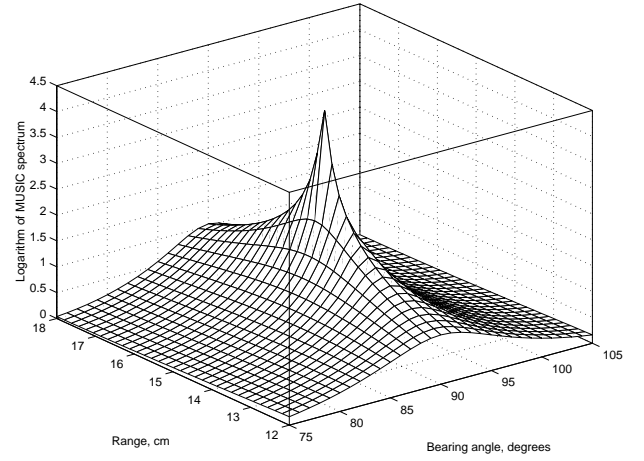
4. CRAMER-RAO BOUND FOR MATCHED FIELD PROCESSING

The Cramer-Rao Bound (CRB) provides very valuable information about the lower limit for the variance of any unbiased estimator. In order to find CRB, however, one should have a closed form expression of the log-likelihood function. In this section, we will extend the results in [14] to include the previously described problem of object localization in [13]. Since the additive noise in (1) is white, Gaussian distributed, the log-likelihood function can be written as [14]:

$$\ln \mathcal{L} = \text{constant} - 2NL \ln \sigma - \frac{1}{\sigma^2} \sum_{l=1}^L [\mathbf{y} - \mathbf{E}_s(r, \theta)\mathbf{x}]^H [\mathbf{y} - \mathbf{E}_s(r, \theta)\mathbf{x}] \quad (8)$$



(a) Geometry used for MFP Example



(b) Logarithm of MUSIC spectrum given by (7)

Figure 2. MFP localization example: single metallic object in a lossy, homogeneous background

where σ^2 is the noise variance, N is the number of receivers and L is the number of data vectors used for estimating the covariance matrix $\hat{\mathbf{R}}$.

Given the log-likelihood function, the Fisher Information Matrix (FIM) can be written as:

$$J = \begin{bmatrix} \Gamma_{rr} & \Gamma_{r\theta} \\ \Gamma_{\theta r} & \Gamma_{\theta\theta} \end{bmatrix} \quad (9)$$

where $\Gamma_{pq} = E[\frac{\partial^2 \ln \mathcal{L}}{\partial p \partial q}]$, $p = \{r, \theta\}$ and $q = \{r, \theta\}$. Following the same steps as [14], the entries of FIM, can be written as:

$$\Gamma_{pq} = \frac{2}{\sigma^2} \sum_{l=1}^L \text{Re}\{\mathbf{x}^H \mathbf{D}_p^H \mathbf{D}_q \mathbf{x}\}, \quad (10)$$

where $\mathbf{D}_p = \frac{\partial \mathbf{E}_s(r, \theta)}{\partial p}$ and $\mathbf{D}_q = \frac{\partial \mathbf{E}_s(r, \theta)}{\partial q}$ with $p = \{r, \theta\}$ and $q = \{r, \theta\}$. Since \mathbf{x} is constant over L experiments, (10) can be further reduced to:

$$\Gamma_{pq} = 2L \times \text{SNR} \times \text{Re}\{\mathbf{D}_p^H \mathbf{D}_q\} \quad (11)$$

where SNR is defined as $\frac{|x|^2}{\sigma^2}$. The CRB matrix for r and θ , then, can be expressed in terms of FIM entries as:

$$\text{CRB}(r, \theta) = \begin{bmatrix} \text{crb}_{rr} & \text{crb}_{r\theta} \\ \text{crb}_{\theta r} & \text{crb}_{\theta\theta} \end{bmatrix} = \begin{bmatrix} \Gamma_{rr} & \Gamma_{r\theta} \\ \Gamma_{\theta r} & \Gamma_{\theta\theta} \end{bmatrix}^{-1} \quad (12)$$

Finally, inserting (11) into (12), we obtain:

$$\text{CRB}(r, \theta) = \frac{1}{2L \times \text{SNR}} \left\{ \text{Re} \begin{bmatrix} \mathbf{D}_r^H \mathbf{D}_r & \mathbf{D}_r^H \mathbf{D}_\theta \\ \mathbf{D}_\theta^H \mathbf{D}_r & \mathbf{D}_\theta^H \mathbf{D}_\theta \end{bmatrix} \right\}^{-1}. \quad (13)$$

In the rest of the paper, we will refer crb_{rr} and $\text{crb}_{\theta\theta}$ as the Cramer-Rao Bounds of radial and azimuthal coordinate variables, respectively. The proper interpretation of CRB matrix is that if \hat{r} and $\hat{\theta}$ are unbiased estimates of polar variables r and θ , then

$$\begin{bmatrix} E\{(\hat{r} - r)^2\} & E\{(\hat{r} - r)(\hat{\theta} - \theta)\} \\ E\{(\hat{\theta} - \theta)(\hat{r} - r)\} & E\{(\hat{\theta} - \theta)^2\} \end{bmatrix} - \text{CRB}(r, \theta) \quad (14)$$

is positive semi-definite [16].

5. EXAMPLES

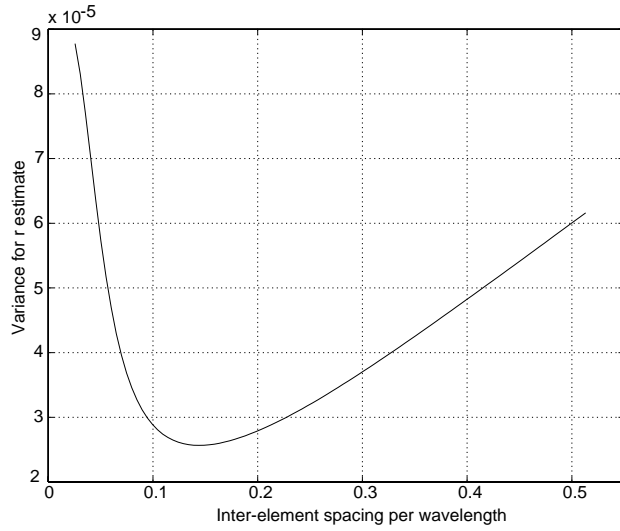
In this section we numerically evaluate the CRB derived in previous section for the geometry depicted in Fig. 2(a). The object to be localized is a metallic mine-like scatterer modeled as an infinitely long circular cylinder, with 7.5 cm diameter in cross-section. The object is placed 15 cm beneath a 33-element receiver array which spans an aperture of 1.5 meters. Center of the array is used as the origin of polar coordinate system that defines the object's radial and azimuthal coordinates. All sensors are assumed to be ideal, isotropic receivers, and form a linear, uniform receiver array. The operating frequency is 1.0 GHz and the plane wave is incident with 90 degrees. The calculation of scattered electromagnetic field and its derivatives is further simplified by ignoring the air-earth interface. The lossy, homogeneous background has the same electrical characteristics of 5% moist San Antonio clay loam or 10% moist Puerto Rico clay loam ($\epsilon_b = 6\epsilon_0$, $\sigma_b = 5 \times 10^{-2} \text{ S/m}$) at around 1.0 GHz [15]. The signal to noise ratio (SNR) is defined as $\frac{|x|^2}{\sigma^2}$, with $x = 1$, see the discussion in Section 2. In all examples, the data model (1) is generated $L = 250$ times to obtain the sample covariance matrix $\hat{\mathbf{R}}$ using (3).

We present several examples in this section: the first one is the evaluation of theoretical CRB given by (13), the second one is a comparison of analytical CRB with Monte-Carlo simulations, and the last one is about the effect of model mismatch on detection performance.

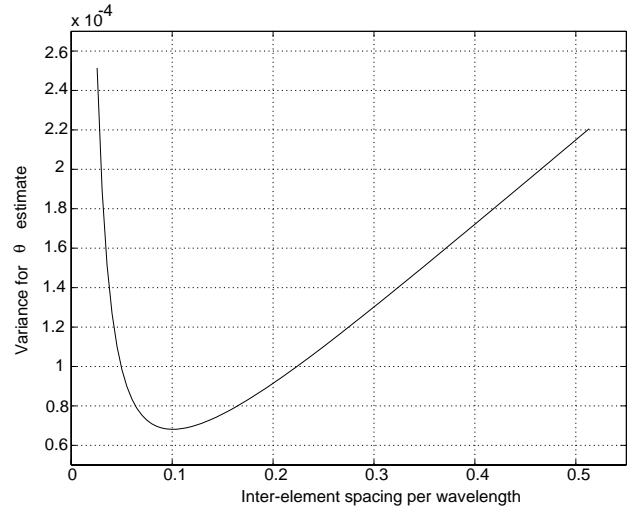
5.1. Evaluation of CRB

In order to determine CRB, one needs to calculate the derivatives of the scattered electric field E_s with respect to radial coordinate r and azimuthal coordinate θ . Finding these derivatives for single object is quite straightforward but tedious, the reader is referred to Appendix A for details of determining fields and their derivatives.

Fig. 3 depicts the dependence of CRB on inter-element spacing for the localization geometry given in Fig. 2(a). Here, the number of receivers is $N = 33$, and the length of array aperture is varied from 10 cm to 2 meters. For each array length, the object is placed under the center of the array. SNR is 0 decibels. Fig. 3(a) shows the variance of estimated radial coordinate \hat{r} , and Fig. 3(b) shows that for estimated azimuthal coordinate $\hat{\theta}$. Axes corresponding to inter-element spacing are normalized to the wavelength in the medium of propagation. It is interesting to note that these curves reach minimum for different inter-element spacings. This observation suggests that if one wants to minimize the variance of \hat{r} , there exists an estimator which approaches this minimum variance for a certain value of inter-element spacing, and this value is distinctly different from the one that minimizes the variance of $\hat{\theta}$. Fig. 3 also reveals that in order to achieve the minimum point on CRB curves both for \hat{r} and $\hat{\theta}$, inter-element spacing should be much smaller than the half-wavelength criterion suggested by the sampling theorem.

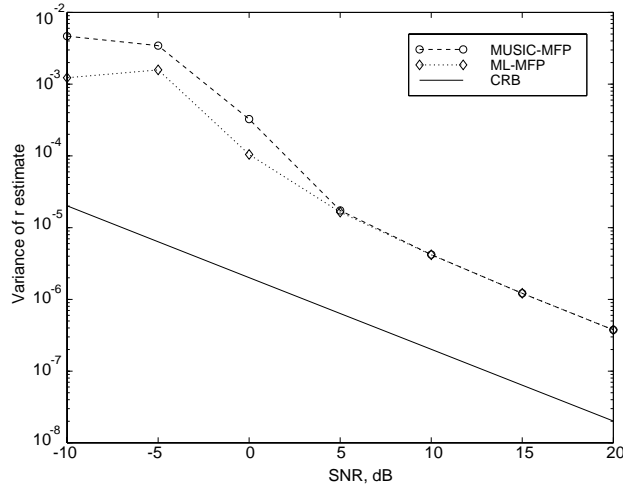


(a) Variance of \hat{r}

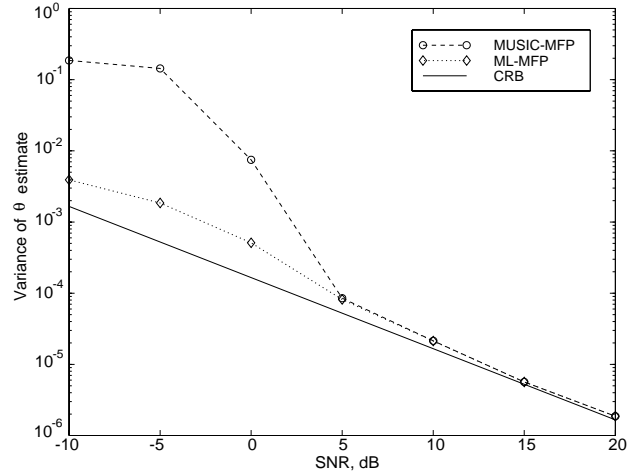


(b) Variance of $\hat{\theta}$

Figure 3. Analytical CRB *vs.* inter-element spacing



(a) Variance of \hat{r}



(b) Variance of $\hat{\theta}$

Figure 4. Comparison of analytical CRB with sample variances of MUSIC-MFP and ML-MFP

5.2. Comparison of CRB with Monte-Carlo Simulations

In this example Monte-Carlo simulated sample variances are compared with the analytical CRB given in (13). We have run the MFP algorithm defined in Section 3 for the geometry given in Fig. 2(a) 1000 times, each time obtaining an estimate of radial (\hat{r}) and azimuthal ($\hat{\theta}$) coordinates of the object. At the end,

sample variances of \hat{r} and $\hat{\theta}$ are calculated and compared with the CRB variances crb_{rr} and $\text{crb}_{\theta\theta}$ obtained analytically. Fig. 4 shows the comparison of variance of estimates obtained using MUSIC-MFP (circles) against the CRB (solid line). As expected, increasing signal to noise ratios results in decreasing variances. In addition, we compared the performance of MUSIC-MFP, and CRB with the maximum likelihood (ML) estimator. For this purpose, MFP is carried out by replacing MUSIC spectrum of (7) with ML cost [14] in the algorithm:

$$F(r, \theta) = \text{tr}\{[I - \mathbf{E}_s(r, \theta)(\mathbf{E}_s^H(r, \theta)\mathbf{E}_s(r, \theta))^{-1}\mathbf{E}_s^H(r, \theta)]\hat{\mathbf{R}}\} \quad (15)$$

where tr denotes the sum of diagonal entries. By minimizing $F(r, \theta)$, maximum likelihood estimates ($\hat{r}_{ML}, \hat{\theta}_{ML}$) of r and θ can be found. In Fig. 4, diamonds denote the variances of these quantities obtained with 1000 Monte-Carlo runs of ML-MFP. As these figures depict, ML-MFP performs better than MUSIC-MFP for low signal to noise ratios, and with increasing SNR both converge to the same value.

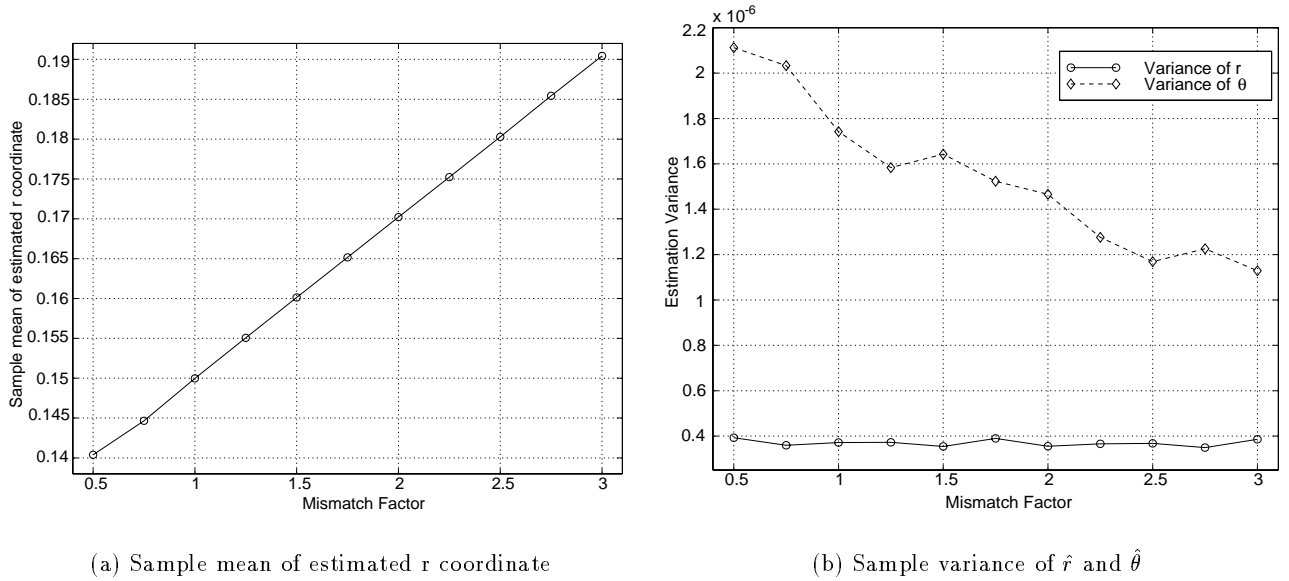


Figure 5. Effects of mismatch on estimated object location and estimation variances

5.3. Model Mismatch in Object Radius

As the last example, we consider the model mismatch in the object radius. Understanding the system performance is quite important in model-based detection algorithms. For this purpose we have defined a mismatch factor (MF), such that the MUSIC-MFP searches for an object of radius $\text{MF} \times \rho$, while the actual object is of radius ρ . The system described at the beginning of this section is used in the experiment with the object geometry given by Fig. 2(a). Signal to noise ratio is fixed at 20 decibels. The sample covariance matrix in (3) is evaluated using the scattered field from the object with radius ρ . Then, while estimating \hat{r} and $\hat{\theta}$, the direction vector $\mathbf{a}(r, \theta)$ in (7) is filled with the scattered field from an object whose radius is mismatched, i.e. $\rho_{\text{mismatch}} = \text{MF} \times \rho$. This experiment is repeated 500 times, and sample mean and variances of \hat{r} and $\hat{\theta}$ are plotted in Fig. 5. Part (a) of this figure depicts the sample mean of \hat{r} . It is interesting to note that, as the mismatch factor increases, MUSIC-MFP localizes the object with increasing bias. Therefore, for this example model mismatch causes bias in the radial coordinate's estimate. Even though not plotted here, sample mean of $\hat{\theta}$ is insensitive to the mismatch factor. This would be expected

since the changes in radius does not alter the symmetry around the center of the object, thus azimuthal coordinate's estimate is not affected. Fig. 5(b) shows that the variance of $\hat{\theta}$ decreases with increasing radius. This is expected given the plot in Fig. 5(a), since MUSIC-MFP thinks that an object of radius $MF \times \rho$ is buried at a depth different from the true depth. The variance of \hat{r} , on the other hand, remains almost constant over all mismatch factors.

6. CONCLUSIONS

In this paper, performance bounds for detection of objects using the matched field processing is derived and analytical results are verified with simulations. The object geometry is simplified with infinitely long circular cylinders embedded in a homogeneous background whose electrical characteristics are known. The uniform, linear array with ideal, isotropic sensors is used as the receiver array while the medium is illuminated by a plane wave.

In previous work [13], we have demonstrated that it is possible to use the matched field processing (MFP) to detect and localize mine and drum sized objects for various background and SNR scenarios. Here, we have derived analytical expressions of the performance bounds on estimates of object coordinates. The performance bounds basically indicate the best case scenario with the best estimator. Therefore, with these bounds available, it is possible to know if an object can be located with specified accuracy. Analysis of theoretical bounds have revealed that there exists an optimum inter-element spacing for which the bound is minimum. Thus, if an estimator is known to be bounded by CRB tightly, choosing the optimum inter-element spacing would minimize the estimation error. The analytical performance bounds are, then, verified with two simulation schemes: MUSIC based MFP, and ML based MFP. Simulations have verified that sample variances always lie above the values established by CRB. It has been observed that the difference between variances of estimates of object coordinates and their respective lower bounds gets smaller as the signal to noise ratio increases. Finally, it is shown that model mismatch in the object radius causes bias in estimating the object's radial coordinate. Furthermore, it is observed that if the object radius in the model is larger(smaller) than the true radius, variance of sample azimuthal coordinate $\hat{\theta}$ is smaller(larger).

As for the future work, we would like to extend present results in three directions. In the theoretical performance bound area, we will derive analytical expressions for the variances of estimation errors in MUSIC-MFP and extend the results of CRB analysis for multiple objects. In terms of simulations, we will run Monte-Carlo experiments to observe the change in error variances as the object location under the receiver array changes. For the model mismatch issue, we will extend the analysis to include mismatches in electrical properties of background and object. In order to account for mismatches due to object shapes, we would also like to analyze shape mismatch where the model and true object shapes are different.

APPENDIX A.

The scattered electromagnetic field from a cylindrical object with circular cross-section due to a plane wave incidence can be calculated using the harmonic expansions of the fields [17]. Given the center of the object (x_0, y_0) and incidence angle of the planewave ϕ_{inc} , the scattered field at the i th receiver location $\mathbf{r}_i = (r_i, \phi_i)$ in the coordinate system of object (Fig. 6) is given by:

$$E_s(\mathbf{r}_i) = \sum_{m=-\infty}^{\infty} j^{-m} c_m H_m^{(2)}(\beta r_i) e^{jm(\phi_i + \pi - \phi_{inc})} \quad (16)$$

where

$$c_m = -\frac{J_m(\beta \rho)}{H_m^{(2)}(\beta \rho)}$$

for metallic scatterers. The radius of the object is ρ and the wavenumber in the medium of propagation is β . The coefficients c_m decay quite rapidly as the number of harmonics increases. Thus, infinite sum in (16) can be truncated at M . The value of M is proportional to the radius of the object, $M \propto \beta\rho$. The variable M representing the number of harmonics here should not be confused with the same variable representing the dimension of the signal subspace in the body of the paper. It should be noted that the origin of coordinate system while calculating (16) is the center of the object, and coordinates of all sensors should be translated[†] to the object's coordinate system in polar form. Therefore, for every object location, this translation has to be carried out repeatedly.

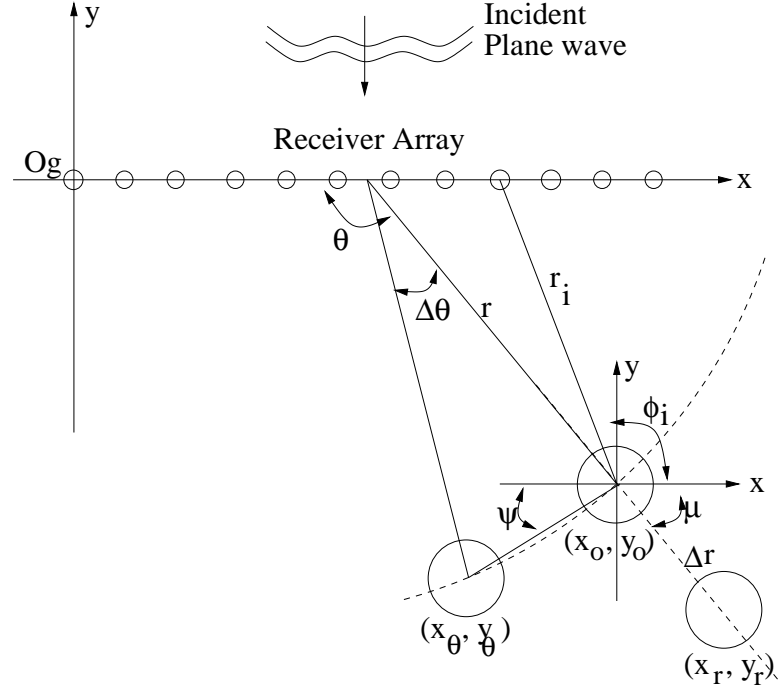


Figure 6. Geometry for calculation of the scattered field and its derivatives

The derivatives with respect to object's polar coordinates (r, θ) have to be taken carefully, since we have to translate coordinates of receivers into object's local coordinate system, and then calculate the scattered field in that coordinate system. Therefore, we use the definition of the derivative operation to find the derivatives of the fields:

$$\frac{\partial \mathbf{E}_s(r, \theta)}{\partial r} = \lim_{\Delta r \rightarrow 0} \frac{\mathbf{E}_s(r + \Delta r, \theta) - \mathbf{E}_s(r, \theta)}{\Delta r} \quad (17)$$

and

$$\frac{\partial \mathbf{E}_s(r, \theta)}{\partial \theta} = \lim_{\Delta \theta \rightarrow 0} \frac{\mathbf{E}_s(r, \theta + \Delta \theta) - \mathbf{E}_s(r, \theta)}{\Delta \theta}. \quad (18)$$

Here $\mathbf{E}_s(r, \theta) = [E_s(\mathbf{r}_1), E_s(\mathbf{r}_2), \dots, E_s(\mathbf{r}_N)]^T$, where N is the number of sensors in the array. To find $\mathbf{E}_s(r + \Delta r, \theta)$ and $\mathbf{E}_s(r, \theta + \Delta \theta)$, the center of the object should be expressed in terms of its original center (x_0, y_0) when it is moved by an infinitesimal amount in r and θ direction. When the object is moved in r -direction by Δr , the new center (x_r, y_r) becomes:

$$x_r = x_0 + \Delta r \cos(\mu) \quad y_r = y_0 - \Delta r \sin(\mu).$$

[†]Translation of coordinates between two points is straightforward and is not discussed here

The angle μ can be expressed in terms of θ . When the object is moved in θ -direction by $\Delta\theta$, the new center (x_θ, y_θ) becomes:

$$x_\theta = x_0 - r\Delta\theta \cos(\psi) \quad y_\theta = y_0 - r\Delta\theta \sin(\psi)$$

where ψ can be expressed in terms of θ and $\Delta\theta$. Once the new object centers are found, all receiver coordinates can be translated to the new object centers where the scattered field is calculated. Having obtained $\mathbf{E}_s(r + \Delta r, \theta)$ and $\mathbf{E}_s(r, \theta + \Delta\theta)$, the derivatives of the scattered fields with respect to r and θ can be determined using (17) and (18), respectively.

REFERENCES

1. A. Witten and J. Molyneux, "Ground penetrating radar tomography: algorithms and case studies," *IEEE Trans. Geoscience and Remote Sensing* **32**, pp. 461–467, March 1994.
2. E. Miller and A. Willsky, "A multiscale, statistically based inversion scheme for linearized inverse scattering problems," *IEEE Trans. on Geoscience and Remote Sensing* **34**, pp. 346–357, March 1996.
3. E. Miller and A. Willsky, "Multiscale, statistical anomaly detection analysis and algorithms for linearized inverse scattering problems," *Multidimensional Systems and Signal Processing* **8**, pp. 151–184, 1996.
4. A. Tikhonov and V. Arsenin, *Solutions to Ill-Posed Problems*, V.H. Winston & Sons, 1977.
5. Y. Bresler, J. Fessler, and A. Macovski, "Model-based estimation techniques for 3-d reconstruction from projections," *Machine Vision and Applications* **1**, pp. 115–126, 1988.
6. D. Rossi and A. Willsky, "Reconstruction from projections based on detection and estimation of objects— parts i and ii: Performance analysis and robustness analysis," *IEEE Trans. on Acoustics, Speech and Signal Processing* **32**, pp. 886–906, August 1984.
7. A. Devaney and G. Tsihrintzis, "Maximum likelihood estimation of object location in diffraction tomography," *IEEE Trans. on Signal Processing* **39**, pp. 672–682, March 1991.
8. H. Krim and M. Viberg, "Sensor array signal processing: Two decades later," Tech. Rep. CICS-P-448, Center for Intelligent Control Systems, January 1995.
9. R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation* **AP-34**, pp. 276–280, March 1986.
10. A. Baggeroerr, W. Kupperman, and P. Mikhalevsky, "An overview of matched field methods in ocean acoustics," *IEEE J. of Oceanic Engineering* **18**, pp. 401–424, October 1993.
11. D. Gingras, P. Gerstoft, and N. Gerr, "Electromagnetic matched-field processing: Basic concepts and tropospheric simulations," *IEEE Trans. on Antennas and Propagation* **45**, pp. 1536–1544, October 1997.
12. J. Choi and Y. Kim, "Spherical beam-forming and music methods for the estimation of location and strength of spherical sound sources," *Mechanical Systems and Signal Processing* **9**(5), pp. 569–588, 1995.
13. A. Şahin and E. Miller, "Object-based localization of buried objects using high resolution array processing techniques," in *Proceedings of SPIE—International Symposium on AeroSense*, vol. 2765 of *Detection and Remediation Technologies for Mines and Minelike Targets*, (Orlando, FL), 1996.
14. P. Stoica and A. Nehorai, "Music, maximum likelihood, and cramer-rao bound," *IEEE Trans. Acoust., Speech, Signal Processing* **37**, pp. 720–741, May 1989.
15. J. Hipp, "Soil electromagnetic parameters as functions of frequency, soil density, and soil moisture," *Proceedings of the IEEE* **62**, pp. 98–103, January 1974.
16. S. Silvey, *Statistical Inference*, Chapman and Hall Ltd., 1975.
17. C. Balanis, *Advanced Engineering Electromagnetics*, John Wiley and Sons, Inc, New York, 1989.

Performance Bounds for Matched Field Processing in Subsurface Object Detection Applications*

Adnan Şahin and Eric L. Miller

Center for Electromagnetics Research,
235 Forsyth Building, 360 Huntington Ave.
Northeastern University, Boston, MA 02115
Telephone: (617) 373-8386
Telefax : (617) 373-8627
E-mail: adnan@cdsp.neu.edu

ABSTRACT

In recent years there has been considerable interest in the use of ground penetrating radar (GPR) for the non-invasive detection and localization of buried objects. In a previous work, we have considered the use of high resolution array processing methods for solving these problems for measurement geometries in which an array of electromagnetic receivers observes the fields scattered by the subsurface targets in response to a plane wave illumination. Our approach uses the MUSIC algorithm in a matched field processing (MFP) scheme to determine both the range and the bearing of the objects.

In this paper we derive the Cramer-Rao bounds (CRB) for this MUSIC-based approach analytically. Analysis of the theoretical CRB has shown that there exists an optimum inter-element spacing of array elements for which the CRB is minimum. Furthermore, the optimum inter-element spacing minimizing CRB is smaller than the conventional half wavelength criterion. The theoretical bounds are then verified for two estimators using Monte-Carlo simulations. The first estimator is the MUSIC-based MFP and the second one is the maximum likelihood based MFP. The two approaches differ in the cost functions they optimize. We observe that Monte-Carlo simulated error variances always lie above the values established by CRB. Finally, we evaluate the performance of our MUSIC-based algorithm in the presence of model mismatches. Since the detection algorithm strongly depends on the model used, we have tested the performance of the algorithm when the object radius used in the model is different from the true radius. This analysis reveals that the algorithm is still capable of localizing the objects with a bias depending on the degree of mismatch.

Keywords: Matched Field Processing, MFP, Subsurface Detection, MUSIC, Cramer-Rao Bound, Performance Bound

* This work was supported in part DOE contract DE-FC07-95ID13395 and Army Research Office Demining MURI under Grant DAAG55-97-1-0013

1. INTRODUCTION

Non-invasive detection and localization of buried objects using ground penetrating radar (GPR) type geometries have been of interest to many researchers in recent years. Some of the most promising application areas for this technology include mine detection and hazardous waste remediation.

In subsurface object detection most often a pixel-by-pixel map of the region of interest is produced and the detection is then made by post-processing the image [1–3]. However, since this inverse problem is inherently ill-posed, the image-then-detect approach is quite sensitive to noise. While the ill-posedness can be off-set through regularization [4], typical regularization methods result in smooth images thereby making the detection all the more difficult. On the other hand, making use of the fact that the ultimate goal of the problem is the localization of objects with known structures, we can reparameterize the problem within an object-based detection framework [5–7]. Rather than looking for a large number of pixels, we extract the information about a small number of unknowns from the data: specifically the coordinates of the buried objects. By constraining the degrees of freedom in this manner, ill-posedness is substantially reduced and the resulting algorithms prove to be robust to noise.

The inherent array structure of GPR measurement geometry in Fig. 1 suggests that the high resolution array processing techniques [8], in particular, multiple signal classification (MUSIC) algorithm [9], quite popular in the signal processing community would be well suited for the subsurface detection problem. However, direction finding array processing techniques assume that the sources are infinitely far away so that the waveform received on the array is planar. For subsurface object detection case, on the other hand, the objects are located relatively closer to the receiver array, thus the scattered field from these objects across the array is not planar. In addition, for such near field objects, range as well as the direction of arrival has to be determined in order to localize the object. In ocean acoustics, the matched field processing (MFP) [10] has been successfully used for localization of sources modeled as point radiators. MFP is an array processing technique that uses the spatial complexities of the fields to localize sources, and thus allows for estimation of both the range and the bearing of the objects. Recently, MFP has been adapted to other application areas such as stratospherical electromagnetics [11], acoustics [12], and subsurface object detection [13]. In [13], we have demonstrated that MUSIC-based MFP can be used to detect and localize simple shaped metallic and dielectric objects over a wide range of soil characteristics and signal to noise ratios quite accurately.

In this paper, we extend the results of [13] to include the performance analysis for the matched field processing. Since we use a model based approach, it is imperative to know the performance of the algorithm against changes in problem parameters. For this purpose, we derive the Cramer-Rao performance bound for an unbiased estimator of the object position similar to derivations in [14]. Analytical bounds of range and bearing variables are then validated by running Monte-Carlo experiments for two estimators. The first one is MUSIC-based MFP described in [13], and the second one is the maximum likelihood (ML) based MFP. We observe that asymptotically MUSIC and ML estimators perform very similarly as reported in [14]. In addition, plotting the analytical CRB against the array length with fixed number of receivers reveals that there are optimal array lengths for estimating the range and bearing of an object, which minimizes the lower bound. Therefore, if one aims to determine the target's range or bearing with minimum variance, there exists an optimal array length, corresponding to an optimal inter-element spacing. Finally, we show the sensitivity of model mismatch on object radius by using a different radius in MFP model than the true radius of the object. This analysis demonstrates that under model mismatches, MFP can still localize the objects with a bias dependent on the degree of mismatch.

The remainder of the paper is organized as follows. In Section 2, we introduce the problem geometry, and brief overview of plane wave MUSIC. In Section 3, we review the matched field processing for subsurface

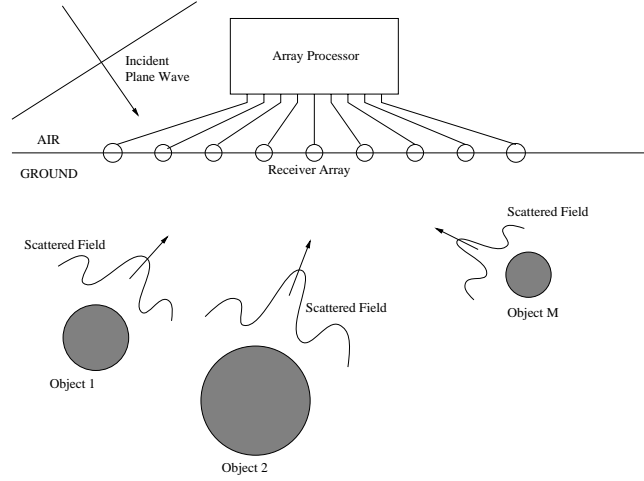


Figure 1. Problem geometry

detection problems. The Cramer-Rao lower bound for MFP is derived in Section 4. Numerical examples are presented in Section 5, and in Section 6, conclusions will be drawn and future directions will be presented.

2. PROBLEM GEOMETRY AND BACKGROUND

The ground penetrating radar (GPR) type measurement scheme depicted in Fig. 1, is considered in this paper. The objects are buried in a background for which the constant electrical characteristics (relative permittivity and conductivity) are assumed known. A transverse magnetic (TM) polarized plane wave, $E_i(\mathbf{r})$, impinges on the objects, inducing surface and volume currents which in turn radiate a scattered field, $E_s(\mathbf{r})^\dagger$. An important assumption in our approach is that the material properties and the shapes of the objects are known a priori. The scattered electric field from these objects is spatially sampled by a uniformly spaced, linear array with isotropic receiver characteristics. The measured data at the sensor outputs are:

$$\mathbf{y} = \mathbf{E}_s \mathbf{x} + \mathbf{n}, \quad (1)$$

where $\mathbf{E}_s = [E_s(\mathbf{r}_1) \ E_s(\mathbf{r}_2) \ \cdots E_s(\mathbf{r}_N)]^T$, \mathbf{r}_i , $i = 1, 2, \dots, N$, is the vector from the origin to the i th receiver location and \mathbf{n} is spatially and temporally white Gaussian noise. If M objects scatter the incident field, then \mathbf{E}_s becomes an $N \times M$ matrix with each column formed by the scattered field due to a different object. For time domain applications, the vector \mathbf{x} contains the time variations. Since we do our analysis in the frequency domain, and suppress $e^{j\omega t}$, $\mathbf{x} = 1$ for multiple scatterers, and $x = 1$ for single object. For the sake of simplicity, this paper deals with the case where there is only one scatterer in the region of interest. Thus, the data model can simply be written as:

$$\mathbf{y} = \mathbf{E}_s + \mathbf{n}. \quad (2)$$

The structure of the receivers in Fig. 1 coupled with the underlying problem of target detection suggests the use of array processing methods for localizing buried targets. In this paper we consider the MUSIC (Multiple Signal Classification) [9] algorithm. Traditionally, MUSIC and other direction finding techniques

[†]All analysis is in frequency domain, thus the $e^{j\omega t}$ time dependence will be dropped.

are used to determine the directions of arrival (DOA) of plane waves to a receiver array. In the rest of this section, we will briefly review the plane wave MUSIC.

The experiment as represented by the data model in (1) is repeated many times in order to determine the statistics of \mathbf{y} . In particular, if L experiments are performed, then the maximum likelihood estimate of the spatial autocovariance matrix $\mathbf{R} = E\{\mathbf{y}\mathbf{y}^H\}$ is given by:

$$\hat{\mathbf{R}} = \frac{1}{L} \sum_{l=1}^L \mathbf{y}_l \mathbf{y}_l^H \quad (3)$$

where \mathbf{y}_l is the data measured at the l th experiment, and superscript H denotes conjugate transpose. The sample covariance matrix $\hat{\mathbf{R}}$ is used in MUSIC algorithm to separate signal and noise subspaces through the eigenspace decomposition [8]:

$$\hat{\mathbf{R}} = \hat{\mathbf{U}}_s \hat{\mathbf{\Lambda}}_s \hat{\mathbf{U}}_s^H + \hat{\mathbf{U}}_n [\hat{\sigma}^2 \mathbf{I}] \hat{\mathbf{U}}_n^H \quad (4)$$

where $\hat{\mathbf{U}}_s$ is the estimated signal subspace matrix and contains the M signal eigenvectors, and $\hat{\mathbf{U}}_n$ is the estimated noise subspace matrix and contains $N - M$ noise eigenvectors of multiple noise eigenvalue $\hat{\sigma}^2$. The projection operator onto the noise subspace is defined as [8]:

$$\hat{\mathbf{\Pi}}_n = \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H. \quad (5)$$

The basic idea behind the planewave MUSIC algorithm is that the reciprocal of the “distance” between the estimated noise subspace and the true noise subspace has sharp peaks around the DOAs. Thus, if one plots this quantity versus all possible angles, estimates of DOAs can be determined by the maxima of the angular spectrum. The spatial spectrum of the MUSIC algorithm is given by [8]:

$$P_{MUSIC}(\theta) = \frac{\mathbf{a}(\theta)^H \mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \hat{\mathbf{\Pi}}_n \mathbf{a}(\theta)} \quad (6)$$

where $\mathbf{a}(\theta) = [1 \ e^{-j\beta d \cos \theta} \ e^{-j2\beta d \cos \theta} \ \dots \ e^{-j(N-1)\beta d \cos \theta}]^T$ is the direction vector that accounts for a plane wave impinging on the array, β is the wave number in the medium of propagation, and d is the distance between two receivers.

It is important to realize that the formulation of the array processing problem presented in this section implicitly assumes that the radiator is infinitely distant so that the scattered field has planar wavefronts at the sensor array. Thus, the elements of the direction vector $\mathbf{a}(\theta)$ are complex exponentials indicative of plane wave signals. However in many applications, including GPR, the receivers are in the near-field region of the radiating sources, resulting in non-planar wavefronts. The target localization problem, therefore, not only requires the DOA relative to the array but also the range of the target from a point on the array. The MUSIC-based MFP described in the next section supplements the plane wave MUSIC to address the near field localization.

3. MATCHED FIELD PROCESSING

As mentioned in the introduction, the matched field processing [10] uses the spatial complexities of the fields to localize sources in underwater acoustics. In a similar manner, plane wave MUSIC outlined in Section 2 can be modified so that the direction vector is filled with the type of the wavefront impinging on the array. As a result, the problem of non-planar wavefronts caused by near field scattering is resolved, and both the range and angle of the scatterer can be estimated. Therefore, MUSIC-based MFP forms the following spectrum:

$$P_{MUSIC}(r, \theta) = \frac{\mathbf{a}(r, \theta)^H \mathbf{a}(r, \theta)}{\mathbf{a}(r, \theta)^H \hat{\mathbf{\Pi}}_n \mathbf{a}(r, \theta)} \quad (7)$$

where the direction vector $\mathbf{a}(r, \theta)$ is now filled with the type of wavefront expected to impinge on the receiver array. For example, if one is concerned with the localization of independent point sources in the near field of the array, spherical MUSIC [12] can be employed to find both the DOAs and the distance of the object from the array. Spherical MUSIC uses $e^{-j\beta r_{mn}}/r_{mn}$ to fill in the direction vector, as opposed to plane wave MUSIC which uses $e^{-j\beta r_n \cos \theta_m}$, where r_{mn} is the distance between the n th array receiver and the m th source (object), r_n is the distance between the origin and the n th receiver, θ_m is the DOA of m th source, and β is the wavenumber in the medium of propagation. Since the detection problem is parameterized by an angle and a distance variable, a two dimensional search in this parameter space is necessary.

For the inverse scattering problem, however, the direction vector $\mathbf{a}(r, \theta)$ is filled with the scattered fields from possible object locations, and the location $(\hat{r}, \hat{\theta})$ maximizing the MUSIC spectrum in (7) is selected as the estimated object center. Because a two dimensional search requires that the exact scattered field be calculated at each point of the parameter mesh, MFP is in general computationally intensive. In the event that the object to be detected is modeled as a simple shape, however, computing the exact scattered field can be relatively simple. In this paper, we simplify the shapes of the objects by replacing them with infinitely long circular cylinders. With simple shapes, the forward scattering problem can easily be solved using harmonic expansions of the fields, see Appendix A for details. In these field calculations, the air-earth interface is ignored in order to simplify the formulation.

To summarize, the MUSIC-based MFP algorithm proceeds as follows:

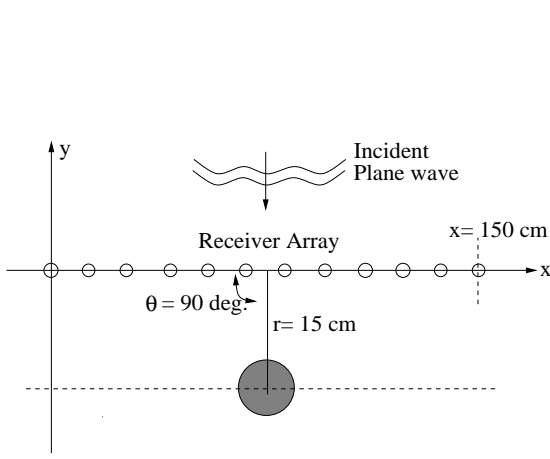
1. Using the data model in (1), perform L , single-frequency scattering experiments each producing a snapshot vector, \mathbf{y}_l , composed of the observed scattered fields over the receiver array
2. Estimate the autocovariance matrix $\hat{\mathbf{R}}$ using (3)
3. Perform an eigenanalysis on $\hat{\mathbf{R}}$ to find the noise-subspace projector $\hat{\mathbf{\Pi}}_n$ (4) and (5)
4. For each point (r_k, θ_k) on a predefined range and angle grid, fill in the direction vector $\mathbf{a}(r_k, \theta_k)$ with the scattered field due to an object placed in that location, and calculate the associated value of $P_{MUSIC}(r_k, \theta_k)$ in (7)
5. Choose $(\hat{r}, \hat{\theta})$ as the estimate of the target location that grid point with the largest P_{MUSIC} .

Fig. 2(b) shows the MUSIC spectrum when a metallic mine-like object in 7.5 cm diameter is placed 15 cm below the array as depicted in Fig. 2(a). The 33-element linear, uniform receiver array spans an aperture of 1.5 meters. All sensors are assumed to be ideal, isotropic receivers. The operating frequency is 1.0 GHz and the plane wave is incident with 90 degrees. The lossy, homogeneous background has the same electrical characteristics of 5% moist San Antonio clay loam or 10% moist Puerto Rico clay loam ($\epsilon_b = 6\epsilon_0$, $\sigma_b = 5 \times 10^{-2} \text{ S/m}$) at around 1.0 GHz [15]. The signal to noise ratio (SNR) is fixed at 0 decibels. As Fig. 2(b) depicts, the location of the object ($r = 15\text{cm}$, $\theta = 90^\circ$) is indicated with a very sharp peak. For relatively high SNR values, the peak remains prominent, and the estimation error is usually on the order of grid size. For lower SNR levels, the peak gets flatter, and as a result estimation errors increase.

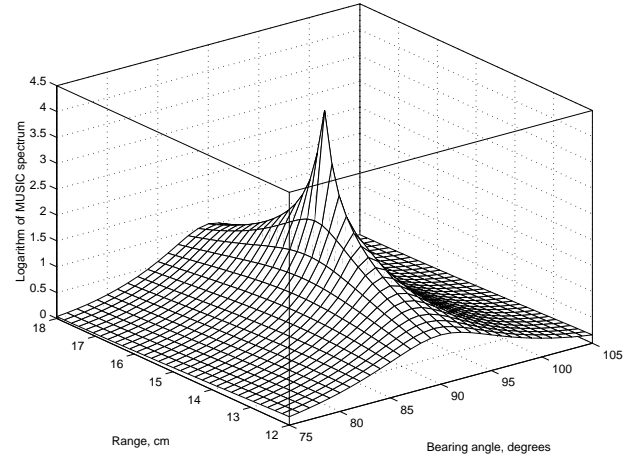
4. CRAMER-RAO BOUND FOR MATCHED FIELD PROCESSING

The Cramer-Rao Bound (CRB) provides very valuable information about the lower limit for the variance of any unbiased estimator. In order to find CRB, however, one should have a closed form expression of the log-likelihood function. In this section, we will extend the results in [14] to include the previously described problem of object localization in [13]. Since the additive noise in (1) is white, Gaussian distributed, the log-likelihood function can be written as [14]:

$$\ln \mathcal{L} = \text{constant} - 2NL \ln \sigma - \frac{1}{\sigma^2} \sum_{l=1}^L [\mathbf{y} - \mathbf{E}_s(r, \theta)\mathbf{x}]^H [\mathbf{y} - \mathbf{E}_s(r, \theta)\mathbf{x}] \quad (8)$$



(a) Geometry used for MFP Example



(b) Logarithm of MUSIC spectrum given by (7)

Figure 2. MFP localization example: single metallic object in a lossy, homogeneous background

where σ^2 is the noise variance, N is the number of receivers and L is the number of data vectors used for estimating the covariance matrix $\hat{\mathbf{R}}$.

Given the log-likelihood function, the Fisher Information Matrix (FIM) can be written as:

$$\mathbf{J} = \begin{bmatrix} \Gamma_{rr} & \Gamma_{r\theta} \\ \Gamma_{\theta r} & \Gamma_{\theta\theta} \end{bmatrix} \quad (9)$$

where $\Gamma_{pq} = E[\frac{\partial^2 \ln \mathcal{L}}{\partial p \partial q}]$, $p = \{r, \theta\}$ and $q = \{r, \theta\}$. Following the same steps as [14], the entries of FIM, can be written as:

$$\Gamma_{pq} = \frac{2}{\sigma^2} \sum_{l=1}^L \text{Re}\{\mathbf{x}^H \mathbf{D}_p^H \mathbf{D}_q \mathbf{x}\}, \quad (10)$$

where $\mathbf{D}_p = \frac{\partial \mathbf{E}_s(r, \theta)}{\partial p}$ and $\mathbf{D}_q = \frac{\partial \mathbf{E}_s(r, \theta)}{\partial q}$ with $p = \{r, \theta\}$ and $q = \{r, \theta\}$. Since \mathbf{x} is constant over L experiments, (10) can be further reduced to:

$$\Gamma_{pq} = 2L \times \text{SNR} \times \text{Re}\{\mathbf{D}_p^H \mathbf{D}_q\} \quad (11)$$

where SNR is defined as $\frac{|x|^2}{\sigma^2}$. The CRB matrix for r and θ , then, can be expressed in terms of FIM entries as:

$$\text{CRB}(r, \theta) = \begin{bmatrix} \text{crb}_{rr} & \text{crb}_{r\theta} \\ \text{crb}_{\theta r} & \text{crb}_{\theta\theta} \end{bmatrix} = \begin{bmatrix} \Gamma_{rr} & \Gamma_{r\theta} \\ \Gamma_{\theta r} & \Gamma_{\theta\theta} \end{bmatrix}^{-1} \quad (12)$$

Finally, inserting (11) into (12), we obtain:

$$\text{CRB}(r, \theta) = \frac{1}{2L \times \text{SNR}} \left\{ \text{Re} \begin{bmatrix} \mathbf{D}_r^H \mathbf{D}_r & \mathbf{D}_r^H \mathbf{D}_\theta \\ \mathbf{D}_\theta^H \mathbf{D}_r & \mathbf{D}_\theta^H \mathbf{D}_\theta \end{bmatrix} \right\}^{-1}. \quad (13)$$

In the rest of the paper, we will refer crb_{rr} and $\text{crb}_{\theta\theta}$ as the Cramer-Rao Bounds of radial and azimuthal coordinate variables, respectively. The proper interpretation of CRB matrix is that if \hat{r} and $\hat{\theta}$ are unbiased estimates of polar variables r and θ , then

$$\begin{bmatrix} E\{(\hat{r} - r)^2\} & E\{(\hat{r} - r)(\hat{\theta} - \theta)\} \\ E\{(\hat{\theta} - \theta)(\hat{r} - r)\} & E\{(\hat{\theta} - \theta)^2\} \end{bmatrix} - \text{CRB}(r, \theta) \quad (14)$$

is positive semi-definite [16].

5. EXAMPLES

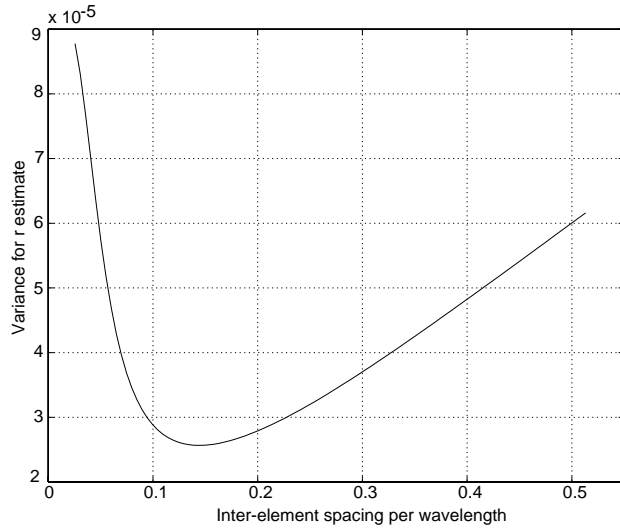
In this section we numerically evaluate the CRB derived in previous section for the geometry depicted in Fig. 2(a). The object to be localized is a metallic mine-like scatterer modeled as an infinitely long circular cylinder, with 7.5 cm diameter in cross-section. The object is placed 15 cm beneath a 33-element receiver array which spans an aperture of 1.5 meters. Center of the array is used as the origin of polar coordinate system that defines the object's radial and azimuthal coordinates. All sensors are assumed to be ideal, isotropic receivers, and form a linear, uniform receiver array. The operating frequency is 1.0 GHz and the plane wave is incident with 90 degrees. The calculation of scattered electromagnetic field and its derivatives is further simplified by ignoring the air-earth interface. The lossy, homogeneous background has the same electrical characteristics of 5% moist San Antonio clay loam or 10% moist Puerto Rico clay loam ($\epsilon_b = 6\epsilon_0$, $\sigma_b = 5 \times 10^{-2} \text{ S/m}$) at around 1.0 GHz [15]. The signal to noise ratio (SNR) is defined as $\frac{|x|^2}{\sigma^2}$, with $x = 1$, see the discussion in Section 2. In all examples, the data model (1) is generated $L = 250$ times to obtain the sample covariance matrix $\hat{\mathbf{R}}$ using (3).

We present several examples in this section: the first one is the evaluation of theoretical CRB given by (13), the second one is a comparison of analytical CRB with Monte-Carlo simulations, and the last one is about the effect of model mismatch on detection performance.

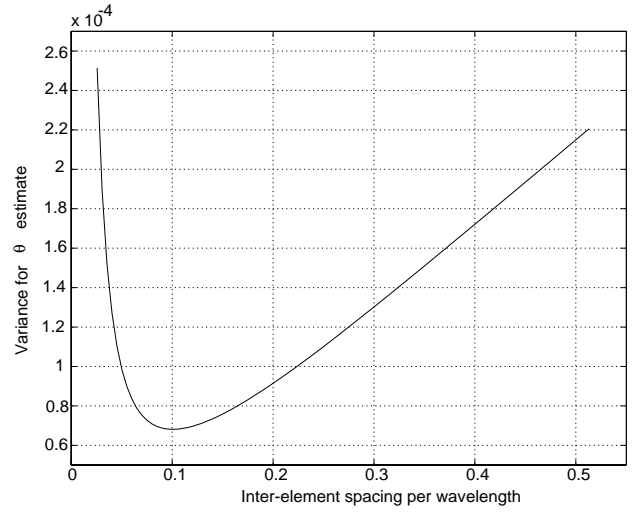
5.1. Evaluation of CRB

In order to determine CRB, one needs to calculate the derivatives of the scattered electric field E_s with respect to radial coordinate r and azimuthal coordinate θ . Finding these derivatives for single object is quite straightforward but tedious, the reader is referred to Appendix A for details of determining fields and their derivatives.

Fig. 3 depicts the dependence of CRB on inter-element spacing for the localization geometry given in Fig. 2(a). Here, the number of receivers is $N = 33$, and the length of array aperture is varied from 10 cm to 2 meters. For each array length, the object is placed under the center of the array. SNR is 0 decibels. Fig. 3(a) shows the variance of estimated radial coordinate \hat{r} , and Fig. 3(b) shows that for estimated azimuthal coordinate $\hat{\theta}$. Axes corresponding to inter-element spacing are normalized to the wavelength in the medium of propagation. It is interesting to note that these curves reach minimum for different inter-element spacings. This observation suggests that if one wants to minimize the variance of \hat{r} , there exists an estimator which approaches this minimum variance for a certain value of inter-element spacing, and this value is distinctly different from the one that minimizes the variance of $\hat{\theta}$. Fig. 3 also reveals that in order to achieve the minimum point on CRB curves both for \hat{r} and $\hat{\theta}$, inter-element spacing should be much smaller than the half-wavelength criterion suggested by the sampling theorem.

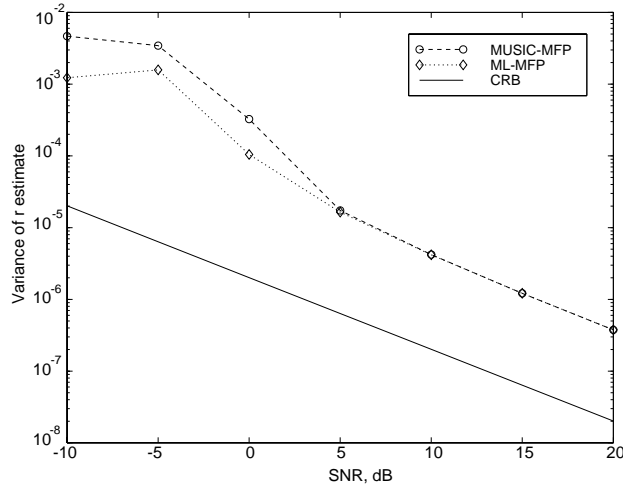


(a) Variance of \hat{r}

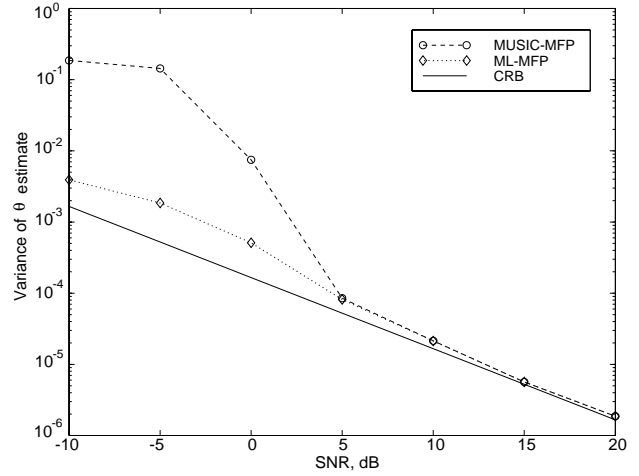


(b) Variance of $\hat{\theta}$

Figure 3. Analytical CRB *vs.* inter-element spacing



(a) Variance of \hat{r}



(b) Variance of $\hat{\theta}$

Figure 4. Comparison of analytical CRB with sample variances of MUSIC-MFP and ML-MFP

5.2. Comparison of CRB with Monte-Carlo Simulations

In this example Monte-Carlo simulated sample variances are compared with the analytical CRB given in (13). We have run the MFP algorithm defined in Section 3 for the geometry given in Fig. 2(a) 1000 times, each time obtaining an estimate of radial (\hat{r}) and azimuthal ($\hat{\theta}$) coordinates of the object. At the end,

sample variances of \hat{r} and $\hat{\theta}$ are calculated and compared with the CRB variances crb_{rr} and $\text{crb}_{\theta\theta}$ obtained analytically. Fig. 4 shows the comparison of variance of estimates obtained using MUSIC-MFP (circles) against the CRB (solid line). As expected, increasing signal to noise ratios results in decreasing variances. In addition, we compared the performance of MUSIC-MFP, and CRB with the maximum likelihood (ML) estimator. For this purpose, MFP is carried out by replacing MUSIC spectrum of (7) with ML cost [14] in the algorithm:

$$F(r, \theta) = \text{tr}\{[I - \mathbf{E}_s(r, \theta)(\mathbf{E}_s^H(r, \theta)\mathbf{E}_s(r, \theta))^{-1}\mathbf{E}_s^H(r, \theta)]\hat{\mathbf{R}}\} \quad (15)$$

where tr denotes the sum of diagonal entries. By minimizing $F(r, \theta)$, maximum likelihood estimates ($\hat{r}_{ML}, \hat{\theta}_{ML}$) of r and θ can be found. In Fig. 4, diamonds denote the variances of these quantities obtained with 1000 Monte-Carlo runs of ML-MFP. As these figures depict, ML-MFP performs better than MUSIC-MFP for low signal to noise ratios, and with increasing SNR both converge to the same value.

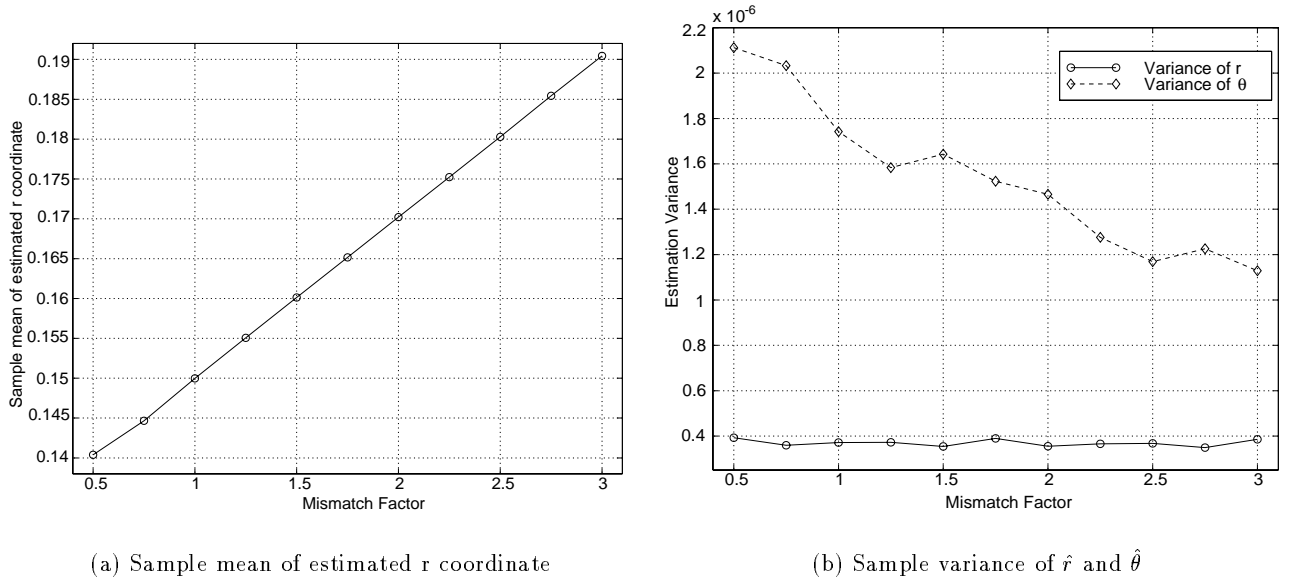


Figure 5. Effects of mismatch on estimated object location and estimation variances

5.3. Model Mismatch in Object Radius

As the last example, we consider the model mismatch in the object radius. Understanding the system performance is quite important in model-based detection algorithms. For this purpose we have defined a mismatch factor (MF), such that the MUSIC-MFP searches for an object of radius $\text{MF} \times \rho$, while the actual object is of radius ρ . The system described at the beginning of this section is used in the experiment with the object geometry given by Fig. 2(a). Signal to noise ratio is fixed at 20 decibels. The sample covariance matrix in (3) is evaluated using the scattered field from the object with radius ρ . Then, while estimating \hat{r} and $\hat{\theta}$, the direction vector $\mathbf{a}(r, \theta)$ in (7) is filled with the scattered field from an object whose radius is mismatched, i.e. $\rho_{\text{mismatch}} = \text{MF} \times \rho$. This experiment is repeated 500 times, and sample mean and variances of \hat{r} and $\hat{\theta}$ are plotted in Fig. 5. Part (a) of this figure depicts the sample mean of \hat{r} . It is interesting to note that, as the mismatch factor increases, MUSIC-MFP localizes the object with increasing bias. Therefore, for this example model mismatch causes bias in the radial coordinate's estimate. Even though not plotted here, sample mean of $\hat{\theta}$ is insensitive to the mismatch factor. This would be expected

since the changes in radius does not alter the symmetry around the center of the object, thus azimuthal coordinate's estimate is not affected. Fig. 5(b) shows that the variance of $\hat{\theta}$ decreases with increasing radius. This is expected given the plot in Fig. 5(a), since MUSIC-MFP thinks that an object of radius $MF \times \rho$ is buried at a depth different from the true depth. The variance of \hat{r} , on the other hand, remains almost constant over all mismatch factors.

6. CONCLUSIONS

In this paper, performance bounds for detection of objects using the matched field processing is derived and analytical results are verified with simulations. The object geometry is simplified with infinitely long circular cylinders embedded in a homogeneous background whose electrical characteristics are known. The uniform, linear array with ideal, isotropic sensors is used as the receiver array while the medium is illuminated by a plane wave.

In previous work [13], we have demonstrated that it is possible to use the matched field processing (MFP) to detect and localize mine and drum sized objects for various background and SNR scenarios. Here, we have derived analytical expressions of the performance bounds on estimates of object coordinates. The performance bounds basically indicate the best case scenario with the best estimator. Therefore, with these bounds available, it is possible to know if an object can be located with specified accuracy. Analysis of theoretical bounds have revealed that there exists an optimum inter-element spacing for which the bound is minimum. Thus, if an estimator is known to be bounded by CRB tightly, choosing the optimum inter-element spacing would minimize the estimation error. The analytical performance bounds are, then, verified with two simulation schemes: MUSIC based MFP, and ML based MFP. Simulations have verified that sample variances always lie above the values established by CRB. It has been observed that the difference between variances of estimates of object coordinates and their respective lower bounds gets smaller as the signal to noise ratio increases. Finally, it is shown that model mismatch in the object radius causes bias in estimating the object's radial coordinate. Furthermore, it is observed that if the object radius in the model is larger(smaller) than the true radius, variance of sample azimuthal coordinate $\hat{\theta}$ is smaller(larger).

As for the future work, we would like to extend present results in three directions. In the theoretical performance bound area, we will derive analytical expressions for the variances of estimation errors in MUSIC-MFP and extend the results of CRB analysis for multiple objects. In terms of simulations, we will run Monte-Carlo experiments to observe the change in error variances as the object location under the receiver array changes. For the model mismatch issue, we will extend the analysis to include mismatches in electrical properties of background and object. In order to account for mismatches due to object shapes, we would also like to analyze shape mismatch where the model and true object shapes are different.

APPENDIX A.

The scattered electromagnetic field from a cylindrical object with circular cross-section due to a plane wave incidence can be calculated using the harmonic expansions of the fields [17]. Given the center of the object (x_0, y_0) and incidence angle of the planewave ϕ_{inc} , the scattered field at the i th receiver location $\mathbf{r}_i = (r_i, \phi_i)$ in the coordinate system of object (Fig. 6) is given by:

$$E_s(\mathbf{r}_i) = \sum_{m=-\infty}^{\infty} j^{-m} c_m H_m^{(2)}(\beta r_i) e^{jm(\phi_i + \pi - \phi_{inc})} \quad (16)$$

where

$$c_m = -\frac{J_m(\beta \rho)}{H_m^{(2)}(\beta \rho)}$$

for metallic scatterers. The radius of the object is ρ and the wavenumber in the medium of propagation is β . The coefficients c_m decay quite rapidly as the number of harmonics increases. Thus, infinite sum in (16) can be truncated at M . The value of M is proportional to the radius of the object, $M \propto \beta\rho$. The variable M representing the number of harmonics here should not be confused with the same variable representing the dimension of the signal subspace in the body of the paper. It should be noted that the origin of coordinate system while calculating (16) is the center of the object, and coordinates of all sensors should be translated[†] to the object's coordinate system in polar form. Therefore, for every object location, this translation has to be carried out repeatedly.

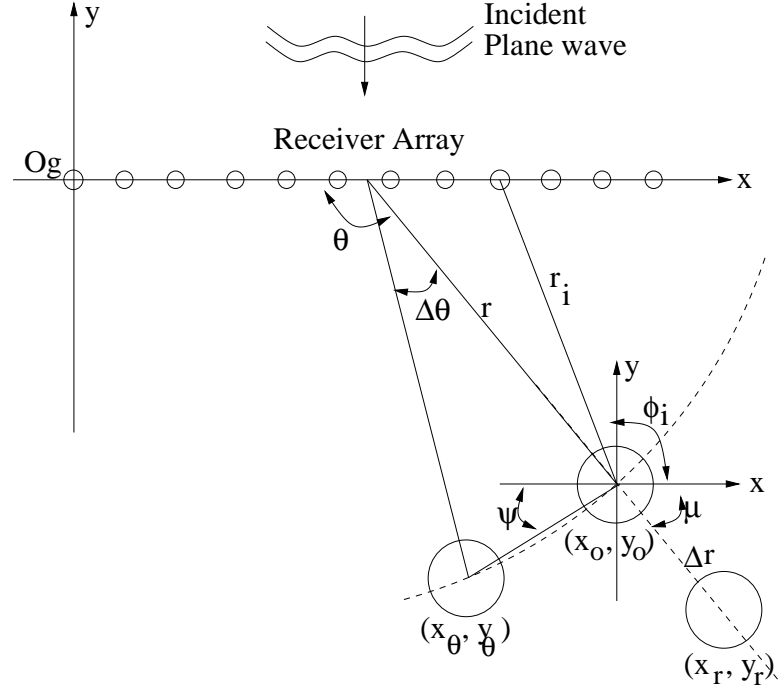


Figure 6. Geometry for calculation of the scattered field and its derivatives

The derivatives with respect to object's polar coordinates (r, θ) have to be taken carefully, since we have to translate coordinates of receivers into object's local coordinate system, and then calculate the scattered field in that coordinate system. Therefore, we use the definition of the derivative operation to find the derivatives of the fields:

$$\frac{\partial \mathbf{E}_s(r, \theta)}{\partial r} = \lim_{\Delta r \rightarrow 0} \frac{\mathbf{E}_s(r + \Delta r, \theta) - \mathbf{E}_s(r, \theta)}{\Delta r} \quad (17)$$

and

$$\frac{\partial \mathbf{E}_s(r, \theta)}{\partial \theta} = \lim_{\Delta \theta \rightarrow 0} \frac{\mathbf{E}_s(r, \theta + \Delta \theta) - \mathbf{E}_s(r, \theta)}{\Delta \theta}. \quad (18)$$

Here $\mathbf{E}_s(r, \theta) = [E_s(\mathbf{r}_1), E_s(\mathbf{r}_2), \dots, E_s(\mathbf{r}_N)]^T$, where N is the number of sensors in the array. To find $\mathbf{E}_s(r + \Delta r, \theta)$ and $\mathbf{E}_s(r, \theta + \Delta \theta)$, the center of the object should be expressed in terms of its original center (x_0, y_0) when it is moved by an infinitesimal amount in r and θ direction. When the object is moved in r -direction by Δr , the new center (x_r, y_r) becomes:

$$x_r = x_0 + \Delta r \cos(\mu) \quad y_r = y_0 - \Delta r \sin(\mu).$$

[†]Translation of coordinates between two points is straightforward and is not discussed here

The angle μ can be expressed in terms of θ . When the object is moved in θ -direction by $\Delta\theta$, the new center (x_θ, y_θ) becomes:

$$x_\theta = x_0 - r\Delta\theta \cos(\psi) \quad y_\theta = y_0 - r\Delta\theta \sin(\psi)$$

where ψ can be expressed in terms of θ and $\Delta\theta$. Once the new object centers are found, all receiver coordinates can be translated to the new object centers where the scattered field is calculated. Having obtained $\mathbf{E}_s(r + \Delta r, \theta)$ and $\mathbf{E}_s(r, \theta + \Delta\theta)$, the derivatives of the scattered fields with respect to r and θ can be determined using (17) and (18), respectively.

REFERENCES

1. A. Witten and J. Molyneux, "Ground penetrating radar tomography: algorithms and case studies," *IEEE Trans. Geoscience and Remote Sensing* **32**, pp. 461–467, March 1994.
2. E. Miller and A. Willsky, "A multiscale, statistically based inversion scheme for linearized inverse scattering problems," *IEEE Trans. on Geoscience and Remote Sensing* **34**, pp. 346–357, March 1996.
3. E. Miller and A. Willsky, "Multiscale, statistical anomaly detection analysis and algorithms for linearized inverse scattering problems," *Multidimensional Systems and Signal Processing* **8**, pp. 151–184, 1996.
4. A. Tikhonov and V. Arsenin, *Solutions to Ill-Posed Problems*, V.H. Winston & Sons, 1977.
5. Y. Bresler, J. Fessler, and A. Macovski, "Model-based estimation techniques for 3-d reconstruction from projections," *Machine Vision and Applications* **1**, pp. 115–126, 1988.
6. D. Rossi and A. Willsky, "Reconstruction from projections based on detection and estimation of objects— parts i and ii: Performance analysis and robustness analysis," *IEEE Trans. on Acoustics, Speech and Signal Processing* **32**, pp. 886–906, August 1984.
7. A. Devaney and G. Tsihrintzis, "Maximum likelihood estimation of object location in diffraction tomography," *IEEE Trans. on Signal Processing* **39**, pp. 672–682, March 1991.
8. H. Krim and M. Viberg, "Sensor array signal processing: Two decades later," Tech. Rep. CICS-P-448, Center for Intelligent Control Systems, January 1995.
9. R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation* **AP-34**, pp. 276–280, March 1986.
10. A. Baggeroerr, W. Kupperman, and P. Mikhalevsky, "An overview of matched field methods in ocean acoustics," *IEEE J. of Oceanic Engineering* **18**, pp. 401–424, October 1993.
11. D. Gingras, P. Gerstoft, and N. Gerr, "Electromagnetic matched-field processing: Basic concepts and tropospheric simulations," *IEEE Trans. on Antennas and Propagation* **45**, pp. 1536–1544, October 1997.
12. J. Choi and Y. Kim, "Spherical beam-forming and music methods for the estimation of location and strength of spherical sound sources," *Mechanical Systems and Signal Processing* **9**(5), pp. 569–588, 1995.
13. A. Şahin and E. Miller, "Object-based localization of buried objects using high resolution array processing techniques," in *Proceedings of SPIE—International Symposium on AeroSense*, vol. 2765 of *Detection and Remediation Technologies for Mines and Minelike Targets*, (Orlando, FL), 1996.
14. P. Stoica and A. Nehorai, "Music, maximum likelihood, and cramer-rao bound," *IEEE Trans. Acoust., Speech, Signal Processing* **37**, pp. 720–741, May 1989.
15. J. Hipp, "Soil electromagnetic parameters as functions of frequency, soil density, and soil moisture," *Proceedings of the IEEE* **62**, pp. 98–103, January 1974.
16. S. Silvey, *Statistical Inference*, Chapman and Hall Ltd., 1975.
17. C. Balanis, *Advanced Engineering Electromagnetics*, John Wiley and Sons, Inc, New York, 1989.

Multiscale, Statistical Anomaly Detection Analysis and Algorithms for Linearized Inverse Scattering Problems *

Eric L. Miller
The Communications and
Digital Signal Processing Center
Department of Electrical and Computer Engineering
235 Forsyth
Northeastern University
360 Huntington Ave.
Boston, MA 02115
Tel: (617) 373-8386
FAX: (617) 373-8627
email: elmiller@cdsp.neu.edu

Alan S. Willsky
Laboratory for Information and
Decision Systems
Department of Electrical Engineering
and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts, 02139

December 13, 1995

Abstract

In this paper we explore the utility of multiscale and statistical techniques for detecting and characterizing the structure of localized anomalies in a medium based upon observations of scattered energy obtained at the boundaries of the region of interest. Wavelet transform techniques are used to provide an efficient and physically meaningful method for modeling the non-anomalous structure of the medium under investigation. We employ decision-theoretic methods both to analyze a variety of difficulties associated with the anomaly detection problem and as the basis for an algorithm to perform anomaly detection and estimation. These methods allow for a quantitative evaluation of the manner in which the performance of the algorithms is impacted by the amplitudes, spatial sizes, and positions of anomalous areas in the overall region of interest. Given the insight provided by this work, we formulate and analyze an algorithm for determining the number, location, and magnitudes associated with a set of anomaly structures. This approach is based upon the use of a Generalized, M-ary Likelihood Ratio Test to successively subdivide the region as a means of localizing anomalous areas in both space and scale. Examples of our multiscale inversion algorithm are presented using the Born approximation of an electrical conductivity problem formulated so as to illustrate many of the features associated with similar detection problems arising in fields such as geophysical prospecting, ultrasonic imaging, and medical imaging.

*This work was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-92-J-0002, and the Advanced Research Project Agency under Air Force Grant F49620-93-1-0604 and the Office of Naval Research under Grant N00014-91-J-1004. The work of the first author was also supported in part by a US Air Force Laboratory Graduate Fellowship

1 Introduction

The goal of many applied problems is the recovery of information regarding the structure of a physical medium based upon measurements of scattered radiation collected at the boundaries [8, 15, 19, 43, 48]. For some of these tomographic-type inverse problems, one seeks a complete description (in the form of an image in two dimensions or a volumetric rendering in 3D) of the structure of the medium. In other cases, however, the full reconstruction is not needed; rather, the ultimate objective is to extract the structure of areas in the medium which are, in some sense, anomalous; that is, regions where the nature of the medium differs from some prior set of expectations. This *anomaly detection problem* arises, for example, in geophysical prospecting where in many instances the fundamental issue is the determination of oil bearing regions in the earth and medical imaging where tumor detection is of import.

As discussed in [27, 29, 31, 33, 34, 44, 45] for many of the application areas previously cited, methods for solving the anomaly detection problem typically proceed by initially generating the full, pixel-by-pixel reconstruction and subsequently post-processing the results to determine the nature of anomalous structures. The necessity of generating a solution to the so-called “full inverse problem” however makes these schemes rather unattractive. Indeed, for many interesting applications, obtaining a full reconstruction of the medium presents a collection of well-known and extensively studied challenges [2, 3, 40] which suggest that solving this problem as the first step toward localizing anomalies should be avoided. In this paper we demonstrate the utility of a multiscale framework for explicitly solving the spatial anomaly detection problem in the context of linearized inverse scattering (also known as diffraction tomography [15]) applications.

The basis for solving the anomaly detection problem is the use of wavelet transforms and the statistical theories of optimal estimation and detection to develop both efficient algorithms for anomaly detection and localization and analytical insight into the nature of the problem and the limits of performance that result from the fundamental physics relating the characteristics of the medium to the observations. In [39, 40], we introduced the use of wavelet transforms and multiresolutional statistical techniques for overcoming many of the challenges associated with the solution of full reconstruction, linearized inverse electrical conductivity problems. Many of the results in [39, 40] followed from the use of multiscale, statistical regularization methods for the incorporation of prior knowledge into the inversion routine. The use of such prior statistical models automatically implies an assumption of some type of statistical regularity on the field and therefore fails to capture adequately the presence of anomalies or localized inhomogeneities. Thus, roughly stated, the problem considered in this paper is the detection, localization, and estimation of such anomalies superimposed on a background of known statistical structure and observed indirectly through the scattering measurements.

The consideration of the anomaly detection problem raises a variety of questions beyond those arising in the full reconstruction inverse problem. How many anomalies are there? Where are they located? What are their sizes? What are their amplitudes? Given answers to the first three of these problems, the fourth is a variant of the full inverse problem in which we focus our attention on determining the magnitudes of only the previously identified anomalous regions. The determination of the number, sizes and locations of the anomalous regions is, however, a potentially daunting collection of tasks as a result of the vast number of combinations of anomaly structures which, in principle, must be explored in the generation of a solution.

Over the past decade, significant work has been performed in the area of anomaly detection from tomographic-type measurements. In [44], Rossi and Willsky were concerned primarily with the use of estimation-theoretic analysis and algorithmic methods for determining the location of

a single object of known size and structure given noisy and sparse computed tomography (CT) measurements. Recently, these results have been extended by Devaney and co-workers [16, 17, 46] in consideration of diffraction tomography (DT) and exact scattering applications. More closely related to the problem of interest in this paper is the work of Bresler, Fessler and Macovski. In [5], the authors examined a 3D reconstruction problem from CT measurements in which the first step of their algorithm required the localization of an unknown number of anomalies of unknown structure. The solution to this problem presented in [5] was to estimate the required parameters for a pre-determined, maximum number of anomalies knowing that further processing would eliminate falsely identified anomalous regions.

In this paper, we present a scale-recursive algorithm for anomaly detection and characterization given DT-type data. Here, the tools of optimal hypothesis testing are used to make a sequence of anomaly detection and localization decisions starting at coarse scales, thereby allowing for the detection of spatially large anomaly structures and providing coarse localization of finer scale anomalies, and then moving to finer ones. This algorithm is significant for two reasons. First, this approach provides a computationally efficient and accurate means of localizing areas of anomalous behavior. Second, the anomaly characterization algorithm may be viewed as a highly efficient first stage in a larger image processing application. Specifically, the output of the algorithm could be refined (for example via the methods described in [5, 44] generalized to the case of diffraction tomography) by higher level processing stages concerned with issues such as identification, classification, or imaging. Toward this end, in Section 6.3, we present one way in which knowledge of the anomaly structures can be used to supplement the information in the prior statistical model in order to improve the output of a least-squares, pixel-by-pixel reconstruction of the region of interest.

In addition to the development of the scale-recursive processing algorithm, by using these same statistical techniques, we provide analysis of the anomaly detection problem that not only yields overall performance limits, but also guides the detection procedure. For example, we are able to define and determine the statistical distinguishability of a small scale, large amplitude anomaly from a larger scale, but smaller magnitude structure or a pair of closely spaced anomalies from a single, broader anomalous region. The use of the results from this analysis can then tell us at what scale and in which regions to terminate our detection procedure, i.e. when finer scale localization is unwarranted given the available data.

In Section 2, we present an overview of the particular anomaly detection problem of interest in this work. The formal definition of the anomaly detection problem as one of optimal hypothesis testing and a review of results from statistical decision theory is provided in Section 3. In Section 4 we demonstrate the utility of our framework in characterizing the *detectability* of an anomaly. Section 5 is devoted to the question of the *distinguishability* of anomalies as a function of their relative positions and structures. In Section 6 we develop and analyze a scale-recursive algorithm for anomaly detection, localization, and estimation, and present the results of its performance under a variety of experiment conditions. Conclusions reached in this paper and directions for further work are presented in Section 7.

2 A Multiscale Framework for Inverse Scattering

2.1 The Scattering Problem

The context in which we develop our anomaly detection algorithm is a low-frequency, two-dimensional inverse electrical conductivity problem illustrated in Figure 1 and similar to problems arising in the field of geophysical prospecting [23, 24, 48] and medical imaging using electrical impedance

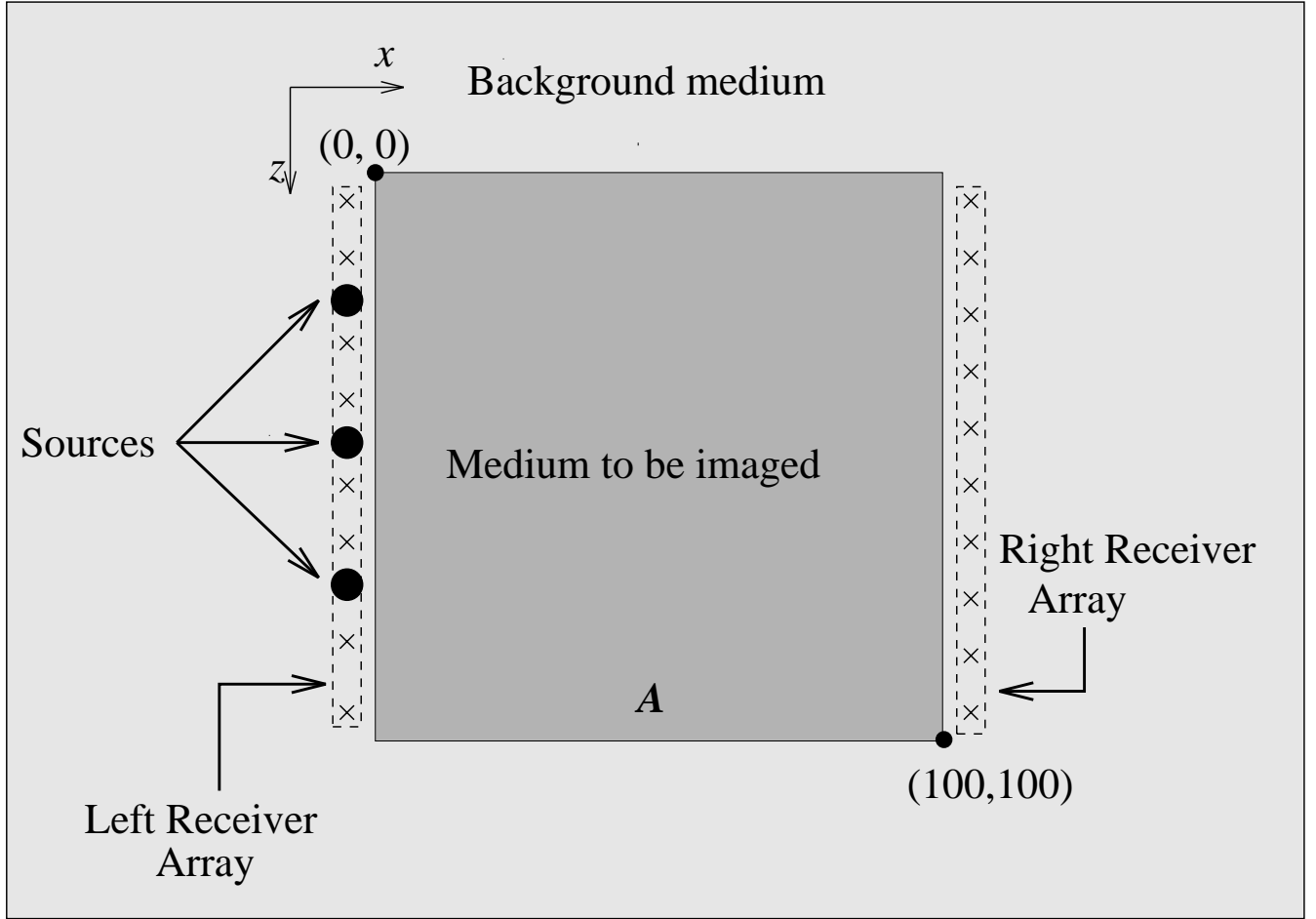


Figure 1: Configuration of inverse conductivity problem. The electromagnetic sources (indicated by the black circles) emit time-harmonic waves into a lossy medium which subsequently are scattered by conductivity inhomogeneities located in the darkly shaded rectangle, A . The secondary fields are observed at one or both receiver arrays located on either vertical edge of region under investigation. Based upon these observations, the objective of the inverse problem is the reconstruction of the conductivity perturbation.

tomography [18–20, 22, 29–31, 43]. Here, we have an array of electromagnetic line-sources oriented perpendicularly to the page emitting time-harmonic, waves into a lossy medium. The electrical properties of this environment are assumed to be decomposed into the sum of an infinite, known, and constant background and a conductivity perturbation, g , with support restricted to region A in Figure 1. The fields from the transmitters are scattered by g , and the secondary fields are observed at one or both of the receiver arrays positioned on the vertical edges of region A . Based upon these observations, the objective of the problem is to detect and localize areas in the region of interest where the structure of g is, in a sense to be defined below, anomalous.

We consider the collection of eighteen scattering experiments defined in Table 1 where each such experiment produces a vector of measurements comprised of the in-phase and quadrature

components of the observed scattered field obtained over one of the two receiver arrays due to energy put into the medium from one of the sources operating at a particular frequency. As is shown in [9], the use of the first Born approximation yields the following linear relationship between the vector of observations associated with the i^{th} scattering experiment, y_i , and a discrete representation of the two dimensional conductivity anomaly, g

$$y_i = T_i g + n_i \quad i = 1, 2, \dots, 18 \quad (1)$$

where the matrices T_i encompass the (linearized) physics and n_i is an additive, zero-mean, uncorrelated, random vector representing the noise in the data. That is, the i^{th} noise is modeled as $n_i \sim \mathcal{N}(0, r_i I)$ where I is an appropriately sized identity matrix.¹ The discrete representation of the conductivity g is constructed using the so-called “pulse” set of basis functions where the conductivity is assumed to be piecewise constant over an $N_{g,x} \times N_{g,z}$ grid of square pixels covering A [26]. For future reference, we define the “stacked” system of data

$$y = Tg + n \quad (2)$$

where $y^T = [y_1^T \ y_2^T \ \dots \ y_{18}^T]$ with T and n defined accordingly.

Experiment number	Source Position	Frequency of source (Hz)	Receiver Array
1 – 6	0:20:100	$f_{HI} = 10,000$	Left
7 – 12	0:20:100	$f_{MID} = 1,000$	Left
13 – 18	0:20:100	$f_{LO} = 100$	Right

Table 1: Data set definitions for observation processes of interest in the paper. The notation $x : y : z$ indicated that the sources are distributed in y increments along a line from x to z .

2.2 A Multiscale Representation of the Problem

The detection techniques developed in Sections 4 – 6 are based upon a linear model relating multiresolution representations of g and n_i to a multiresolution representation the data, y_i . A scale-space representation of the problem has been chosen for two reasons. First, the matrices T_i in (1) are of the class which are made sparse in the wavelet transform domain [1, 4] thereby lowering the computational complexity of the detection algorithm in Section 6. Although not considered extensively in this work, such computational benefits are explored in [41]. Second, as we discuss below, a collection of useful and physically meaningful models for the non-anomalous behavior of the conductivity field are specified easily in the wavelet domain.

Following the work in [39, 40], orthonormal, discrete wavelet transform (DWT) [14] operators (matrices) \mathcal{W}_i and \mathcal{W}_g are used to move from physical to scale space in the following manner

$$\eta_i = \mathcal{W}_i y_i = (\mathcal{W}_i T_i \mathcal{W}_g^T)(\mathcal{W}_g g) + \mathcal{W}_i n_i \equiv \Theta_i \gamma + \nu_i \quad (3)$$

where $\mathcal{W}_g^T \mathcal{W}_g = \mathcal{W}_i^T \mathcal{W}_i = I$ follows from the orthonormality of the wavelet transformation [14, 35]. There are a variety of reasons why we may wish to use different transforms for the data than for g . First, from Figure 1, each data set is to be collected over a 1D array of receivers. Hence, \mathcal{W}_i will act on a one dimensional signal while \mathcal{W}_g is used to transform the 2D conductivity profile. Additionally, it may be the case that the lengths of each data record vary from one observation process to the next. Finally, analogously to the physical space case, we define the stacked systems

$$\eta = \Theta \gamma + \nu \quad (4)$$

where $\eta = [\eta_1^T \ \eta_2^T \ \dots \ \eta_{18}^T]^T$, Θ and ν are defined analogously and $\nu \sim \mathcal{N}(0, R)$ with $R = \text{diag}(r_1 I, r_2 I, \dots, r_{18} I)$.

¹The notation $x \sim \mathcal{N}(m, P)$ indicates that the random vector x has a Gaussian probability distribution with mean m and covariance matrix P .

2.3 Multiscale Prior Models

Recently there has been significant work in the use of fractal models for describing the spatial distribution of geophysical quantities. In [13], Crossley and Jensen explore the propagation of acoustic radiation in the Earth's crust using a velocity model composed of the sum of a deterministic profile and a fractal perturbation. In considering the distribution of hydraulic conductivity, Brewer and Wheatcraft [6] employ a wavelet-based model very similar to the one described below as a means of interpolating coarse scale observations of hydraulic conductivity to finer scales. Brown [7] relates both the electrical and hydraulic conductivities in the earth to a self-similar model for the height distribution in rock fractures and studies the resulting fluid and current flow patterns through such a formation. Finally, the propagation of electromagnetic radiation through media with fractal characteristics has been studied extensively by Jaggard and co-workers [32].

With this work as motivation, we use a stochastic, fractal-type model to describe the spatial distribution of the electrical conductivity in the absence of anomalies. While there are many self-similar models which may be used to describe the conductivity, results of Wornell [50], Tewfik [47], and Chou et al. [10–12] suggest that there exist a wide range of statistical models specified *directly* in the wavelet transform domain possessing the desired modeling characteristics and simple structures thereby making them quite attractive for use in signal and image processing applications.

Under the particular wavelet-based model of interest in this paper, the wavelet coefficients of the non-anomalous conductivity field, denoted by the vector $\tilde{\gamma}$, are taken to be uncorrelated, Gaussian random variables. That is, $\tilde{\gamma}$ is distributed according to

$$\tilde{\gamma} \sim \mathcal{N}(0, P_0) \quad (5)$$

where P_0 is a *diagonal* matrix whose nonzero entries are the variances of the corresponding wavelet coefficients. While a detailed description of the internal structure of P_0 is presented in [35, 50], the fractal-type behavior of the process is obtained by taking the variance of the wavelet coefficients to vary exponentially with scale. Coefficients in $\tilde{\gamma}$ governing the coarsest scale behavior of the conductivity have relatively large variances while fine scale components possess smaller variances.

3 Anomaly Detection as a Hypothesis Testing Problem

3.1 A Model for the Conductivity

The objective of the anomaly detection problem is to determine those areas in A where the behavior of g is anomalous in that in these regions g differs from some prior set of beliefs regarding the manner in which the conductivity is expected to behave. Thus, the conductivity g is decomposed as

$$g = \tilde{g} + \bar{g} \quad (6)$$

where \tilde{g} represent that portion of g consistent with our prior assumptions and \bar{g} encompasses the anomalous behavior of the conductivity; that is, the perturbation of the conductivity away from its non-anomalous structure. In the wavelet transform domain, (6) takes the form

$$\gamma = \mathcal{W}_g g = \mathcal{W}_g \tilde{g} + \mathcal{W}_g \bar{g} \equiv \tilde{\gamma} + \bar{\gamma}. \quad (7)$$

As will be seen in Sections 4–6, considerable insight into the anomaly detection problem is obtained through performance analysis carried out using anomaly structures of varying sizes (i.e. spatial scales) located in different regions of A . Also, the primary intent of the detection algorithm presented in Section 6 is to localize quickly and efficiently regions where anomalies are suspected to exist. As region A is pixelated into an $N_{g,x} \times N_{g,z}$ grid and because we perform anomaly localization through a process of spatial subdivision, we are lead naturally to consider a representation in which anomalous regions are defined to be superpositions of rectangular subsets of A .

Referring to Figure 2, the structure of the i^{th} anomaly in A is defined by its magnitude, a_i , its size, and its location in A . The area of an anomaly defines its *scale* in that small scale anomalies are

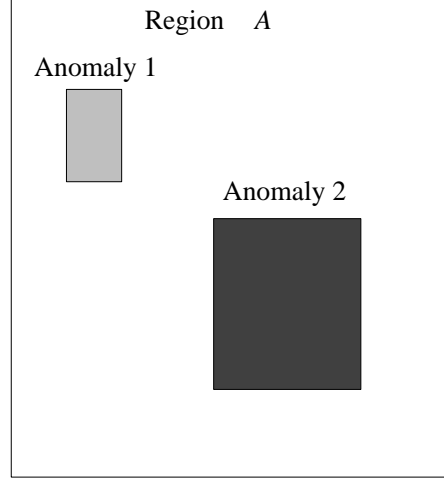


Figure 2: General structure of anomalous regions of interest in this paper. The magnitudes, a_1 and a_2 of the two anomalies shown here are proportional to the color of the corresponding rectangles.

correspondingly small in area and similarly for larger scale anomaly structures. Mathematically, the form for the anomalous behavior of the conductivity over the region A is

$$\bar{g} = \sum_{j=1}^{N_a} b_j a_j = Ba. \quad (8)$$

Here, N_a is the number of anomalous regions located in A , a_j is a scalar defining the magnitude of each anomaly, and b_j represents the discrete indicator function over the j^{th} rectangular region in \bar{g} . In (8), the column vector a represents the collection of anomaly amplitude coefficients while B is the matrix whose j^{th} column is b_j . In the wavelet transform domain, (8) is written as

$$\bar{\gamma} = \sum_{j=1}^{N_a} (\mathcal{W}_g b_j) a_j \equiv Ba \quad (9)$$

where $B = [\mathcal{W}_g b_1 \ \mathcal{W}_g b_2 \ \dots \ \mathcal{W}_g b_{N_a}]$. Finally, use of (7) and (9) in (4) yields the following relationship among the anomaly structures, the non-anomalous background \tilde{g} or $\tilde{\gamma}$, and the data

$$\eta = \Theta \bar{\gamma} + \Theta \tilde{\gamma} + \nu = \Theta Ba + \Theta \tilde{\gamma} + \nu \quad (10)$$

where, because $\tilde{\gamma}$ and ν are taken to be uncorrelated,

$$P_\eta = E[\eta \eta^T] - E[\eta]E[\eta^T] = \Theta P_0 \Theta^T + R. \quad (11)$$

Note that the analysis methods and algorithmic techniques presented in this work are based entirely on an observation model of the form in (10). In particular, the results in this paper are not dependent upon the assumption of rectangular anomalies; rather structures with arbitrary shapes and orientations can be employed in principle through the appropriate specification of the matrix B . Nonetheless, as will be seen in Sections 4–6 of this paper, rectangular structures prove to be highly useful for obtaining significant insight into the nature of the anomaly characterization problem and as the basis for an algorithm designed to extract this information from observed scattered fields.

To provide a normalized notion of the overall size of an anomaly, we define an SNR-type quantity called the *anomaly-to-background ratio* (ABR) which provides a measure of the energy in an anomaly relative to that of \tilde{g} . Mathematically, we have for an anomaly \bar{g} composed of a single rectangular region defined by the column vector b and with amplitude a

$$ABR^2 = \frac{\text{Power in } \bar{g}}{\text{Expected power in } \tilde{g}} = \frac{a^2 (b^T b)}{\text{tr}(\bar{P}_0)} \quad (12)$$

where $\text{tr}(M)$ is the trace of the matrix M and $\bar{P}_0 = \mathcal{W}_g^T P_0 \mathcal{W}_g$ is the covariance matrix of \tilde{g} .

As described in [40], under the Born approximation used to obtain (1), $g = \mathcal{W}_g^T \gamma$ represents a perturbation about a known, constant background conductivity, g_0 . From physical principles, the overall conductivity, $g_0 + g = g_0 + \tilde{g} + Ba$ must be greater than zero. Thus, in theory the elements of a may assume both positive as well as negative values so long as the positivity constraint is satisfied. To simplify matters, in this paper we assume that the a_i are strictly greater than zero corresponding to regions of locally higher conductivity than the background.

3.2 The M-ary Hypothesis Testing Problem

In Section 6, we consider a statistical decision-theoretic methodology for reconstructing $\bar{\gamma}$ which is based upon a sequence of M-ary Generalized Likelihood Ratio Tests (GLRT) as a means of localizing an unknown number of anomalous regions in A . The mathematical description of each such test begins with the formulation of the following M hypotheses, H_i for $i = 0, 1, 2, \dots, M-1$, corresponding to M different configurations of anomalous areas

$$H_i : \quad \eta = \Theta \mathcal{B}_i a_i + \Theta \tilde{\gamma} + \nu \quad i = 0, 1, 2, \dots, M-1. \quad (13)$$

Note that from (13) under H_i we have, $\eta \sim \mathcal{N}(\Theta \mathcal{B}_i a_i, P_\eta)$ where P_η is given by (11).

The hypothesis test is implemented as a rule which when given the data, indicates which of the H_i is true. Because it will be the case in Section 6 that the a_i are taken to be deterministic but unknown parameters, a standard likelihood ratio test solution to the hypothesis testing problem [49] cannot be employed in this context. Rather, we use a Generalized Likelihood Ratio Test (GLRT) [49] for performing the test. This procedure requires first that an estimate of each a_i be computed assuming that H_i is correct. As this problem is, in general, ill-posed, we choose here to use the following regularized, least squares estimate

$$\hat{a}_i = (\mathcal{B}_i^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_i + \alpha I)^{-1} \mathcal{B}_i^T \Theta^T P_\eta^{-1} \eta. \quad (14)$$

where the parameter α is used to control the degree of regularization.

Given \hat{a}_i , the hypothesis testing rule employed in this paper is

$$\text{Choose } H_i \text{ with } i = \begin{cases} 0 & \max_j L_j(\eta) < 0 \\ \arg \max_j L_j(\eta) & \text{otherwise} \end{cases} \quad (15)$$

where

$$L_j(\eta) = l_j(\eta) - l_0(\eta) \quad j = 1, 2, \dots, M-1 \quad (16)$$

and for $j = 0, 1, 2, \dots, M-1$

$$l_j(\eta) = \eta^T P_\eta^{-1} \Theta \mathcal{B}_j \hat{a}_j - \frac{1}{2} \hat{a}_j^T \mathcal{B}_j^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_j \hat{a}_j. \quad (17)$$

3.3 The Binary Hypothesis Testing Case

While the algorithm for extracting anomaly information is based upon the M-ary GLRT, much of the analysis of the anomaly detection problem is performed in the context of the *binary hypothesis testing* (BHT) framework in which two alternatives, $\bar{\gamma}_0 = \mathcal{B}_0 a_0$ and $\bar{\gamma}_1 = \mathcal{B}_1 a_1$, are compared.² Traditionally, the analysis of the BHT centers around the probability of detection, P_d and the false alarm probability, P_f . For the linear-Gaussian model considered in this work, it is shown in [49] that P_d and P_f are related to the various quantities defining the structure of the problem via

$$d = \text{erfc}_*^{-1}(P_f) - \text{erfc}_*^{-1}(P_d) \quad (18)$$

²Note that in the contexts where the binary testing scenario is to be explored, the values of a_0 and a_1 are assumed known so that a generalized test is not required.

where

$$d^2 = (\bar{\gamma}_1 - \bar{\gamma}_0)^T \Theta^T P_\eta^{-1} \Theta (\bar{\gamma}_1 - \bar{\gamma}_0) \quad (19)$$

$$\text{erfc}_*(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt. \quad (20)$$

Thus, based upon (18), we see that our ability to distinguish between two anomaly structures is intimately related to the Fisher discriminant, d , which has the interpretation of a “signal-to-noise” ratio [49]. Note that for a given P_f , larger d results in larger P_d and therefore better performance.

From (19) we observe that the performance of the binary hypothesis test is a function of both the geometric configurations, as captured in the matrices \mathcal{B}_i , and the magnitudes, a_i , of the two candidate anomaly structures. To better understand the role of these two factors, consider the case in which $\bar{\gamma}_i$ corresponds to a single rectangular region so that each \mathcal{B}_i is a column vector and each a_i is a scalar. Substituting (9) into (19) and expanding the quadratic yields

$$\delta_1^2 a_1^2 - 2\delta_{1,0} a_1 a_0 + \delta_0^2 a_0^2 - (\text{erfc}_*^{-1}(P_f) - \text{erfc}_*^{-1}(P_d)) = 0 \quad (21)$$

where

$$\delta_j^2 = \mathcal{B}_j^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_j \quad \text{for } j = 0, 1 \quad (22)$$

$$\delta_{1,0} = \mathcal{B}_1^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_0. \quad (23)$$

In [37], it is shown that when viewed as a function of a_0 and a_1 , (21) defines an ellipse the form of which is illustrated in Figure 3.³ This ellipse indicates that, given the geometry of the candidate anomalies, \mathcal{B}_0 and \mathcal{B}_1 , there are only certain combinations of a_0 and a_1 which will result in performance below that level dictated by a particular P_d and P_f . In fact, these points are precisely those that lie inside the plotted ellipse. Also, there exists a minimum level, $a_{1,0}^{\min}$ (depending on the geometric structures of *both* anomalies) such that for $\bar{\gamma}_1 = \mathcal{B}_1 a_1$ with $a_1 > a_{1,0}^{\min}$, the binary hypothesis test will achieve or exceed the P_d and P_f performance figures *independent* of a_0 .

4 Detectability Analysis

The first issue we address in conjunction with the anomaly detection problem is that of the detectability of an anomaly as a function of location, spatial size, and amplitude. After defining a particular collection of anomaly structures, a set of binary hypothesis testing problems are explored in which H_0 corresponds to there being no anomaly in the region while under H_1 , a particular member of our anomaly collection is assumed to be present. The objective of the detectability analysis is to determine the minimum magnitude each such structure must possess to guarantee a prespecified level of performance from the binary hypothesis test.

Detectability is of interest due to the physics governing the relationship between the observations, η , and the conductivity, γ and the constrained experimental conditions in which data are collected only along the vertical edges of A . From these facts, it is not expected that arbitrarily small (in scale and magnitude) anomalies will be detectable with arbitrary precision throughout A . Rather, we anticipate that small anomalies should be readily detected only close to the observation points while interior to region A small scale structures would require significantly larger magnitudes to be as detectable as their counterparts closer to the edges.

With this intuition in mind, we consider a family of anomaly structures generated by a set of dyadic tessellations of A . For example, with $N_{g,x} = N_{g,z} \equiv N_g = 16$, we take as \mathcal{J}_1 the set of N_g^2 indicator functions which are one over single pixels in A and zero elsewhere. Analogously, \mathcal{J}_2 is the collection of $N_g^2/4$ characteristic functions over disjoint 2×2 sized regions of A . Thus, in general

³For illustrative purposes only, in Figure 3 it is assumed that the major axis of the ellipse is oriented at an angle less than 90° from the a_0 axis. While this is not necessarily the case, the analysis which follows is independent of which axis is the major and which the minor.

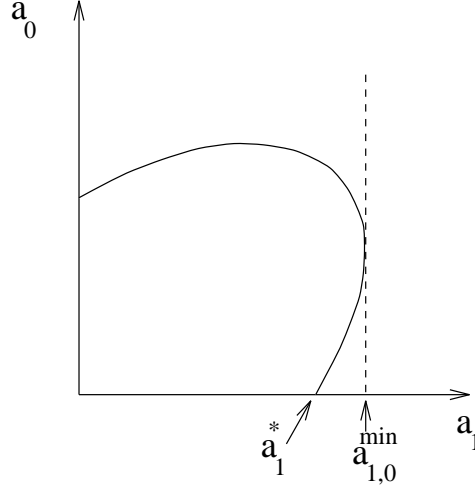


Figure 3: The structure of the ellipse defined by (21). The axes represent the magnitudes of anomaly structures in a binary hypothesis testing problem. As discussed in Section 2, a_0 and a_1 are taken to be nonnegative so that only the first quadrant is shown in this illustration. Here a_1^* is the minimum amplitude of $\tilde{\gamma}_1$ required to detect this structure when the alternate hypothesis is $\tilde{\gamma}_0 = 0$ for a BHT with prespecified P_d and P_f . The value $a_{1,0}^{min}$ is the minimum intensity of $\tilde{\gamma}_1$ required to ensure that for *any* $\tilde{\gamma}_0$ the performance of the resulting BHT meets or exceeds that defined by P_d and P_f .

\mathcal{J}_m (for m an integral power of 2) is the set of $(N_g/m)^2$ non-overlapping square regions of size $m \times m$ completely covering A . Finally, we define \mathcal{J} as the union of all \mathcal{J}_m .

To begin our analysis of detectability, for each anomaly structure in \mathcal{J} , we consider a collection of binary hypothesis testing problems where the two hypotheses in the j^{th} problem correspond to the situations in which no anomaly is present in A or a scaled version of the j^{th} element of \mathcal{J} is in A . Recalling (13), these alternatives take the form

$$H_0 : \eta = \Theta\tilde{\gamma} + \nu \quad (24a)$$

$$H_{1,j} : \eta = \Theta\mathcal{B}_j a_j + \Theta\tilde{\gamma} + \nu. \quad (24b)$$

The goal of our detectability analysis then is to determine for each anomaly structure in \mathcal{J} , the minimum value of a_j , denoted a_j^* , such that the above hypothesis test attains a certain level of performance as specified by P_d and P_f .

The primary quantity used to characterize the performance of the binary hypothesis test in (24a)–(24b) is the Fisher discriminant discussed in the previous section which here takes the form

$$d_j^2 = a_j^2 (\mathcal{B}_j^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_j) \equiv a_j^2 \delta_j^2 \quad (25)$$

where δ_j^2 is defined in (22) and represents the Fisher discriminant for the unit amplitude anomaly over the j^{th} member of \mathcal{J} . Now, for a given P_d and P_f , (18) and (25) are combined to give the following expression for a_j^* :

$$a_j^* = \frac{\text{erfc}^{-1}(P_f) - \text{erfc}^{-1}(P_d)}{\delta_j}. \quad (26)$$

In Figure 4, a_j^* are plotted for all anomalies in \mathcal{J} for the case in which data from the 18 experiments described in Table 1 at an SNR of 10 are available and where P_d is set to 0.95 and P_f is 0.05. In this work, the SNR associated with the anomaly-free observation process $\eta_i = \Theta_i \tilde{\gamma} + \nu_i$ with $\nu_i \sim (0, r_i^2 I)$ and $\gamma \sim (0, P_0)$ is defined as

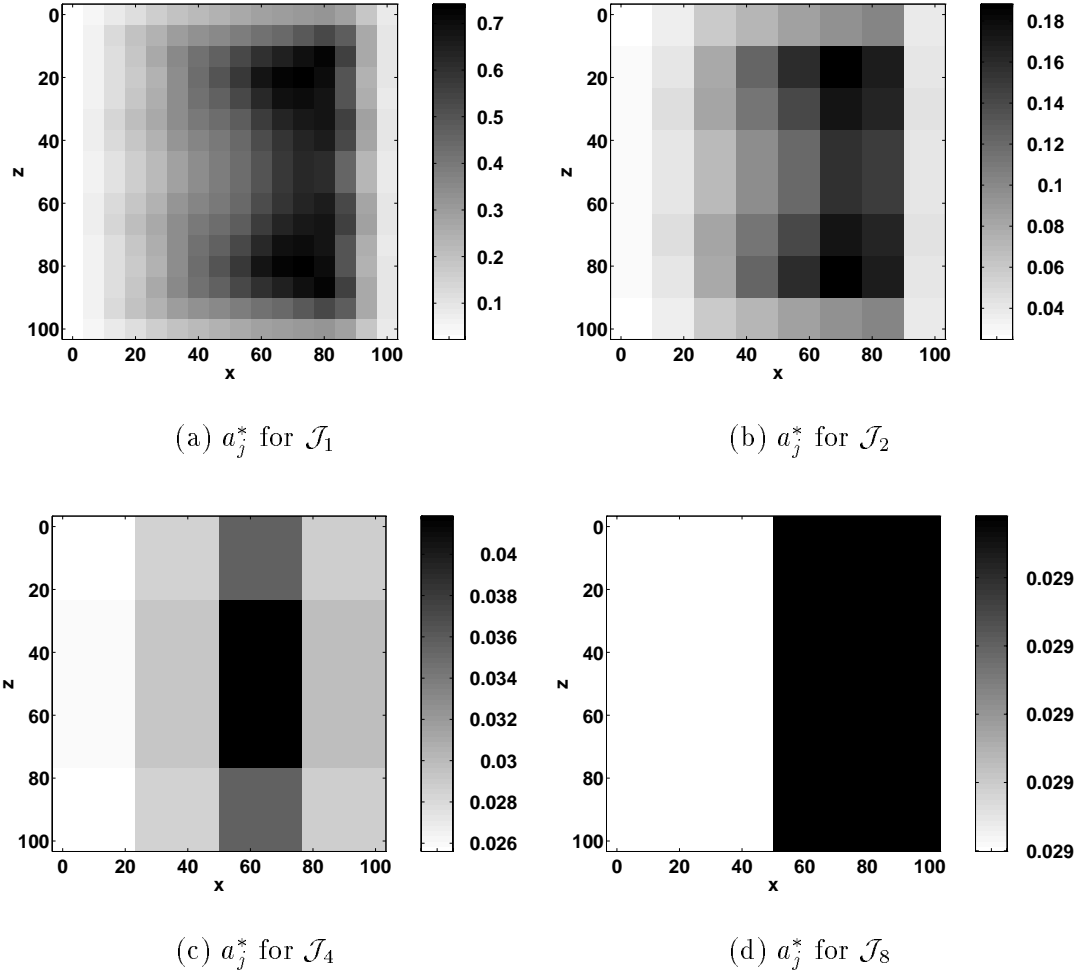


Figure 4: Value of a_j^* for all anomaly structures in \mathcal{J} where the data from the experiments described in Table 1 at an SNR of 10 are used as input to the likelihood ratio test. Here, we have $P_d = 0.95$ and $P_f = 0.05$. Note that the scales in these images are all different with a^* decreasing significantly as the size of the anomalies increases.

$$SNR_i^2 = \frac{\text{Power per pixel in } \Theta_i \gamma}{\text{Power per pixel in } \nu_i} = \frac{\text{tr}(\Theta_i P_0 \Theta_i^T)}{N_g r_i^2}. \quad (27)$$

Thus, each 1×1 pixel in Figure 4(a) corresponds to an anomaly in \mathcal{J}_1 with the intensity of that pixel proportional to a_j^* . In all four cases, we see that near the middle of the region, the magnitude required to obtain the desired level of performance in the binary hypothesis test is significantly larger than that required near the vertical edges i.e. where the sources and receivers are located. For vertical values roughly in the range $40 \leq z \leq 60$, this effect is somewhat smaller. Also, as the areas of the anomalies increase, the required magnitudes decrease. This coincides with the intuition that large scale structures should be easier to detect than their fine scale counterparts. Finally, the ABR values in Figure 4 are quite small with the median values all less than 0.9. This implies that our statistical approach toward anomaly detection should prove quite advantageous in detecting relatively small amplitude conductivity anomalies.

To explain the behavior of a_j^* , we note that as described in [40] the low and medium frequency

kernels are most sensitive to the conductivity structure over the horizontal range $0 \leq x \leq 50$ so that the required magnitude for an anomaly to be “seen” in this area should be relatively low. The smaller values of a_j^* in the region $40 \leq z \leq 60$ are due primarily to the combined coverage of this region by more observation kernels, T_i , than is the case for the top and bottom edges.

5 Distinguishability Analysis

In this section, we explore issues associated with our ability to successfully distinguish between pairs of candidate anomalies in order to obtain quantitative insight into the ambiguity which exists in attempting to differentiate between anomalous structures of differing sizes, locations, and magnitudes. The results of this work then are used both in the formulation as well as the analysis of the detection algorithm in Section 6.

Before proceeding with the analysis, we note that the issue of distinguishability has been considered previously in the context of electrical impedance tomography [18, 22, 29]. In that work, distinguishability was examined in a deterministic setting where observation perturbation was modeled as a bounded but otherwise unknown signal. Under such a model, two conductivity profiles were defined to be distinguishable if the norm of the difference between the data sets produced by each exceeded the noise level. The notion of distinguishability developed below is rather different as it rests upon a statistical model for both the additive measurement noise and background perturbations in the medium’s conductivity.

The mathematical formulation of the distinguishability problem of interest in this work follows directly from Section 3.3. We begin by considering the following binary hypothesis testing problem

$$H_j : \eta = \Theta \mathcal{B}_j a_j + \Theta \tilde{\gamma} + \nu \quad (28a)$$

$$H_i : \eta = \Theta \mathcal{B}_i a_i + \Theta \tilde{\gamma} + \nu. \quad (28b)$$

The primary tool for our distinguishability analysis is the quantity $a_{i,j}^{min}$ defined in Section 3.3 to be the smallest value of a_i such that the performance of the binary hypothesis test in (28a)–(28b) meets or exceeds that defined by $P_{d,i,j}$ and $P_{f,i,j}$ independent of the amplitude of a_j . Finally, for all experiments and for all i and j of interest in this section, $P_{d,i,j}$ is equal to 0.95 and $P_{f,i,j} = 0.05$.

In Figures 6, $a_{i,j}^{min}$ is shown as a function of $j \in \mathcal{J}$ in the case where the geometric structure of anomaly $\tilde{\gamma}_i$ is given in Figure 5(a). Similarly, $a_{i,j}^{min}$ is displayed for the anomaly geometry of Figure 5(b) in Figure 7. Essentially these two examples demonstrate the manner in which the ability to differentiate structures is dependent upon the spatial position of the anomalies in region A . In both cases, we see that the largest values of $a_{i,j}^{min}$ are associated with hypothesis tests in which $\tilde{\gamma}_i$ is compared to a second, relatively close-by anomaly structure; however, these amplitudes are roughly twice as large for the structure located toward the middle of the region than for the anomaly closer to the source/receiver arrays.

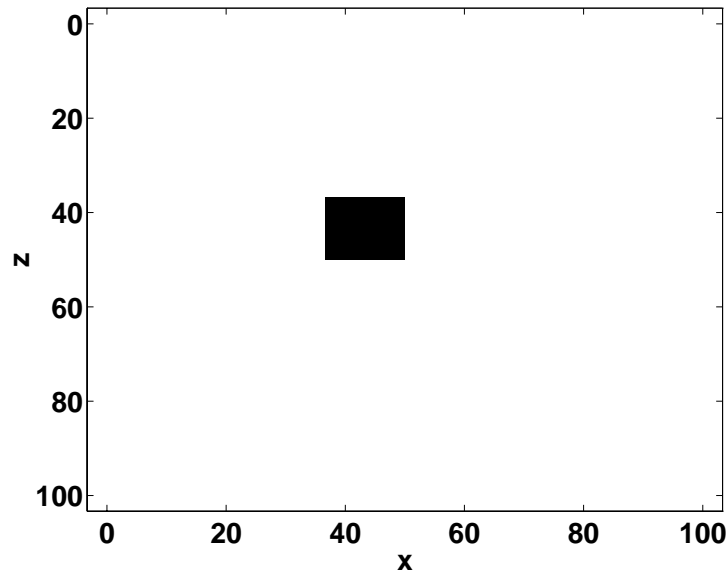
In Table 2, the ABRs corresponding to the largest and smallest values for $a_{i,j}^{min}$ in Figures 6 and 7 are shown. That is for i fixed, the entries in the first column of Table 2 are the anomaly-to-background ratios generated by

$$a_i^{max,min} = \max_j a_{i,j}^{min}$$

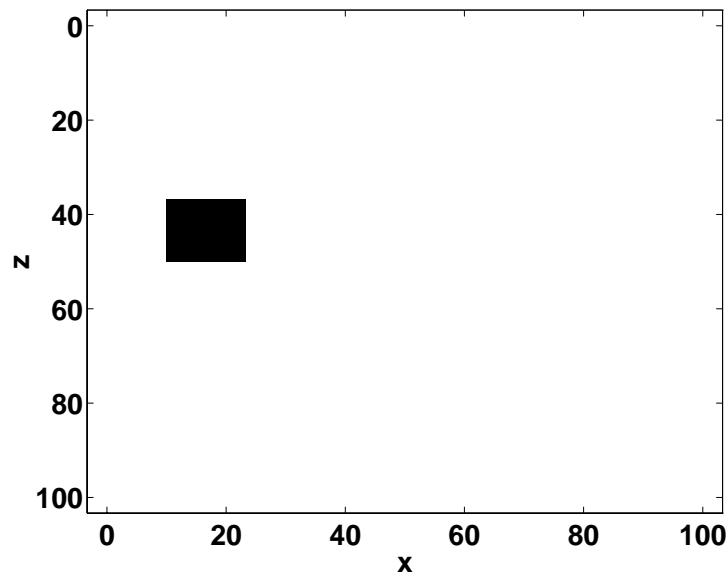
while those of the second column are associated with

$$a_i^{min,min} = \min_j a_{i,j}^{min}.$$

Note that if a_i is greater than $a_i^{max,min}$, a BHT with the anomaly $\tilde{\gamma}_i$ given by $\mathcal{B}_i a_i$ will meet the $P_{d,i,j}$ and $P_{f,i,j}$ specification regardless of both the amplitude as well as the location of γ_j , i.e. the performance will be independent of j . On the other hand if a_i is less than $a_i^{min,min}$ then for every j there will be some range of amplitudes a_j for which the performance specifications will



(a) First anomaly structure to be analyzed in distinguishability problems



(b) Second anomaly structure to be analyzed in distinguishability problems

Figure 5: Anomaly structures to be analyzed in distinguishability problems

not be achieved. Now, from the first row of Table 2, we see that for an anomaly with geometric structure in Figure 5(a), an ABR of 2.11 ensures that any binary hypothesis test in which this structure is compared to a member of \mathcal{J} will meet the performance specifications of $P_{d,i,j} = 0.95$ and $P_{f,i,j} = 0.05$. Alternatively, if the ABR falls below 0.56 then for all structures in \mathcal{J} , (i.e. all

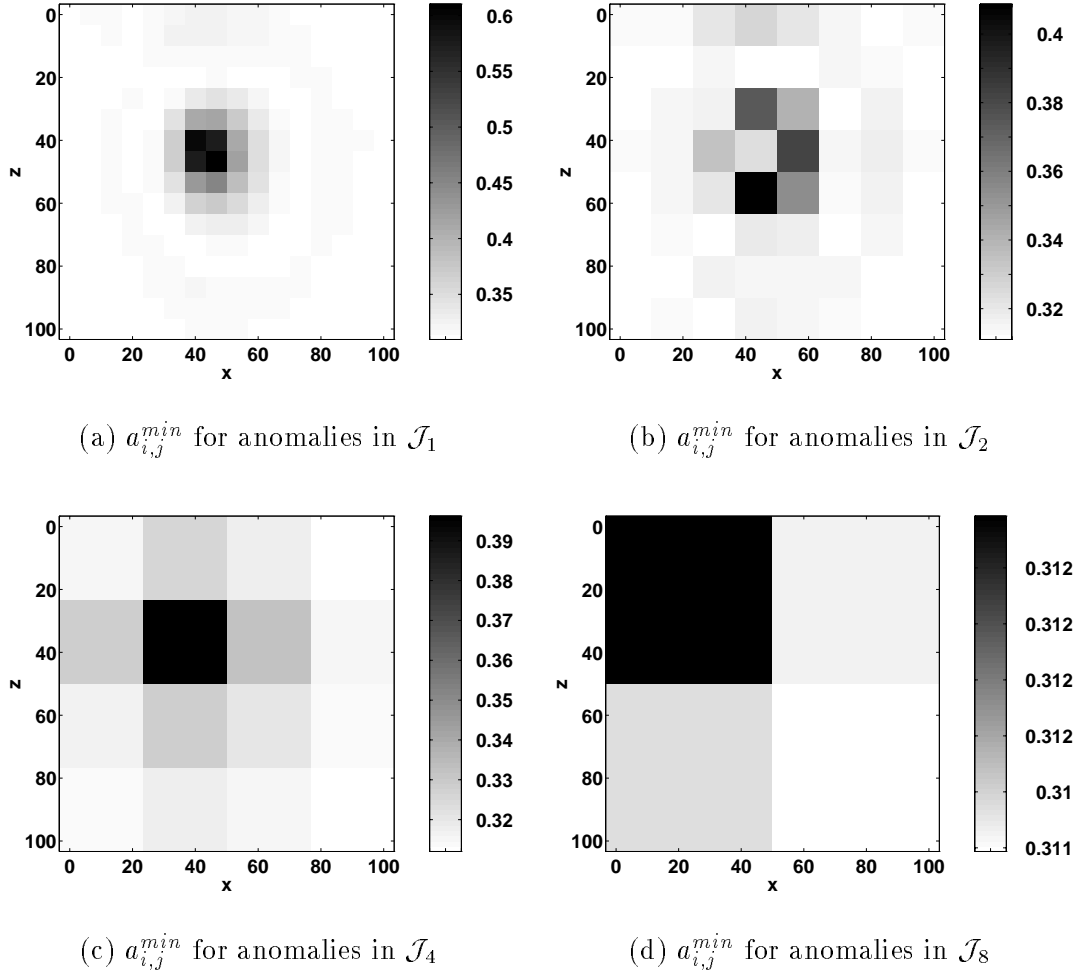


Figure 6: Images of the minimum magnitude of the anomaly in Figure 5(a) to guarantee a $P_d = 0.95$ and $P_f = 0.05$ in binary hypothesis tests involving this anomaly structure and elements of \mathcal{J} . Note that while the scales in these images are different the magnitudes are all less than 2.5.

\mathcal{B}_j) the performance of the BHT will fail to meet the $P_{d,i,j}$ and $P_{f,i,j}$ requirements for some range of a_j . Similar results hold for the second anomaly structure located closer to the left side except that in this case, the required values of the ABR are smaller.

Anomaly $\tilde{\gamma}_i$	Maximum ABR	Minimum ABR
Rightmost (Figure 5(a))	0.49	0.24
Leftmost (Figure 5(b))	2.11	0.56

Table 2: Minimum and maximum anomaly-to-background ratio associated with the smallest and largest values for $a_{i,j}^{min}$ for the anomaly structures in Figure 5(a) and 5(b).

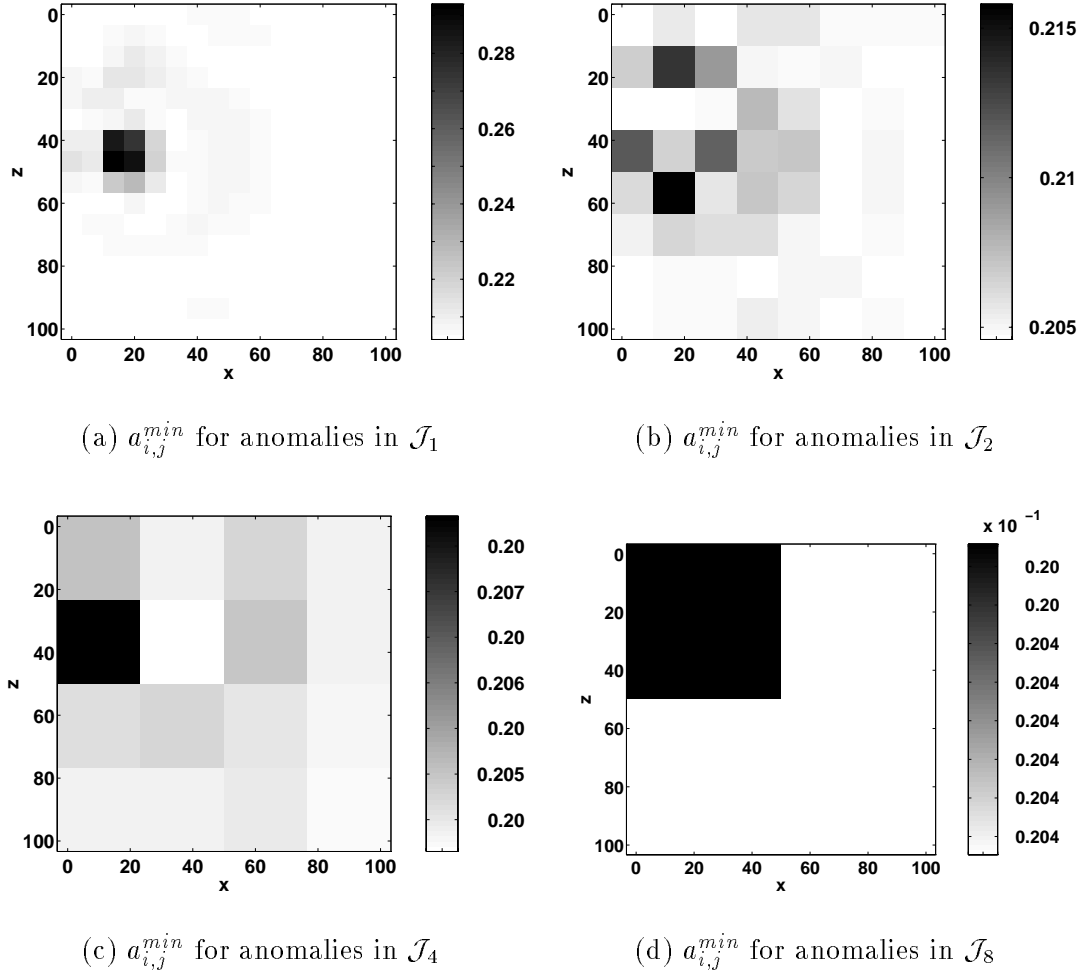


Figure 7: Images of the minimum magnitude of the anomaly in Figure 5(b) to guarantee a $P_d = 0.95$ and $P_f = 0.05$ in binary hypothesis tests involving this anomaly structure and elements of \mathcal{J} . Again, the scales in (a) through (d) are all different; however the overall range of values is between 0.9 and 1.3.

6 A Multiscale Algorithm for Anomaly Characterization

In this section we describe and analyze a multiscale, decision-theoretic algorithm to determine the positions, sizes and magnitudes of an unknown number of anomalous structures in region A . We begin with a small collection of relatively large rectangular areas in which anomalies *may* be located. Each region represents a top-level node in a tree of finer-scale subdivisions of A . We next use a decision-directed procedure for determining how best to move from one level of the tree, corresponding to a collection of coarse-scale hypotheses, to the next level in which anomalies are better localized using smaller-scale rectangles. The result of this procedure is a collection of rectangular areas of varying sizes and positions where we believe anomalies exist. To limit the number of targeted areas which contain no anomalies, the algorithm concludes with a pruning step where we also estimate the magnitudes of the final group of chosen anomaly structures.

6.1 A Scale Recursive, Decision Driven Detection Algorithm

The first step in our detection algorithm involves an M-ary Hypothesis test in which we consider 10 ways to subdivide A in order to better localize anomalous structures. As seen in Figure 8 the first configuration corresponds to the presence of a coarse scale anomaly with support over all of A . This particular structure indicates that *no* further decomposition is warranted. The next four possibilities each allows for a single anomaly localized to the top, bottom, left and right halves of A respectively. Because anomalies might lie both in the left/right as well as the top/bottom halves, the sixth and seventh structures in Figure 8 are included. Since multiple anomalies may be present in the region, the eighth configuration corresponds to the presence of one anomaly located in the left half and one in the right while the ninth presents the analogous situation but for the top and the bottom. Finally, for this initial decomposition only, we consider the last case where we conjecture that *no* anomalous regions exists in A .

Given the 10 choices in Figure 8, we formulate a 10-ary hypothesis testing problem the solution of which is obtained using the Generalized Likelihood Ratio Test (GLRT) discussed in Section 3.2. Using (17) we compute the values of the generalized log-likelihood function for each of the hypotheses under consideration. From Figure 8, if H_0 is chosen, no further decomposition occurs and we conclude that there is a single anomaly covering the entire region of interest. If H_9 is selected, the algorithm terminates with the conclusion that there is no anomaly in region A . Otherwise, we decompose that hypothesis with the largest generalized log-likelihood value.

Our scale-recursive decomposition of A continues by essentially repeating the hypothesis testing procedure for each of the subregions indicated by the initial 10-ary hypothesis test as being of interest. For example, consider the case where H_3 is chosen. Referring to Figure 8, this selection corresponds to an anomaly located in the left half of A . In an effort to better localize the anomalous activity in this region, we consider an M-ary hypothesis test similar to that described in the previous paragraph but where the underlying area involved in the decomposition is now the left half of A rather than all of A . While the subdivision is of a rectangular region as opposed to a square area, the form of the hypotheses fundamentally remains the same as in those displayed in Figure 8 in that we consider the possibilities of anomalies located in the top, bottom, left, and right halves, etc. of this long and thin structure. We note that the first of these nine hypotheses, H_0 , corresponds to the case where no further decomposition of the left half is warranted and thus serves as a means of terminating the scale recursive search over this region of A . Instead of ten, there are only nine hypotheses as we no longer include the possibility that no anomaly exists in the left half of A since the previous iteration indicated that *somewhere* in the left side there exists an anomaly.

This nine-hypothesis GLRT is repeated recursively beginning with the regions selected in the initial decomposition of A . This decision-theoretic localization process continues until no further subdivision in a particular region is warranted based upon the selection of the H_0 hypothesis at some stage of the process or because no addition refinement is possible because the structures under consideration are too small. Thus at the end of our scale-recursive decomposition of A we have a collection of rectangular regions where anomalous structures are likely to exist. We then collect the wavelet-domain representations of these rectangles as columns in a matrix labeled \mathcal{B}_{leaf} .

To limit the number of false alarms generated by our detection algorithm, we retain only those structures in \mathcal{B}_{leaf} corresponding to sufficiently “detectable” anomalies. Specifically, we begin computing \hat{a}_{leaf} , the amplitudes associated with \mathcal{B}_{leaf} , using (14) with \mathcal{B}_i replaced by \mathcal{B}_{leaf} . Next, for each column of \mathcal{B}_{leaf} , we calculate the minimum required amplitude to guarantee a set level of performance from a detectability-type hypothesis test developed in Section 4 (here we use $P_d = 0.80$ and $P_f = 0.10$). The final estimated anomaly structure generated by our algorithm is composed of

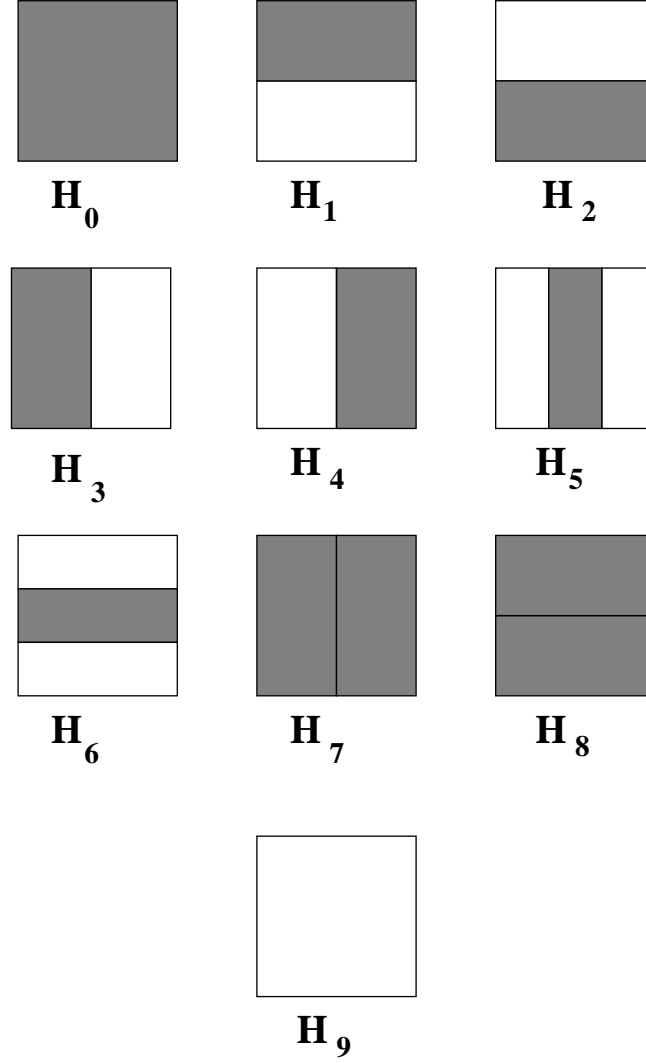


Figure 8: Geometric structures of nine possible decompositions used at each stage of our decomposition of A . The darkly shaded regions indicate the areas where anomalous structures are hypothesized to exist. While the figure illustrates the decomposition of a square region, analogous subdivision schemes are used for rectangular areas as well with the fundamental idea being the presence of anomalies in the top, bottom, left, right, etc.

those columns of \mathcal{B}_{leaf} and elements of a_{leaf} corresponding to anomalies whose amplitudes exceed this required minimum and we label these estimates $\hat{\mathcal{B}}$ and \hat{a} respectively.

6.2 Algorithm Analysis

The scale-recursive detection algorithm described in Section 6.1 requires that we be able to identify successfully large-scale structures covering the true, smaller-scale anomalies. The results of the distinguishability analysis suggest that the correct large-scale structures are likely to be selected. Indeed, Figures 6 and 7 showed that the largest values of $a_{i,j}^{min}$ corresponded to those j in \mathcal{J} which overlap anomaly i . From this, we conclude that small-scale anomalies “look” most like those large-scale counterparts located in the same region of A .

To further verify this intuition, we undertake a more detailed performance analysis of the GLRT used in the detection algorithm. Specifically, we consider the case where a single anomaly, \bar{g}^* , of unknown amplitude exists at some fine scale and we perform a generalized binary hypothesis in which the two hypotheses correspond to coarse scale structures one of which covers \bar{g}^* and one of which does not. We are interested in examining how the probability of correctly choosing the overlapping structure (which we call the probability of detection for these experiments) using the GLRT of Section 3.2 varies with the scale and position of the non-overlapping alternate as well as the amplitude of the true anomaly. High detection probabilities reflect favorably on the GLRT-based approach of the scale-recursive algorithm.

Following the notation of (17), let $l_1(\eta)$ be the statistic associated with the overlapping anomaly hypothesis and $l_0(\eta)$ be the statistic for the non-overlapping case. From (15) and (16), the probability of choosing the overlapping structure given knowledge of $\bar{\gamma}^* = \mathcal{W}_g \bar{g}^*$ is

$$\text{Prob}[L_1(\eta) > 0 | \bar{\gamma}^*] = \text{Prob}[l_1(\eta) - l_0(\eta) > 0 | \bar{\gamma}^*]. \quad (29)$$

Upon substituting (14) into (16) and using (17), straightforward linear algebra demonstrates that the random variable $L_1(\eta)$ may be written as

$$L_1(\eta) = x_1^2(\eta) - x_0^2(\eta) \quad (30)$$

where the two-vector $x(\eta) = [x_1(\eta) \ x_0(\eta)]^T$ is

$$x(\eta) = \mathcal{B}_{10}^T \Theta^T P_\eta^{-1} \eta \sim \mathcal{N}(\mathcal{B}_{10}^T \Theta^T P_\eta^{-1} \Theta \bar{\gamma}^*, \mathcal{B}_{10}^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_{10}) \quad (31)$$

and for $j = 0, 1$

$$\mathcal{B}_{10} = \begin{bmatrix} s_1 \mathcal{B}_1 & s_0 \mathcal{B}_0 \end{bmatrix} \quad (32a)$$

$$s_j^2 = \frac{1}{2} [P_j(1 + \alpha P_j)] \quad (32b)$$

$$P_j = (\mathcal{B}_j^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_j + \alpha)^{-1} \quad (32c)$$

From (30), $\text{Prob}[L_1(\eta) > 0 | \bar{\gamma}^*] = \text{Prob}[|x_1(\eta)| > |x_0(\eta)| | \bar{\gamma}^*]$ which is the integral of the probability density function for $x(\eta)$ defined in (31) over the shaded region in Figure 9.

In Figure 10, detection probabilities are displayed for binary hypothesis tests where \bar{g}^* is the structure in Figure 5(a) and the hypotheses are pairs of structures from \mathcal{J} . For example, the shade of dark region in the lower left corner in Figure 10(a) is $\text{Prob}[L_1(\eta) > 0 | \bar{\gamma}^*]$ for the BHT where the first hypothesis is the large structure overlapping the true, smaller size anomaly (represented by the white region in Figure 10(a)) and the alternate hypothesis is the 8×8 pixel lower left corner of A . Similar interpretations hold for the other two dark areas in Figure 10(a) and for each of the smaller square areas in Figures 10(b)–(c). For all of these images, the ABR for the true, small anomaly is set to 1.5. Figures 10(a)–(c) indicate the manner in which the detection performance of the GLRT-based algorithm depends upon the scale of the hypotheses relative to that of the true anomaly. At the coarsest scale, detection probabilities are about 60%. However, for all finer scales, P_d rises sharply with the lowest values confined to structures which are close to the true anomaly.

In Figure 10(d), we display the minimum P_d at each scale as a function of true anomaly's ABR. For example, the points on each of the three curves at an ABR of 1.5 are the minimum P_d values in each of the three images in Figure 10(a)–(c). From these curves we see that at the coarsest scale, even at high ABRs, the detection probabilities reach about 80%. As expected, when the hypotheses are drawn from the finer scales, the minimum P_d rises quickly to close to 100%.

The results in Figures 10 indicate that if the scale-recursive anomaly detection algorithm developed in Section 6.1 correctly identifies the coarse scale structures overlapping the true anomalies, then the detection performance at finer scales should be quite good even at ABRs less than 1. Also, because the *lowest* detection probabilities at fine scales are associated with structures close to the true structure, it is anticipated that the scale-recursive detection algorithm should be very successful in producing estimates of anomalies which are “sufficiently close” to the truth if not exactly the truth. This idea will be made more precise in Section 6.3.

The analysis in this section indicates that the primary difficulty associated with the algorithm is

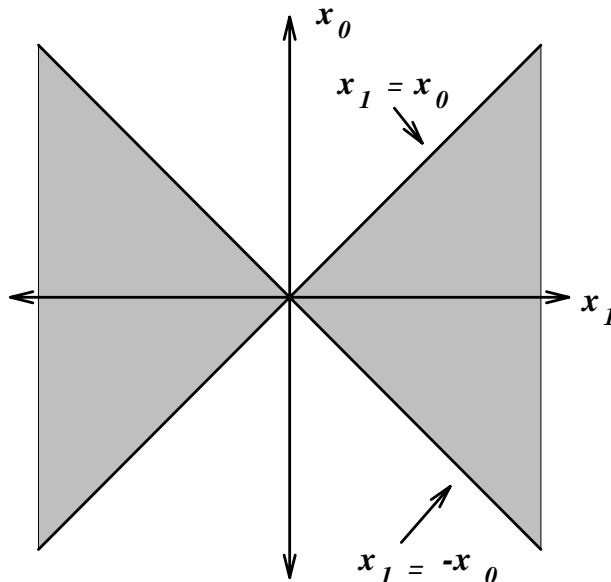


Figure 9: Integration region in $x_1 - x_0$ space for evaluation of $\text{Prob}[L_1(\eta) > 0 | \bar{\gamma}^*]$ in (29)

that coarse scale detection probabilities can be low. To overcome the potential problem of selecting the wrong area or areas of A for further refinement at coarse scales we modify the scale-recursive algorithm in the following manner. At the opening stage of the algorithm, rather than accepting the *single* hypothesis with the largest generalized log-likelihood value, we consider further refinement of A based upon those hypotheses corresponding to the *four* largest log-likelihood values (excluding H_0 and H_9). As will be seen in Section 6.3, despite the additional computational requirements of this approach, the overall complexity of the algorithm remains rather low. Finally, we note that one could extend this strategy of keeping additional structures for further refinement to more than just the first stage of the algorithm and could retain fewer or greater than four alternatives; however for the application of interest here, the choices described above were sufficient.

6.3 Examples

In this section, we examine the performance of the scale-recursive algorithm described in Sections 6.1 and 6.2. First, we use Monte Carlo studies to verify the ability of this approach to detect anomalous structures. The quantities of interest here are the sample probability of detection, \bar{P}_d , the sample average value of the number of false alarms per pixel \bar{P}_f , and the sample probability of error, \bar{P}_e . We say that a particular rectangular anomaly, $\bar{\gamma}^*$, has been detected if there exists a column in \hat{B} which is sufficiently close to $\bar{\gamma}^*$. Specifically, we define a “region of ambiguity” associated with the anomaly structure currently under investigation. This area is constructed such that anomaly structures identified in this region are “essentially indistinguishable” from the true anomaly. More formally, we compute the probability of successfully distinguishing $\bar{\gamma}^*$ from each member of \mathcal{J} in a binary hypothesis test of the form in (28a)–(28b). For each such test, the amplitudes of the two anomalies are chosen so that relative to the anomaly-free background, the two structures are equally detectable (i.e. they individually have the same d^2 value as defined with $P_d = 0.85$ and $P_f = 0.10$ in (18) and (26).) A pixel in A is said to be in the ambiguity region if (1) there exists a member of \mathcal{J} which is nonzero on that pixel and (2) the probability of distinguishing that element of \mathcal{J} from $\bar{\gamma}^*$ is below a given threshold, taken as 0.85 for all problems considered in

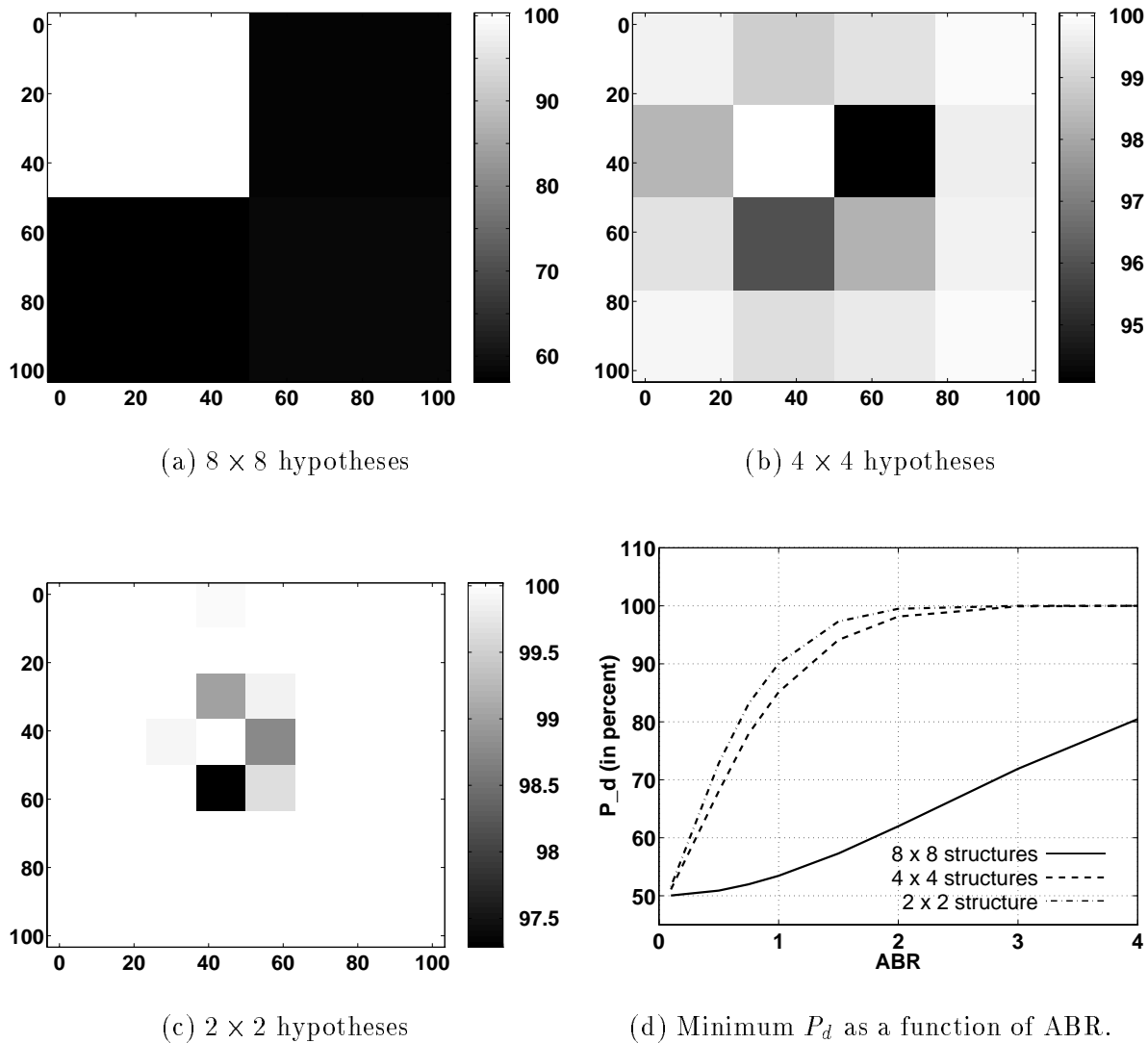


Figure 10: In (a)–(c) detection probabilities are displayed for binary hypothesis tests where \bar{g}^* is the structure in Figure 5(a) and the hypotheses are pairs of structures from \mathcal{J} . For each such test, one of the hypotheses is a larger scale structure overlapping \bar{g}^* while the second structures is from the same scale as the first but is disjoint from the pixels of \bar{g}^* . The shade of each square in (a)–(c) is the probability of correctly choosing the overlapping structure when the alternate is the anomaly occupying the square under investigation. The ABR for the true structure is 1.5. The minimum P_d at each scale as a function of true anomaly's ABR is shown in (d).

this section. Finally, for an estimated structure to be called a detection the area of intersection between it and the region of ambiguity must be at least a quarter of the area of the estimated structure. Such a definition implies a constraint on the localization of an estimated anomaly in both space and scale before we will call it a detection. As an example, the region of ambiguity at $P_d = 0.85$ associated with the anomaly structure in Figure 5 is displayed in Figure 11. The elements of $\hat{\mathcal{B}}$ which do not correspond to detections are taken to be false alarms and the per-pixel false alarm rate, \bar{P}_f , is defined as the total number of false alarm pixels divided by the number of

pixels in region A . Finally, the sample probability of error is $\bar{P}_e = 1 - \bar{P}_d + \bar{P}_f$.

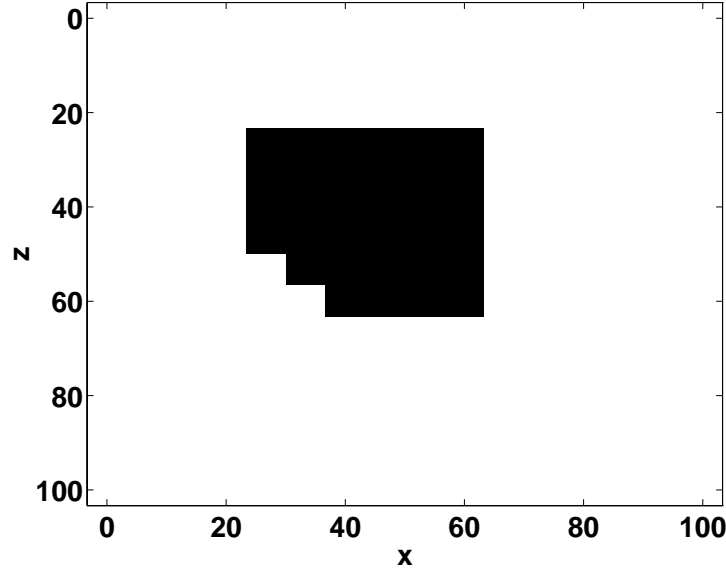


Figure 11: Region of ambiguity for structure shown in Figure 5 for $P_d = 0.85$.

We also examine the computational complexity of the scale-recursive algorithm. The complexity of the algorithm is quantified in terms of the number of Generalized Likelihood Ratio Tests (GLRTs) which must be performed in the processing of the data. As the spatial decomposition of region A is driven by the noisy data, the number of GLRTs will vary from one data set to the next. Thus, for a particular \bar{g}^* , the computational performance is based upon the average number of required GLRTs required per iteration of the corresponding Monte-Carlo.

Finally as discussed in Section 1, the detection algorithm results are used to improve the solution to the full reconstruction inverse problem. From our model for γ in (6), the estimate of the overall conductivity is the sum of the estimates of $\bar{\gamma}$ and $\tilde{\gamma}$, denoted $\hat{\gamma}$ and $\hat{\tilde{\gamma}}$ respectively, where $\hat{\gamma} = \hat{\mathcal{B}}\hat{a}$ is provided by our scale-recursive detection algorithm. Now, the linear least-squares estimate (LLSE) of $\tilde{\gamma}$ developed in [37,40] is based upon the assumption that *no* anomalies exist in the data; however, the output of the detection algorithm provides additional information through $\hat{\tilde{\gamma}}$ as to the structure of the conductivity field. To make use of the information in order to improve the estimate $\hat{\gamma}$, we define $\hat{\tilde{\gamma}}_c$ as the LLSE of $\tilde{\gamma}$ based upon a “corrected” data set in which the effects of $\hat{\tilde{\gamma}}$ have been removed. Mathematically this corrected estimate takes the form

$$\hat{\tilde{\gamma}}_c = P\Theta^T R^{-1} [\eta - \Theta\hat{\mathcal{B}}\hat{a}] \quad (33)$$

where $P = (\Theta^T R^{-1} \Theta + P_0^{-1})^{-1}$ is the error covariance matrix for nominal LLSE. Thus, the estimate of the overall conductivity field is

$$\hat{\gamma} = \hat{\tilde{\gamma}}_c + \hat{\tilde{\gamma}} = P\Theta^T R^{-1} \eta + [I - P\Theta^T R^{-1} \Theta] \hat{\mathcal{B}}\hat{a} \quad (34)$$

where we recognize the term $P\Theta^T R^{-1} \eta$ as the uncorrected LLSE estimate [49].

Unless otherwise stated, the data upon which the examples are based are generated using the Born-based measurements model in (2) for the scattering experiments described in Table 1. For all cases consider, the background conductivity, g_0 , is set to 1 S/m and at the highest ABRs of interest, the anomaly amplitudes are only 0.7 S/m. As discussed in [25], under these circumstances the Born approximation is known to be valid. In Section 7, we discuss issues associated with extending the work in this paper to account for the underlying non-linearity associated with the inverse conductivity problem. Finally, for all experiments the parameter α in (14) is set to 0.25.

6.3.1 The Single Anomaly Case

We begin by considering the case where it is known that there is a single anomaly of unknown amplitude and location in region A . Given that there is only one structure, the combinatorial complexity associated with an “exhaustive search” for the anomaly is sufficiently low that we shall compare both the detection/false-alarm performance as well as the complexity of the scale-recursive approach against an alternate algorithm akin to a multi-scale matched filter. This algorithm detects the single anomaly by computing the GLRT for each of the structures in family \mathcal{J} taking that element of \mathcal{J} associated with the largest GLRT statistic as the estimate. Because this method is multiscale in nature and has a fixed number of GLRTs per Monte-Carlo iteration (since there are a fixed number of structures in \mathcal{J}) it allows for a fair comparison against which we can judge the performance of the scale-recursive algorithm. For the scale-recursive method, we shall account for the knowledge that there is only a single anomaly in A by retaining only the column of \mathcal{B}_{leaf} associated with the most likely anomaly structure. Finally, for this example, the true anomaly structure is shown in Figure 5 and the SNR for all scattering experiments is 10.

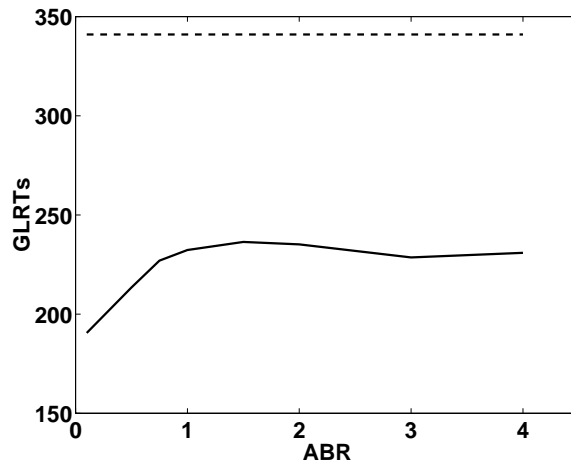
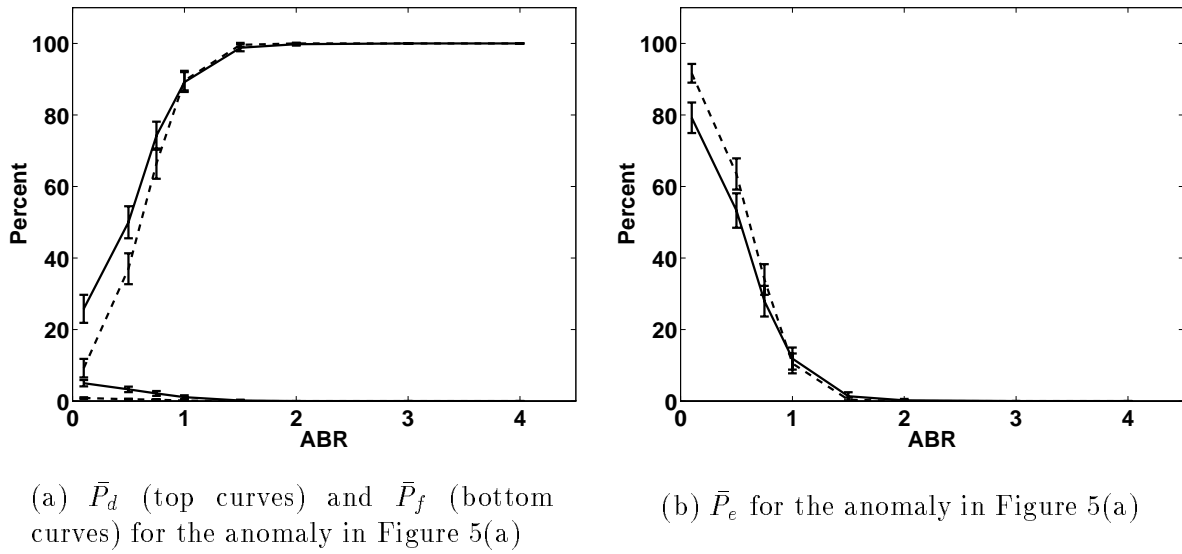
In Figure 12 we show \bar{P}_d , \bar{P}_f , \bar{P}_e and the average number of GLRTs per Monte-Carlo iteration as a function of anomaly-to-background ratio obtained after 500 Monte-Carlo iterations. The solid lines are the results for scale-recursive algorithm and the dashed lines indicate the performance of the multi-scale exhaustive search procedure. Figure 12(a) indicates that at low ABRs, the scale-recursive approach tends to have a higher detection probability than the exhaustive search with a slightly higher probability of false alarm. Even for the low ABR of 0.50, \bar{P}_d is well above 50% and rises to above 90% for ABR values greater than one. At high ABRs the performance of the two algorithms is about the same. Despite the slightly higher \bar{P}_f of the scale-recursive approach, the overall error probability is lower for the scale-recursive method at these small ABRs. Finally, from 12(c) the computational complexity of the scale-recursive characterization algorithm is seen to be roughly constant across the ABR range at 65% that of the exhaustive search.

In Figure 13(a) we display one realization of $g = \bar{g} + \tilde{g}$ obtained in our Monte Carlo process at an ABR of 1.5. Using the LLSE to perform the full reconstruction as in [40] results in the image in Figure 13(b). By incorporating the results of the scale-recursive detection algorithm into the inversion procedure through the use of (34), we obtain the estimate of the overall conductivity field shown in 13(c). Thus, successful identification of the highly parameterized anomaly structures can significantly improve localization both in space and scale and the GLRT procedure results in an accurate estimate of the structure’s amplitude. Also, the details in the remainder of the estimate do in fact reflect the coarse scale, fractal features of the conductivity profile in Figure 13(a).

6.3.2 The Multiple Anomaly Case

We now turn our attention to the case where multiple anomalies exist in region A .⁴ Lifting the single anomaly assumption causes the computational complexity of an exhaustive-search-type of approach to be prohibitive in that one would be required to examine the likelihood of all combinations of all non-overlapping, structures in a collection such as \mathcal{J} assuming separately $n = 1$ then $n = 2$ through $n = N_{max}$ anomalies exist in region A where N_{max} is a pre-determined maximum number of anomalies. Thus, here we present only the results of the scale-recursive detection algorithm. In particular, we explore the performance for the anomaly configuration in Figure 14.

⁴Note that in this multi-anomaly case, the ABR is used to determine the magnitude of each structure individually. For example at an ABR of one, the amplitude of the left anomaly is set so that if it were the only structure in the medium, the ABR would be one.

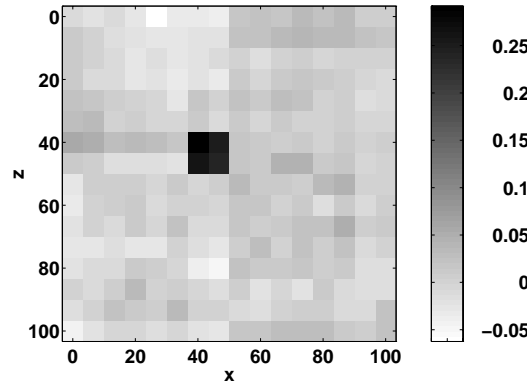


(c) Average number of GLRTs per Monte-Carlo iteration for anomaly detection algorithms for the anomaly in Figure 5(a).

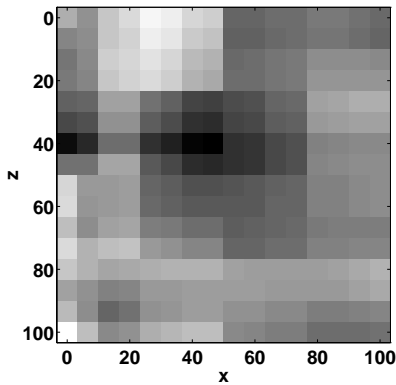
Figure 12: Performance curves as a function of ABR obtained after 500 Monte-Carlo iterations for the anomaly in Figure 5(a). Solid lines = results for scale-recursive algorithm. Dashed lines = results for multi-scale exhaustive search. The error bars are drawn at the plus/minus two standard deviation level.

The Monte-Carlo results for this experiment are displayed in Figure 15 where the top two curves of (a) correspond to the individual \bar{P}_d statistics for the two anomalies and the lowest of the three curves is a plot of \bar{P}_f . Here we see that both structures are quite easily detected with a \bar{P}_d of well over 90% even at the low ABR of one. As is expected, removing the single-anomaly assumption causes the algorithm to retain a greater number of candidate structures (including the true anomalies) thereby raising \bar{P}_f above that seen in Section 6.3.1.

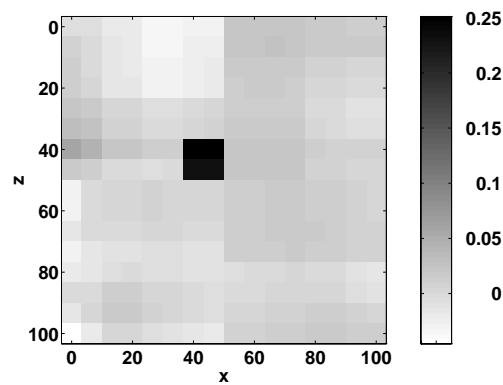
In Figure 15(c) we plot the average number of GLRTs as a function of ABR. Note that at



(a) Anomaly in fractal background
(ABR=1.5)



(b) Reconstruction of conductivity in
(a) using LLSE



(c) Reconstruction of conductivity
in (a) using (34)

Figure 13: Comparison of reconstructed conductivity profile using the LLSE of [40] and an estimate based upon the output of the scale-recursive anomaly detection algorithm. The true conductivity is shown in (a) and contains a single anomaly near the center of the region. The LLSE is shown in (b) and the estimate obtained from (34) is illustrated in (c). Here we see that the use of the information from the detection algorithm allows for the successful localization of the anomaly in space and scale without sacrificing our ability to resolve the fractal features of the conductivity profile in (a). Additionally, the GLRT procedure results in an accurate estimate of the anomaly's amplitude.

worst the complexity of this algorithm is still well below the complexity of the single-anomaly exhaustive search algorithm and only about 30% greater than the complexity of the single-anomaly scale-recursive algorithm. Thus, despite the fact that the multiple anomaly problem is, from a combinatorial viewpoint, significantly more complex than the single anomaly case, we see that the scale-recursive localization method represents a highly efficient and accurate means of localizing an unknown number of structures in the region of interest.

In Figure 16, we compare the full reconstruction results obtained from the LLSE to those where (34) is used to estimate the underlying conductivity for one run of the Monte-Carlo at an ABR of

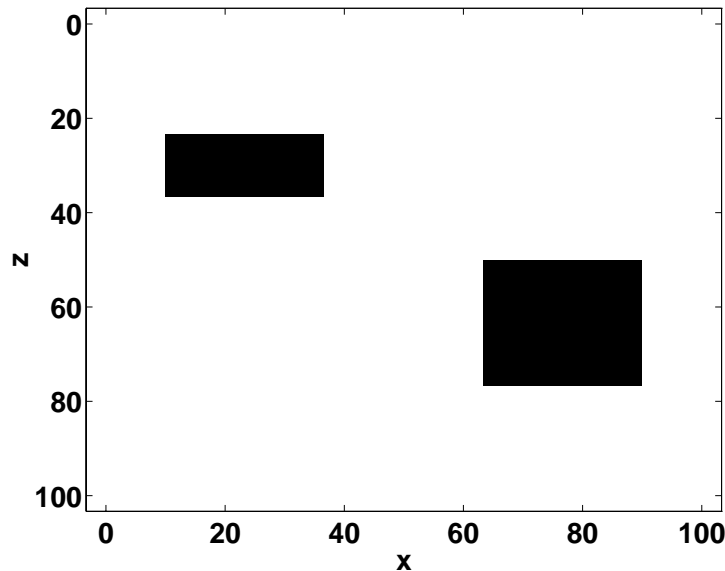


Figure 14: Two-region anomaly structure

1.5. From Figure 16(b) we see that the LLSE is successful in reconstructing the structure on the left; however, the lower amplitude/more pixel anomaly is almost completely undetected. Figure 16(c) indicates that the incorporation of the information from the anomaly detection algorithm significantly improves the localization in space as well as scale of both anomaly structures, especially the rightmost. Finally, the anomaly amplitudes are better estimated using the GLRT method.

7 Conclusion and Future Work

In this paper, we have presented a framework based upon techniques from the areas of multiscale modeling, wavelet transforms, and statistical decision and estimation theory for addressing a variety of issues arising in anomaly detection problems. Beginning with a linear model relating the data and the quantity to be reconstructed, we use the wavelet transform to take the problem from physical space to scale space where computational complexity is reduced for a wide variety of problems [1, 4, 41] and where we are able to take advantage of the rich and useful class of models recently developed for describing the structure of the medium in the absence of anomalous activity [21, 35, 47, 50]. The problems of characterizing the number, positions, and magnitudes of anomaly structures was formulated using the tools of statistical decision theory. To understand how the physics of the problem and the constraints on the geometry of the data collection process affect our ability to isolate anomalous regions, we defined and explored the issues of anomaly detectability and distinguishability. This analysis led to the development of a scale-recursive algorithm employing a sequence of Generalized Likelihood Ratio Tests for extracting anomaly information from data.

This work was presented in the context of a linearized inverse scattering problem arising in geophysical prospecting. The same scattering model is encountered in a variety of other fields where some form of energy is used to probe a lossy medium [18–20, 22, 29, 31]. More generally, the analysis and algorithmic methods developed in this work require only a measurements model of the form in (2) and are thus relevant for any linear inverse problem (eg. computed tomography) in which anomaly characterization is of interest.

An important extension of the work presented here is in development of algorithms and analysis

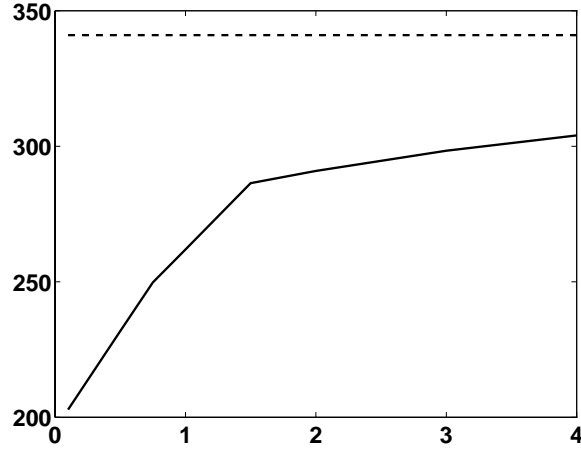
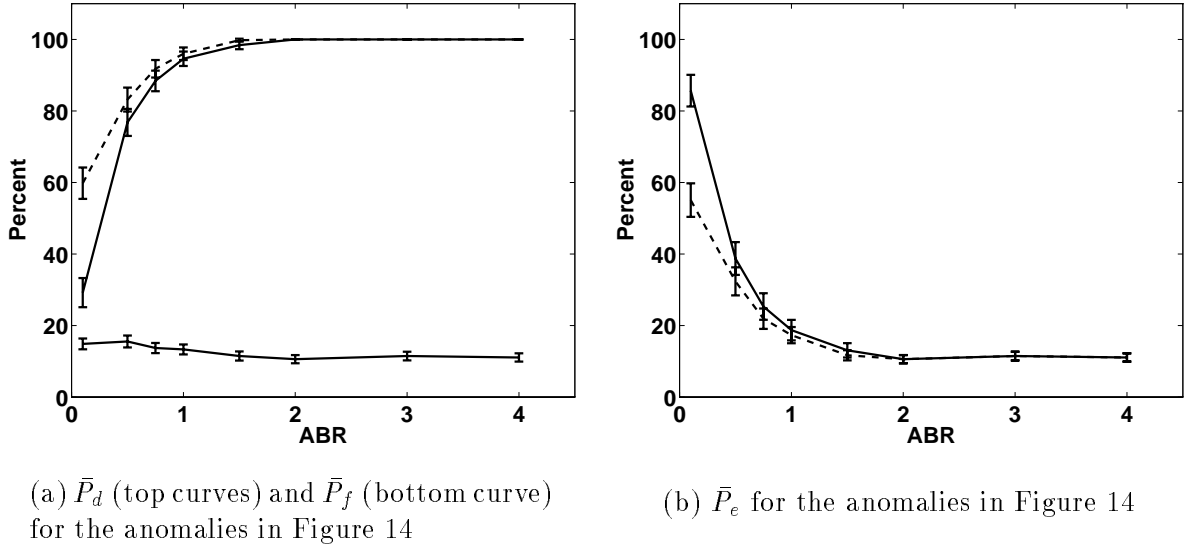


Figure 15: Performance curves obtained after 500 Monte-Carlo iterations of scale-recursive detection algorithm for the anomalies in Figure 14. Solid lines in (a) and (b) are detection and error probabilities for the upper left anomaly while dashed lines are for lower right anomaly. The error bars are drawn at the plus/minus two standard deviation level. In (c), the computational complexity associated with this scenario is shown by the solid line. For comparison, the dashed line is the complexity associated with the single anomaly exhaustive search.

methods for detecting anomaly structures using the *nonlinear* physics governing the relationship between the conductivity and the observed scattered electric field. The primary difficulty here is maintaining or improving the detection/false-alarm performance of the current method while retaining the low computation complexity in an algorithm based upon a significantly more complex scattering model. In [38] we present preliminary results for one form of a scale-recursive anomaly

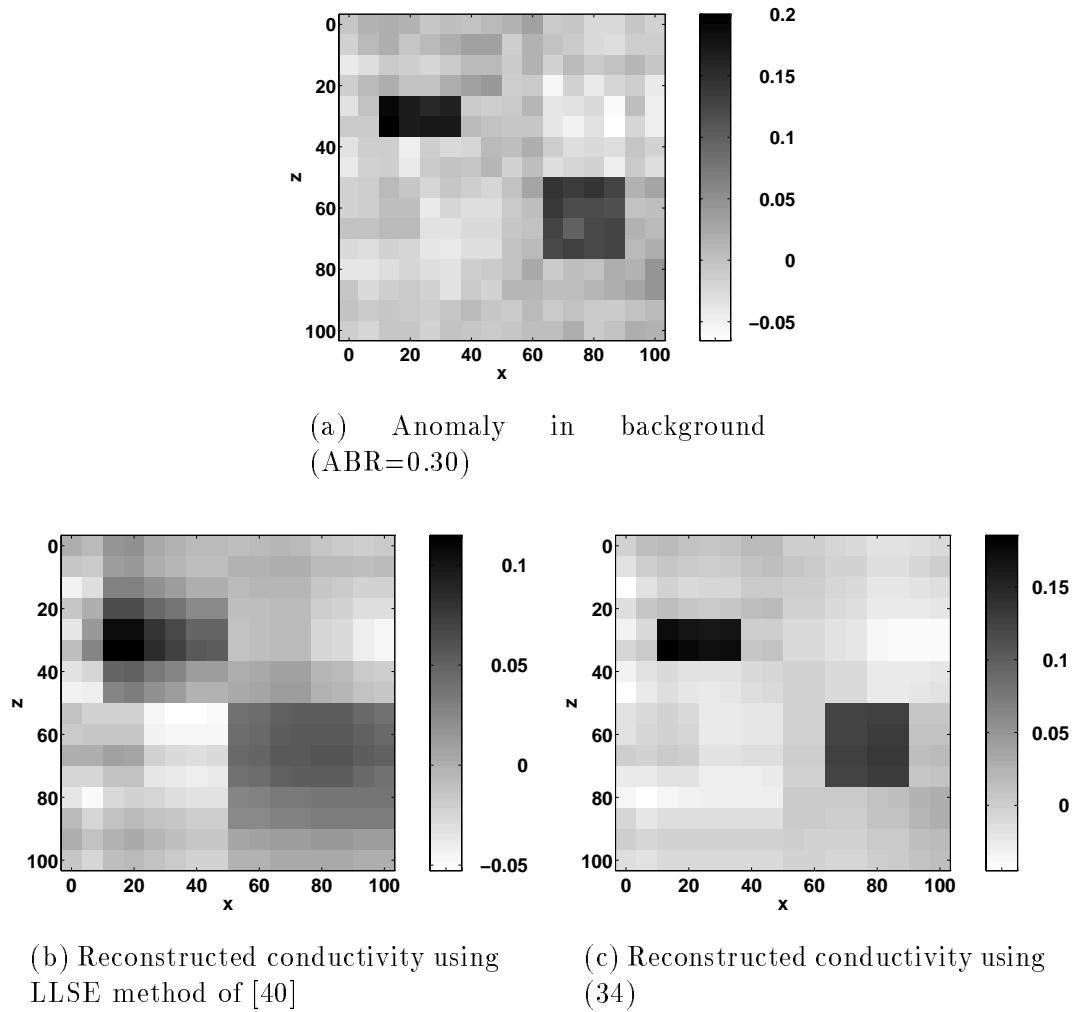


Figure 16: Comparison of reconstructed conductivity profile using the LLSE of [40] and an estimate based upon the output of the scale-recursive anomaly detection algorithm. The true conductivity is shown in (a) and contains a two anomalies. The LLSE is shown in (b) and the estimate obtained from (34) is illustrated in (c). Here we see that the use of the detection information allows for the successful localization of both anomaly structures and offers a significant improvement over the LLSE in localizing the anomaly in the lower right.

characterization algorithm using the computationally efficient, nonlinear Extended Born Approximation [48]. Further work remains in the exploration of detectability and distinguishability in the nonlinear context and the extension of this approach to higher frequency (eg. microwave) problems.

Another avenue of research is in the use of alternate methods for progressively dividing the region of interest. The problem of anomaly detection is similar to that of image segmentation in that the goal of both is to partition a two-dimensional grid of pixels into disjoint regions. The primary difference between these two problems is the data. In the segmentation case the data are the pixels in the image whereas we wish to do the anomaly localization given the significantly less informative observations of scattered radiation. For the segmentation problem, significant work has been performed in the use of hierarchical methods for performing this decomposition. For example,

segmentation techniques have been developed where (a) small structures are merged into larger regions [36, 42] and (b) both splitting as well as merging operations are used in the segmentation process [28]. Examining the utility of merge- and split/merge-algorithms for the anomaly detection problem would be of considerable interest especially as a means of overcoming the difficulties of detecting small-scale structures using large scale hypotheses.

References

- [1] B. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin. Wavelets for the fast solution of second-kind integral equations. *SIAM J. on Scient. Comput.*, 14(1):159–184, 1993.
- [2] M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data. I: General formulation and singular system analysis. *Inverse Problems*, 1:301–330, 1985.
- [3] M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data. II: Stability and regularisation. *Inverse Problems*, 4:573–594, 1988.
- [4] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, 44:141–183, 1991.
- [5] Yoram Bresler, Jeffrey A. Fessler, and Albert Macovski. A Bayesian approach to reconstruction from incomplete projections of a multiple object 3D domain. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(8):840–858, November 1989.
- [6] Kevin E. Brewer and Stephen W. Wheatcraft. Including multi-scale information in the characterization of hydraulic conductivity distributions. In Efi Foufoula-Georgiou and Praveen Kumar, editors, *Wavelets in Geophysics*, volume 4 of *Wavelet Analysis and its Applications*, pages 213–248. Academic Press, 1994.
- [7] Stephen R. Brown. Transport of fluid and electric current through a single fracture. *Journal of Geophysical Research*, 94(B7):9429–9438, July 1989.
- [8] W.C. Chew and Y. M. Wang. Reconstruction of two-dimensional permittivity distribution using the distorted born iterative method. *IEEE Trans. Medical Imaging*, 9(2):218–225, June 1990.
- [9] Weng Cho Chew. *Waves and Fields in Inhomogeneous Media*. Van Nostrand Reinhold, New York, 1990.
- [10] Kenneth C. Chou, S.A. Golden, and Alan S. Willsky. Multiresolution stochastic models, data fusion and wavelet transforms. Technical Report LIDS-P-2110, MIT Laboratory for Information and Decision Systems, May 1992.
- [11] Kenneth C. Chou, Alan S. Willsky, and Ramine Nikoukhah. Multiscale recursive estimation, data fusion, and — regularization. *IEEE Trans. Automatic Control*, 39(3):464–478, March 1994.
- [12] Kenneth C. Chou, Alan S. Willsky, and Ramine Nikoukhah. Multiscale systems, Kalman filters, and Riccati equations. *IEEE Trans. Automatic Control*, 39(3):479–492, March 1994.

- [13] David J. Crossley and Oliver G. Jensen. Fractal velocity models in refraction seismology. In Christopher H. Scholtz and Benoit B. Mandelbrot, editors, *Fractals in Geophysics*, pages 61–76. Birkhauser, 1989.
- [14] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996, 1988.
- [15] A. J. Devaney. Geophysical diffraction tomography. *IEEE Trans. on Geoscience and Remote Sensing*, GE-22(1):3–13, January 1984.
- [16] A. J. Devaney and G. A. Tsihrintzis. Maximum likelihood estimation of object location in diffraction tomography. *IEEE Trans. ASSP*, 39(3):672–682, March 1991.
- [17] A. J. Devaney and G. A. Tsihrintzis. Maximum likelihood estimation of object location in diffraction tomography, part II: Strongly scattering objects. *IEEE Trans. ASSP*, 39(6):1466–1470, June 1991.
- [18] David C. Dobson. Estimates on resolution and stabilization for the linearized inverse conductivity problem. *Inverse Problems*, 8:71–81, 1992.
- [19] David C. Dobson and Fadil Santosa. An image-enhancement technique for electrical impedance tomography. *Inverse Problems*, 10:317–334, 1994.
- [20] David C. Dobson and Fadil Santosa. Resolution and stability analysis of an inverse problem in electrical impedance tomography: Dependence on the input current patterns. *SIAM J. Appl. Math.*, 54(6):1542–1560, December 1994.
- [21] Patrick Flandrin. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Information Theory*, 38(2):910–917, March 1992.
- [22] D. G. Gisser, D. Isaacson, and J. C. Newell. Electric current computed tomography and eigenvalues. *SIAM J. Appl. Math.*, 50(6):1623–1634, December 1990.
- [23] T. M. Habashy, W. C. Chew, and E. Y. Chow. Simultaneous reconstruction of permittivity and conductivity profiles in a radially inhomogeneous slab. *Radio Science*, 21(4):635–645, July–August 1986.
- [24] Tarek M. Habashy, Edward Y. Chow, and Donald G. Dudley. Profile inversion using the renormalized source-type integral equation approach. *IEEE Transactions on Antennas and Propagation*, 38(5):668–682, May 1990.
- [25] Tarek M. Habashy, Ross W. Groom, and Brian R. Spies. Beyond the Born and Rytov approximations: A nonlinear approach to electromagnetic scattering. *Journal of Geophysical Research*, 98(B2):1759–1775, February 1993.
- [26] R. F. Harrington. *Field Computations by Moment Methods*. Macmillan Publ. Co., 1968.
- [27] Jorg H. Hippler, Helmut Ermert, and Ludwig von Bernus. Broadband holography applied to eddy current imaging using signals with multiplied phases. *Journal of Nondestructive Evaluation*, 12(3):153–162, 1993.
- [28] Steven L. Horowitz and Theodosios Pavlidis. Picture segmentation by a tree traversal algorithm. *Journal of the ACM*, 23(2):368–388, April 1976.

- [29] David Isaacson. Distinguishability of conductivities by electrical current computed tomography. *IEEE Trans. on Medical Imaging*, MI-5(2):91–95, June 1986.
- [30] David Isaacson and Margaret Cheney. Current problems in impedance imaging. In David Colton, Richard Ewing, and William Rundell, editors, *Inverse Problems in Partial Differential Equations*, chapter Chapter 9, pages 141–149. SIAM, 1990.
- [31] David Isaacson and Margaret Cheney. Effects of measurement precision and finite numbers of electrodes on linear impedance imaging algorithms. *SIAM J. Appl. Math.*, 51(6):1705–1731, December 1991.
- [32] Dwight L. Jaggard. On fractal electrodynamics. In H. N. Kritikos and D. L. Jaggard, editors, *Recent Advances in Electromagnetic Theory*, pages 183–224. Springer-Verlag, 1990.
- [33] Jonathan M. Lees and Peter E. Malin. Tomographic images of p wave velocity variation at Parkfield, California. *Journal of Geophysical Research*, 95(B13):21,793–21,804, December 10 1990.
- [34] Valdis Liepa, Fadil Santosa, and Michael Vogelius. Crack determination from boundary measurements - Reconstruction using experimental data. *Journal of Nondestructive Evaluation*, 12(3):163–174, 1993.
- [35] Stephane G. Mallat. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Trans. PAMI*, 11(7):674–693, July 1989.
- [36] Jean Marie Beaulieu and Morris Goldberg. Hierarchy in picture segmentation: A stepwise optimization approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(2):150–163, February 1989.
- [37] Eric L. Miller. The application of multiscale and statistical techniques to the solution of inverse problems. Technical Report LIDS-TH-2258, MIT Laboratory for Information and Decision Systems, Cambridge, MA., August 1994.
- [38] Eric L. Miller. A scale-recursive, statistically-based method for anomaly characterization in images based upon observations of scattered radiation. In *1995 IEEE International Conference on Image Processing*, 1995. Washington D.C.
- [39] Eric L. Miller and Alan S. Willsky. A multiscale approach to sensor fusion and the solution of linear inverse problems. *Applied and Computational Harmonic Analysis*, 2:127–147, 1995.
- [40] Eric L. Miller and Alan S. Willsky. Multiscale, statistically-based inversion scheme for the linearized inverse scattering problem. to appear in *IEEE Trans. on Geoscience and Remote Sensing*, 1995.
- [41] Eric L. Miller and Alan S. Willsky. Wavelet-based, stochastic inverse scattering methods using the extended born approximation. In *Progress in Electromagnetics Research Symposium*, July 1995. Seattle, Washington.
- [42] Jacqueline Le Moigne and James C. Tilton. Refining image segmentation by integration of edge and region data. *IEEE Trans. on Geoscience and Remote Sensing*, 33(3):605–615, May 1995.

- [43] John E. Molyneux and Alan Witten. Impedance tomography: imaging algorithms for geophysical applications. *Inverse Problems*, 10:655–667, 1994.
- [44] David J. Rossi and Alan S. Willsky. Reconstruction from projections based on detection and estimation of objects—parts I and II: Performance analysis and robustness analysis. *IEEE Trans. on ASSP*, ASSP-32(4):886–906, August 1984.
- [45] Ken Sauer, James Sachs Jr., and Catherine Klifa. Bayesian estimation of D objects from few radiographs. *IEEE Trans. Nuclear Science*, 41(5):1780–1790, October 1994.
- [46] Alon Schatzberg, Anthony J. Devaney, and Alan J. Witten. Estimating target location from scattered field data. *Signal Processing*, 40:227–237, 1994.
- [47] A. H. Tewfick and M. Kim. Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Trans. Information Theory*, 38(2):904–909, 1992.
- [48] Carlos Torres-Verdín and Tarek M. Habashy. Rapid 2.5-D forward modeling and inversion via a new nonlinear scattering approximation. *Radio Science*, pages 1051–1079, July-August 1994.
- [49] Harry L. Van Trees. *Detection, Estimation and Modulation Theory: Part I*. John Wiley and Sons, New York, 1968.
- [50] G. W. Wornell. A Karhunen-Loeve-like expansion for $1/f$ processes via wavelets. *IEEE Transactions on Information Theory*, 36:859–861, July 1990.

Statistically-Based Methods for Anomaly Characterization in Images from Observations of Scattered Radiation*

Eric L. Miller
235 Forsyth Building
Northeastern University
360 Huntington Ave.
Boston, MA 02215
Tel: (617) 373-8386
email: elmiller@cdsp.neu.edu

February 14, 1997

Abstract

In this paper we present an algorithm for the detection, localization, and characterization of anomalous structures in an overall region of interest given observations of scattered electromagnetic fields obtained along the boundary of the region. Such anomaly detection problems are encountered in applications including medical imaging, radar signal processing, and geophysical exploration. The techniques developed in this work are based on a non-linear scattering model relating the anomalous structures to the observed data. A sequence of M-ary hypothesis test are employed first to localize anomalous behavior to large areas and then to refine these initial estimates to better characterize the true target structures. We introduce a method for the incorporation of prior information into the processing which reflects constraints relevant directly to the anomaly detection problem such as the number, shapes, and sizes of anomalies present in the region. The algorithm is demonstrated using a low-frequency, inverse conductivity problem found in geophysical applications.

Submitted to *IEEE Trans. on Image Processing*
EDICS: IP 2.5 Geophysical and Seismic Imaging

*This work was supported in part by the Department of Energy under grant DOE-FC07-95ID13395 and the Center for Electromagnetics Research at Northeastern University

1 Introduction

A common problem arising in a variety of application areas is the non-invasive detection, localization, and characterization of anomalous areas in an overall region of interest given observations of scattered acoustic or electromagnetic radiation collected along a portion of the region's boundary [9, 15]. For example, tumor detection in otherwise healthy tissue arises frequently in many areas of medical signal processing [1]. In geophysical exploration, particularly low values of electrical conductivity in a region of the earth provide one indication as to the presence of oil [18]. Finally, there has been significant interest recently in the use of ground penetrating radar for the detection of buried objects such as land mines and metallic drums containing hazardous waste [14].

Typical methods for solving this *anomaly detection* problem proceed by first forming an image of distribution of a particular physical quantity (eg. sound speed or electrical permittivity) internal to the medium and subsequently post-processing the reconstructed profile to locate and classify anomalous areas [9, 15]. The problem of image formation from scattered field measurements (also known as the inverse scattering problem) presents a collection of difficulties which makes this image-then-detect approach rather unattractive. First, because the reconstructions are obtained typically over a fine scale grid of pixels, these problems are computationally intensive. The complexity of many practically interesting problems is further increased by the non-linear relationship between the physical parameter of interest and the observed scattered fields. The resulting imaging algorithms require the solution of a large-scale, non-convex, non-linear optimization problem [18].

Additional problems arise from the fact that one often is restricted to probing the medium with sources operating at only a few frequencies and observing the scattered fields over a limited portion of the region's periphery. As discussed in [11, 13] under these circumstances the imaging problem is highly ill-posed in that the information content in the the data does not support a stable fine-scale reconstruction of a pixelated representation of the physical parameter. Commonly, this problem

of ill-posedness is solved through the use of a regularization scheme which enforces a degree of smoothness in the reconstruction [10]. The use of such a regularizer for anomaly characterization serves to *increase* the difficulty of detection and identification after image formation. Typical targets represent regions where the medium's structure is locally *not* smooth. Imaging with a smoothness regularizer results in a low-pass image in which these distinguishing anomaly characteristics are softened thereby increasing the difficulty of subsequent detection and identification.

Over the past decade, significant research has been performed on the detection of anomalies directly from scattered field data. Rather than solving a large, ill-posed imaging problem, these techniques employ statistical signal processing methods to extract from the data a relatively small number of geometric parameters describing the distribution of anomalous structures. In [15], Rossi and Willsky considered the use of estimation-theoretic analysis and algorithmic methods for determining the location of a single object of known size and structure given noisy and sparse computed tomography (CT) measurements. Devaney and co-workers [3, 17] extended these results to diffraction tomography (DT) and exact scattering applications. In [2], Bressler, Fessler and Macovski presented an algorithm for determining the locations and geometric structures of an unknown number of anomalies based on CT measurements. A similar problem is considered in [16, 20], where the authors were concerned with the detection of an unknown number of circular objects of unknown radii given noisy CT observations. The iterative algorithms in [16, 20] began by identifying a large collection of likely targets with subsequent steps (a) pruning away structures which were deemed unlikely to be actual objects (b) refining the estimates of the radii and locations for the remaining objects.

A decision-theoretic, multiscale algorithm based on a sequence of M-ary, hypothesis tests was presented and analyzed in [11] for the characterization of an unknown number of anomalous regions of unknown size, shape and amplitude for problem in which a linear model related the anomalies to

the measurements. The approach in [11] began by identifying a small number of large sized regions in which anomalies were likely to exist. Subsequent processing steps refined these estimates and provided for the re-partitioning of each coarse scale region to account for the existence of multiple structures. The application of interest was a diffraction tomography problem arising in geophysical exploration.

In this paper, the problem of detecting and localizing an unknown number of anomalous areas is addressed for problems where linear scattering models are not valid so that the full, nonlinear physics must be incorporated into the processing. As in [11], the fundamental idea behind the approach presented here is the use of a sequence of hypothesis tests which begin by localizing anomalous behavior to a few, large sized areas and subsequently “zoom in” to better localize the true target structures. It should be noted that the spatial subdivision process considered here is quite different from that in [11]. In addition to the use of a non-linear scattering model, here we introduce a decision-theoretic regularization method which allows for the inclusion of prior information specifically tailored to the anomaly detection problem. Expectations regarding the number, sizes and shapes of anomalous areas are easily and naturally incorporated into the hypothesis testing process. While we do not examine the issue extensively in this work, the anomaly detection algorithm has a structure making it well tailored for a computationally efficient, parallel implementation. As in [11], we concentrate on a low-frequency geophysical inverse conductivity problem.

In Section 2 of this paper the scattering model and problem of interest are presented. An overview of the relevant detection methods is provided in Section 3. The multiscale anomaly detection algorithm is described in Section 4 and examples provided in Section 5. Section 6 is devoted to conclusions.

2 Problem Formulation

2.1 The Scattering Model

We consider a two-dimensional inverse conductivity problem illustrated in Fig. 1 where there exist a set of electromagnetic line-sources oriented perpendicularly to the page emitting time-harmonic waves into a lossy medium. The medium is characterized by the free space values for electric permittivity, ϵ_0 , and magnetic permeability, μ_0 and a conductivity profile modeled as the sum of a constant background, g_0 , and a collection of anomalies described by the function $g(\mathbf{r})$. The goal of the problem is to characterize $g(\mathbf{r})$ from noisy, pointwise observations of the component of the scattered electric field oriented perpendicularly to the page. These measurements are obtained along receiver arrays positioned on the vertical edges of A from K scattering experiments. Each experiment produces a vector of measurements, \mathbf{y}_i , comprised of the in-phase and quadrature components of the scattered field obtained over a single receiver array due to energy put into the medium from one of the sources operating at a frequency, $f_i = \omega_i/2\pi$.

As discussed in [18], the relationship between $g(\mathbf{r})$ and the j th element of \mathbf{y}_i ; that is, the measured scattered field at position \mathbf{r}_j is

$$y_i(\mathbf{r}_j) = \imath\omega_i\mu_0 \int_A G_i(\mathbf{r}_j, \mathbf{r}') E_i(\mathbf{r}') g(\mathbf{r}') d\mathbf{r}' + n_i(\mathbf{r}_j) \quad (1)$$

where $n_i(\mathbf{r}_j)$ represents additive noise, $G_i(\mathbf{r}, \mathbf{r}')$ is the Green's function for the problem and $E_i(\mathbf{r})$ is the component of the total electric field perpendicular to the page. From [18], $E_i(\mathbf{r})$ satisfies

$$E_i(\mathbf{r}) = \bar{E}_i(\mathbf{r}) + \imath\omega_i\mu_0 \int_A G_i(\mathbf{r}, \mathbf{r}') g(\mathbf{r}') E_i(\mathbf{r}') d\mathbf{r}' \quad (2)$$

with $\bar{E}_i(\mathbf{r}) = I_i \imath\omega_i\mu_0 G_i(\mathbf{r}, \mathbf{r}_i)$ the background field generated by a line source with current density I_i positioned at \mathbf{r}_i . Finally $G_i(\mathbf{r}, \mathbf{r}') = \imath/4H_0^{(1)}(k_{i,0}|\mathbf{r} - \mathbf{r}'|)$ is the Green's function for this problem with $k_{i,0}^2 = \omega_i^2\mu_0(\epsilon_0 + \imath g_0/\omega_i)$ and $H_0^{(1)}(z)$ the zeroth order Hankel function of the first kind.

A discrete representations of (1) is obtained using the method of moments (MOM) [8] where $g(\mathbf{r})$ and $E_i(\mathbf{r})$ are expanded in pulse bases (i.e. zeroth order splines). A Galerkin scheme is employed

to discretize (2) using the same pulse bases. Upon discretization, (1) and (2) reduce to

$$\mathbf{y}_i = \mathbf{G}_{i,s} \mathcal{D}(\mathbf{E}_i) \mathbf{g} + \mathbf{n}_i \quad (3a)$$

$$\mathbf{E}_i = \bar{\mathbf{E}}_i + \mathbf{G}_i \mathcal{D}(\mathbf{g}) \mathbf{E}_i \quad (3b)$$

where \mathbf{g} is a lexicographically ordered vector of expansion coefficients for $g(\mathbf{r})$, \mathbf{E}_i (resp. $\bar{\mathbf{E}}_i$) is a vector of coefficients for $E_i(\mathbf{r})$ (resp. $\bar{E}_i(\mathbf{r})$), $\mathbf{G}_{i,s}$ (resp. \mathbf{G}_i) is a matrix representation of the integral kernel in (1) (resp. (2)), and $\mathcal{D}(\mathbf{x})$ is the diagonal matrix whose (i, i) th element is the i th component of the vector \mathbf{x} . Solving for \mathbf{E}_i in (3b) and substituting the result in (3a) gives

$$\mathbf{y}_i = \mathbf{G}_{i,s} \mathcal{D} \left\{ [\mathbf{I} - \mathbf{G}_i \mathcal{D}(\mathbf{g})]^{-1} \bar{\mathbf{E}}_i \right\} \mathbf{g} + \mathbf{n}_i \equiv \mathbf{h}_i(\mathbf{g}) + \mathbf{n}_i. \quad (4)$$

In this paper, \mathbf{n}_i is taken to be additive, white Gaussian noise uncorrelated from one observation vector to the next. Thus, $\mathbf{n}_i \sim \mathcal{N}(\mathbf{0}, r_i \mathbf{I})$ and $E[\mathbf{n}_i \mathbf{n}_j^T] = r_i \mathbf{I} \delta(i - j)$ where \mathbf{I} is an appropriately sized identity matrix and $\delta(n)$ is the Kronecker delta.¹ Finally, we collect the data from the K scattering experiments into a single vector, \mathbf{y} so that the overall observation model is given by

$$\mathbf{y} = \mathbf{h}(\mathbf{g}) + \mathbf{n} \quad (5)$$

where $\mathbf{y}^T = [\mathbf{y}_1^T \ \mathbf{y}_2^T \ \dots \ \mathbf{y}_K^T]$, $\mathbf{h}(\mathbf{g})$ and \mathbf{n} are defined accordingly, and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ with $\mathbf{R} = \text{diag}(r_1 \mathbf{I}, r_2 \mathbf{I}, \dots, r_K \mathbf{I})$.

One difficulty with the use of (4) in the context of an inverse problem is that of computational complexity. A basic component of any inverse problem is the evaluation of \mathbf{y}_i for various values of \mathbf{g} . According to (4), each such evaluation requires the inversion of the large, dense matrix $\mathbf{I} - \mathbf{G}_i \mathcal{D}(\mathbf{g})$. In fact, for a discretization of A into an $N_g \times N_g$ array of pixels, this matrix is of size $N_g^2 \times N_g^2$. To overcome this computational burden, the algorithm discussed in Section 4 is implemented using the nonlinear, extended Born approximation (EBA) [6] to the true scattering model. That is, for all computations, $\mathbf{h}_i(\mathbf{g})$ is replaced by $\mathbf{h}_i^{EBA}(\mathbf{g})$ where a detailed discussion of $\mathbf{h}_i^{EBA}(\mathbf{g})$ may be found in [6, 13, 18]. Like the true scattering physics, the EBA is a nonlinear

¹The notation $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{P})$ indicates that the random vector $M\mathbf{x}$ has a Gaussian probability distribution with mean \mathbf{m} and covariance matrix \mathbf{P} .

function of the conductivity. For the geophysical application of interest here, the EBA has been shown to provide a highly accurate approximation of $\mathbf{h}_i(\mathbf{g})$ even for conductivity profiles where the more common Born or Rytov linearized models are no longer valid [6, 18]. The principle advantage of the EBA is that the computational complexity of the approximation is a small fraction of that associated with the exact scattering model [18].

2.2 The Anomaly Model

Because the MOM discretization procedure pixelates A into an array of rectangular cells and because the algorithm in Section 4 performs anomaly localization through a process of spatial subdivision, we are lead naturally to consider a representation for \mathbf{g} in which anomalous regions are defined to be superpositions of rectangular subsets of A . Referring to Fig. 2, the structure of the i th anomaly in A is defined by its magnitude, its size, and its location in A . Mathematically, the form for the anomalous behavior of the conductivity over the region A is

$$\mathbf{g} = \sum_{j=1}^{N_a} \mathbf{b}_j a_j = \mathbf{B}\mathbf{a}. \quad (6)$$

Here, N_a is the number of anomalous regions located in A , a_j is a scalar defining the magnitude of each anomaly, and \mathbf{b}_j represents the discrete indicator function over the j th rectangular region. In (6), the column vector \mathbf{a} represents the collection of anomaly amplitude coefficients while \mathbf{B} is the matrix whose j th column is the lexicographically ordered form of \mathbf{b}_j . We note that the algorithm in Section 4 is not dependent upon the assumption of rectangular anomalies; rather structures with arbitrary shapes and orientations can be employed in principle through the appropriate specification of the matrix \mathbf{B} . Finally, the anomaly detection problem may now be stated in the following manner: Given the data and the noise statistics, determine the structure of the matrix \mathbf{B} and the vector \mathbf{a} , i.e. find the number, locations, sizes, and the amplitudes of the anomalous regions in A .

3 M-Ary Hypothesis Testing

The anomaly detection algorithm to be presented and analyzed in Sections 4 and 5 is based on a sequence of M-ary hypothesis tests [19] in which each hypothesis corresponds to a particular anomaly distribution in A . That is, the m th hypothesis for one such test takes the form

$$H_m : \quad \mathbf{g} = \mathbf{B}_m \mathbf{a}_m \quad (7)$$

for $m = 0, 1, \dots, M - 1$. In (7) \mathbf{B}_m describes the geometric characteristics of the hypothesized anomaly configuration and \mathbf{a}_m is the associated amplitude vector.

Given a data vector, the hypothesis test itself may be viewed as a rule for selecting the “best” anomaly configuration from among the M possible choices. In this paper, the following maximum *a posteriori* (MAP) procedure is employed

$$\text{Choose } H_m \text{ with } m = \arg \max_j \log [p_{\mathbf{y}|H_j}(\mathbf{Y}|H_j)] + \log(p_j) \quad (8)$$

where $p_{\mathbf{y}|H_j}(\mathbf{Y}|H_j)$ is the probability density function for the data under H_j and p_j represents the prior probability that H_j is in fact true. The Gaussian model for \mathbf{n} allows (8) to be written as

$$\begin{aligned} \text{Choose } H_m \text{ with } m &= \arg \min_j \frac{1}{2} \|\mathbf{y} - \mathbf{h}(\mathbf{B}_j \mathbf{a}_j)\|_{\mathbf{R}^{-1}}^2 + \pi_j \\ &\equiv \arg \min_j \ell_j(\mathbf{y}) + \pi_j \end{aligned} \quad (9)$$

where $\|\mathbf{x}\|_{\mathbf{A}} \equiv \mathbf{x}^T \mathbf{A} \mathbf{x}$ and $\pi_j = -\log(p_j)$.

For the anomaly detection algorithm in Section 4, the amplitude vectors, \mathbf{a}_m , are assumed to be deterministic, but unknown quantities. Thus, a generalized hypothesis test is employed in which \mathbf{a}_m in (7) is replaced by the following estimate of \mathbf{a}_m given the data

$$\hat{\mathbf{a}}_m = \arg \min_{\mathbf{a}_m > -g_0} \frac{1}{2} \|\mathbf{y} - \mathbf{h}(\mathbf{B}_m \mathbf{a}_m)\|_{\mathbf{R}^{-1}}^2 + \alpha^2 \|\mathbf{a}_m\|. \quad (10)$$

The notation $\mathbf{a}_m > -g_0$ implies that each element of the vector \mathbf{a}_m is greater than the negative of the background conductivity and is required to ensure that the overall conductivity in region A , is greater than zero [18]. As in [11], (10) represents a regularized, maximum likelihood estimate of the amplitude vector where the second term is used to offset any ill-posedness in the estimation problem.

The solution to this constrained non-linear, least squares optimization problem is obtained using a Gauss-Newton algorithm [4] modified to account for the bound constraint in \mathbf{a}_m .

The evaluation of the each likelihood statistic, $\ell_m(\mathbf{y})$, requires the solution of a non-linear optimization problem; however, the dimension of each problem is equal to the number of anomalies associated with H_m which is typically far less than the number degrees of freedom found in imaging application [13, 18]. Thus, the computational burden of each optimization problem is quite small relative to an image-then-detect approach. For all examples presented in Section 5, the Gauss-Newton method typically converged in well under 10 iterations. Finally, while the hypothesis test does require the solution of many small optimization problems, each $\ell_m(\mathbf{y})$ may be evaluated independently so that complexity can in principle be kept low through a parallel implementation of the test.

4 The Anomaly Characterization Algorithm

Here we present a multiscale algorithm for the detection and characterization of an unknown number of anomalous areas in region A based on a sequence of MAP hypothesis tests. Two classes of tests are employed in this algorithm where each test takes as input the current estimate of the anomaly structures in the form of a geometry matrix, $\hat{\mathbf{B}}$, and an estimated amplitude vector $\hat{\mathbf{a}}$. The overall algorithm is a sequence of *localization hypothesis tests* which “zoom in” on the true anomalous regions coupled with *pruning tests* designed to eliminate false alarms thereby controlling the amount of work required to identify true anomalous structures.

In Fig. 7, a pseudo-code listing of the algorithm is displayed. The algorithm begins with a single localization test in which anomalous activity is identified in a collection of relatively large size areas in region A . After this coarse-scale detection procedure, the basic processing step is comprised of a single localization-type test followed by a sequence of one or more pruning tests.

The algorithm terminates when successive null hypotheses are chosen for both a localization test and then a pruning test; that is, when it is determined that no further subdivision is warranted and no structures should be removed from the current configuration. Essentially, this approach grows a tree of hypothesized anomaly configurations where the root of the tree is a single anomaly over all of region A and the leaves are either new anomaly structures or increasingly localized estimates of current anomalies.

A key feature of this procedure is the specification of the non-negative penalty terms, π_j , in (9) (or equivalently the prior probabilities $p_j = e^{-\pi_j}$) to control the manner in which the tree is both grown (i.e. new anomalies are added or old ones further refined) as well as pruned (i.e. structures are removed from processing.) For example, hypotheses may be penalized to a greater or lesser degree depending on features such as the number of anomalies associated with \mathbf{B}_m , the spatial distribution of these structures, or geometric characteristics such as area or shape. Thus, in their ability to mathematically capture prior expectations and constraints, the π_j play the same role in the hypothesis testing framework as that of a smoothness regularizer in a more traditional imaging-type problem [10]. In the anomaly detection context, the property of smoothness is not of concern, but rather geometric characteristics of the anomaly distribution are the natural features for which prior models are required. The remainder of this section is devoted to a description of both classes of hypothesis tests.

4.1 The Localization Hypothesis Test

Given an estimated configuration of anomalies in region A , the spatial refinement hypothesis test used here indicates either that no further localization is deemed necessary or further refinement is warranted in which case only a single structure in the current anomaly collection is further subdivided. Formally, suppose that N regions in A have been identified as containing anomalies so that $\hat{\mathbf{B}}$ contains N columns, $\hat{\mathbf{b}}_j$, $j = 1, 2, \dots, N$. Localization to smaller regions of A is obtained

through the use of a test with $8N + 1$ hypotheses. The first hypothesis corresponds to retaining the current $\hat{\mathbf{B}}$ indicating that no further localization is warranted. Under the next 8 hypotheses, the first column of $\hat{\mathbf{B}}$ is replaced by one or more columns corresponding to anomalies with support over a subset of the region corresponding to $\hat{\mathbf{b}}_1$. This same construction is repeated with $\hat{\mathbf{b}}_1$ replaced by $\hat{\mathbf{b}}_j$ for $j = 2, 3, \dots, N$. Thus, the non-null hypotheses correspond to $8N$ ways in which anomalies can be better localized by perturbing only one of the N elements of $\hat{\mathbf{B}}$.

In Fig. 4, the decompositions of $\hat{\mathbf{b}}_j$ of interest in this paper are displayed. The first six choices correspond to the true anomaly being located in the top, bottom, left, right, vertical middle, and horizontal middle of the region given by $\hat{\mathbf{b}}_j$. The last two decompositions allow for the presence of two anomalies in $\hat{\mathbf{b}}_j$: one in the top and one in the bottom or one in the left and one in the right. By choosing one of these last two hypotheses, we are able to identify multiple structures in A .

As discussed in Section 3, we associate with each hypothesis a matrix, \mathbf{B}_m for $m = 0, 1, \dots, 8N$ where $\mathbf{B}_0 = \hat{\mathbf{B}}$ and for $m > 0$, \mathbf{B}_m is equal to $\hat{\mathbf{B}}$ with one column replaced either by a single column, as is the case for the first six hypotheses in Fig. 4, or two column vectors for the last two hypotheses. In either case, we define $\tilde{\mathbf{b}}_m$ to be the column(s) in \mathbf{B}_m different from those in $\hat{\mathbf{B}}$.

In addition to specifying \mathbf{B}_m , the MAP test requires that the penalty terms π_m , also be specified. For the cases where $\tilde{\mathbf{b}}_m$ contains one column, we begin by defining a collection of quantities $\tilde{\pi}_m$ as

$$\tilde{\pi}_0 = 10^\sigma f_0 \left(\max_k |\hat{\mathbf{b}}_k| \right) \quad (11)$$

and for $m > 0$

$$\tilde{\pi}_m = \begin{cases} +\infty & |\tilde{\mathbf{b}}_m| < 4 \text{ or } \epsilon(\tilde{\mathbf{b}}_m) > 4 \\ 10^\sigma \text{card}(\tilde{\mathbf{b}}_m) f_1(|\tilde{\mathbf{b}}_m|) & \text{otherwise} \end{cases} \quad (12)$$

where $|\tilde{\mathbf{b}}_m|$ is the number of non-zero pixels in $\tilde{\mathbf{b}}_m$; $\epsilon(\tilde{\mathbf{b}}_m)$ is the ratio of the length of longer to the shorter side of the anomaly and provides a measure of the eccentricity of the structure; $\text{card}(\tilde{\mathbf{b}}_m)$ is the number of columns in $\tilde{\mathbf{b}}_m$ (in this case one); and $f_0(|\mathbf{b}_m|)$ and $f_1(|\mathbf{b}_m|)$ are plotted in Fig. 5.

When $\tilde{\mathbf{b}}_m$ contains two columns, the number of nonzero pixels in each column are equal and the geometric structures of the underlying anomalies are the same (only their positions in A differ), so that there is no ambiguity in assuming that $|\tilde{\mathbf{b}}_m|$ and $\epsilon(\tilde{\mathbf{b}}_m)$ refer to the first column in these cases.

Finally, π_m are obtained from $\tilde{\pi}_m$ via

$$\pi_m = \tilde{\pi}_m - \log \left(\sum_{k=0}^{8N} \tilde{\pi}_k \right) \quad (13)$$

thereby ensuring that $\sum_{m=0}^{8N} p_m = \sum_{m=0}^{8N} e^{-\pi_m} = 1$.

The parameter σ in (11) and (12) determines the overall magnitude of the penalty terms in the hypothesis test. Thus, selection of σ is analogous to the determination of the regularization parameter in many imaging type inverse problems [5, 7]. For σ too small the prior information will have negligible impact on the hypothesis test while an excessively large value of σ negates the influence of the data. In this paper, σ is chosen using the following somewhat *ad hoc* approach

$$\sigma = \left\lfloor \log_{10} \frac{1}{2} \|\mathbf{y} - \mathbf{h}(\mathbf{1}\hat{\mathbf{a}})\|_{\mathbf{R}^{-1}}^2 \right\rfloor - 1 \quad (14)$$

where $\mathbf{1}$ is the indicator function over all of region A , $\hat{\mathbf{a}}$ is the ML amplitude for an anomaly with this structure, and $\lfloor x \rfloor$ is the greatest integer less than or equal to x . Thus, σ is selected so that the largest of the $\tilde{\pi}_m$ is roughly an order of magnitude less than the size of the covariance weighted error between the data and the initial estimated anomaly structure used in the algorithm.

Fig. 5 and eqs. (11) and (12) demonstrate that the penalty terms are chosen to reflect prior expectations regarding the sizes, shapes and number of anomalies in region A . The structure of $f_0(|\mathbf{b}_m|)$ indicates that as the size of the largest structure in $\hat{\mathbf{B}}$ increases, the penalty for *not* performing a decomposition increases thereby encouraging the decomposition of larger structures. The first condition in (12) is used to ensure that no structures are identified if they contain few than four pixels or if they are too long and fat or tall and thin. That is, we require anomalies to have a minimum area and not be excessively elongated. The second condition in (12) and the form of $f_1(|\mathbf{b}_m|)$ indicate that the penalty increases as the number of columns in $\tilde{\mathbf{b}}_m$ increases and the size

of the structures decreases. Thus, the prior expectations serve to prevent the identification of many small anomalies in region A . Clearly, depending on the application, other prior modeling choices are both possible and interesting. The selection made here is useful for the inverse conductivity problem under consideration and provides a successful demonstration of the utility of prior modeling in an hypothesis testing framework for capturing information not commonly accessible to traditional regularization schemes.

4.2 The Pruning-Type Hypothesis Test

The second class of hypothesis tests used here are designed to eliminate from consideration previously identified regions which are unlikely to actually contain anomalies. Given an estimated geometry matrix $\hat{\mathbf{B}}$ with N columns, the anomaly removal hypothesis test consists of $N + 1$ hypotheses with $\mathbf{B}_0 = \hat{\mathbf{B}}$ and for $m = 1, 2, \dots, N$, \mathbf{B}_m equals $\hat{\mathbf{B}}$ with the m th column removed. Thus, at most one structure is eliminated per hypothesis test.

With $\hat{\mathbf{b}}_m$ the column of $\hat{\mathbf{B}}$ removed for the m th test, $\tilde{\pi}_m$ for the pruning-type hypothesis test is

$$\tilde{\pi}_m = \begin{cases} 10^\sigma f_2(N) & m = 0 \\ 10^\sigma f_3(|\hat{\mathbf{b}}_m|) & m = 1, 2, \dots, N \end{cases} \quad (15)$$

where $f_2(N)$ and $f_3(|\hat{\mathbf{b}}_m|)$ are shown in Figs. 5(a) and (b), σ is chosen as in (14), and π_m are obtained from $\tilde{\pi}_m$ using (13). As seen in Fig. 5, the form of $f_2(N)$ indicates an increasing penalty as the number of estimated structures rises. Also, Fig. 5(a) shows that $\tilde{\pi}_m$ increases as the size of the anomaly being removed from consideration rises thereby penalizing against the premature removal of large-scale structures.

4.3 Initial Localization

For the algorithm presented in this section to detect and localize relatively small anomalies, during the initial processing stages the correct large scale structures which overlap the true, smaller sized anomalies must be chosen for further processing. In [11], this coarse-scale localization issue

was analyzed in depth for a decision-theoretic anomaly detection algorithm applied to a scattering problem identical to the one considered here but using a linearized model rather than the non-linear model of Section 2. There, it was shown that the probability of a coarse scale miss was unacceptably high. In that case, performance was improved simply by retaining a greater number of coarse-scale regions for further decomposition than would otherwise be the case.

With these results in mind, in this paper we address the problem of coarse scale localization by modifying the hypothesis set used in the first localization test. Specifically, we supplement the decompositions of region A presented in Fig. 4 with those shown in Fig. 6 making for a 15-ary initial hypothesis test rather than the 8-ary test described in Section 4.1. Additionally, after performing this expanded localization-type test, we retain for further processing *all other* areas in region A not part of the chosen decomposition. For example, if the selection labeled “Horizontal middle two” is chosen from Fig. 6, then *four* anomalies are identified for further processing: the two square structures in the middle of region A as well as the two long, narrow regions at the top and bottom. While alternate means may be developed for addressing this issue of coarse scale localization, the method described here balances the conflicting requirements of an exhaustive, fine scale search over region A for likely targets and the desire to limit the computational overhead of the algorithm.

5 Examples

We consider the problem of characterizing conductivity anomalies given data from the scattering experiments described in Table 1. Region A is $100\text{ m} \times 100\text{ m}$ and discretized into a 16×16 array of pixels. The background conductivity is 0.1 S/m . The source frequencies are 100 Hz (used in a cross-well configuration) as well as 1 and 10 kHz (used to obtain information near the left edge of A). Both receiver arrays consist of 32 equally spaced elements extending from $z = -0.05\text{ m}$ to $z = 100.05\text{ m}$. The left array is located at $x = -0.05\text{ m}$ and the right at $x = 100.05\text{ m}$. The six

sources are located along the line $x = -0.05$ m equally spaced from $z = 0.05$ m to $z = 99.95$ m.

The anomalies in all experiments are taken to have amplitudes of 1 S/m representing a 10:1 contrast against the background conductivity. Under conditions such as these, it is well known that the Born and Rytov linearizations of the scattering physics fail to hold thereby necessitating the use of a more complete, nonlinear scattering model. While the detection algorithm is based on the extended Born approximation (EBA) described in Section 2.1, the data vectors for all simulations are generated using the exact scattering model in (3a) and (3b).

The performance of the anomaly detection algorithm is assessed using a collection of different anomaly configurations under a variety of signal to noise conditions. For the model $\mathbf{y}_i = \mathbf{h}_i(\mathbf{g}) + \mathbf{n}_i$ with $\mathbf{y}_i \in \mathbb{R}^{N_i}$ and $\mathbf{n}_i \sim \mathcal{N}(0, r_i^2 \mathbf{I})$, the signal-to-noise ratio (SNR) in decibels is defined as

$$SNR_i = 10 \log_{10} \left\{ \frac{[\mathbf{h}_i(\mathbf{g})]^T [\mathbf{h}_i(\mathbf{g})]}{N_i r_i^2} \right\}. \quad (16)$$

For each anomaly configuration performance is examined at SNRs of 40, 20, 10, and 0 dB. Given an anomaly and SNR combination, the detection algorithm is executed for 10 independent realizations of the measurement noise process with the results of these runs forming the basis for our evaluation.

The information for each anomaly configuration is analyzed both quantitatively as well as qualitatively. From a quantitative perspective, we are concerned with the detection and false alarm probabilities associated with the algorithm. An anomaly is classified as having been detected if the $\hat{\mathbf{B}}$ matrix produced at the last stage of the algorithm contains at least one structure which overlaps the true anomaly structure. All structures in the final $\hat{\mathbf{B}}$ matrix not corresponding to detections are taken to false alarms. The statistic of interest here is the rate of false alarms per unit area defined as number of falsely identified pixels divided the total number of pixel in A . Finally, the results of the anomaly detection algorithm are compared against those of an imaging-type inverse scattering algorithm described in [13]. This reconstruction procedure requires the solution of a nonlinear least squares problem. The method of [13] is based on a form of the Levenburg-Marquardt algorithm

and makes use of the EBA to lower the algorithm's computational complexity.

While it is not the objective of this paper to provide an exhaustive analysis of the computational complexity of the anomaly detection algorithm, this approach is generally more efficient than the imaging technique of [13]. All algorithms in this paper have been implemented in Matlab and executed on a dual processor, SparcStation 20 with 64Mb of internal RAM. The time required for a single run of the anomaly detection algorithm varied from 4 to about 7 hours depending on the underlying anomaly configuration and the noise conditions. For the imaging algorithm, 50 iterations of the Levenburg-Marquardt technique were used to process each data set at which point the algorithm had converged to a minimum of the least-squares cost function. In contrast to the detection-based approach, the imaging algorithm required between 24 and 36 hours of processing.

5.1 Experiment I: A Single Anomaly

The first anomaly configuration of interest here is shown in Fig. 7(a) and consists of a single structure located near the middle of region *A*. As discussed in [11–13], for inverse problems of the type considered in this paper, the ability to resolve anomalous behavior is greatest near either vertical edge of region *A* where the sources and receivers are located. For locations in the middle of the region, it becomes increasingly difficult to detect structures or resolve fine scale detail in the conductivity profile. Finally, the structure in Fig. 7(a) is aligned on the underlying grid of 16×16 pixels in such a manner that it is not possible for a decomposition scheme based upon the hypotheses in Figs. 4 and 6 to produce an estimated anomaly which is *exactly* of the same shape as the truth. Thus, this example represents a particularly challenging anomaly configuration for the algorithm.

In Fig. 7(b), the detection and false alarm probabilities are displayed for this case where the error bars indicate plus and minus two standard deviations from the sample mean. For the 40, 20, and 10 dB SNR cases, the anomaly was detected 100% of the time. At 0 dB, the detection

rate drops to 80%; however for the two cases in which the target was missed, the output of the algorithm did include an estimated anomaly located directly above the true structure. Thus, the target essentially was localized for all runs at all SNRs. Finally, the false alarm rate is at most 5% or 12 falsely identified pixels out of the 256. Sample estimates of anomaly configuration at 40dB and 10 dB SNR are presented in Fig. 8(a)–(b). The dashed lines overlaying these images indicate the region of A occupied by the true structure. We note that at both SNRs, the anomaly is detected. At 40 dB, there is also a small amplitude, falsely identified structure while the 10 dB example is free of such false alarms.

In Fig. 9, we display intermediate results of the algorithm for one trial at 0 dB. In Fig. 9(a), the estimated anomaly structure is shown after the first location test. Here we see that the algorithm has chosen to examine the 4 quarters of region A . Even at this coarse processing stage the structure with the largest amplitude is the quarter in which the true anomaly resides. In Figs. 9(b)–(e), the identified anomaly configuration is shown after one quarter, one half, three quarters and all of the processing for this trial. It is evident that well before the algorithm has terminated, the true anomaly is rather well localized. While there do exist falsely identified structures in the left side of Fig. 9(e) their amplitudes are far smaller than that of correctly identified anomaly. Thus, post-processing the results of the anomaly detection algorithm to remove structures whose magnitudes are in some sense small would further lower the false alarm rate. Finally, in Fig. 9(f) the results of the imaging algorithm are displayed for the same data set used to generate Figs. 9(a)–(e). In this case, the imaging algorithm does indicate the existence of a structure in the correct location. However, the geometric and amplitude information in this image is inferior to that generated by the detection algorithm, and as noted previously, the generation of this image took significantly longer than that of Fig. 9(e).

5.2 Experiment II: Two anomalies close to the left edge

In Fig. 7(c), the second anomaly configuration of interest in this paper is presented. Here we are interested in recovering the composition of two structures both of which are located close to the left vertical edge of region A . As discussed in Section 5.1, previous work has demonstrated that anomalous behavior is most easily detected near the sources and receivers. Thus, the structures in Fig. 7(c) are meant to test the ability of the detection algorithm to successfully isolate multiple anomalies.

For all trials and at all SNRs, Fig. 7(c) indicates that both structures are detected with a false alarm rate again less than 5%. Sample estimates are displayed for trials at 40 dB and 10 dB SNR respectively in Fig. 8(c)–(d). The reconstruction at 40 dB is basically identical to the true structures both in terms of geometry as well as amplitude. For the 10 dB case, both structures are correctly identified with the geometry of the lower anomaly rotated relative to that of the true structure. Again, the amplitudes are close to the true values of 1 S/m. In Fig. 10(a)–(e), the results of the detection algorithm at various stages of processing are displayed for one trial at 0 dB SNR. After the opening localization test, the most likely region for further decomposition corresponds to a subset of A containing both structures. As the algorithm proceeds, this area is further decomposed to better localize the two individual structures. Like the previous example, the falsely identified regions at the end of the algorithm are also those with the smallest amplitudes. Finally, in Fig. 10(f), the results of the imaging algorithm are displayed. For this particular data set, the imaging approach provides a severely degraded representation of the bottom structure and entirely misses the top anomaly.

5.3 Experiment III: Two separated anomalies

The anomaly configuration for the last example considered in this paper is shown in Fig. 7(e). From the previous experiments, the anomaly located close to the left edge of A should be relatively

easy to detect while the second structure positioned closer to the right vertical edge should be less detectable. This intuition is born out by the curves in Fig. 7(f) which show that both structures were detected for all trials at SNRs of 10, 20 and 40 dB. At 0 dB SNR, the smaller structure was identified in all 10 trials which the larger one was detected in 8 out of the 10 experiments. As in the first example, for the two cases in which this structures was not detected there was an estimated anomaly whose support was directly adjacent to that of the true structure. Finally, the false alarm rate for this example is again below 5%.

Sample results of the algorithm at 40 and 10 dB SNR are displayed in Fig. 8(e)–(f) where it is seen that at both noise levels, the structures are well localized. The amplitude of of the smaller structure is quite close to 1 while that of the second structure is about 1.5 for the 40 dB case and 2 at 10 dB. In Fig. 11, the results of one experiment at 0 dB SNR are plotted. Unlike the previous two cases, at the opening stages of the algorithm, the anomaly with the largest amplitude is located in a region not containing one of the true structures. Nonetheless, in the final estimate, the smaller structure is exactly identified and one of the estimated anomalies fully overlaps the second, larger structure. In contrast to the results of the detection algorithm, the imaging results for this problem, shown in Fig. 11(f) demonstrate that only a severely blurred representation of the left anomaly is available with essentially no information present regarding the structure on the right of region A .

6 Conclusion

In this paper, we have developed and tested an algorithm for the detection, localization, and characterization of a collection of anomalous areas in an overall region of interest from observations of scattered radiation obtained along the region’s boundary. Our approach employs a full, non-linear measurement model relating the anomalous structures to the observed fields and thus provides a

method for solving the target detection problem in cases where more common models based on the Born or Rytov linearizations are not valid. The algorithm makes no assumption concerning the number of anomalies present in the region or the amplitudes of these structures. While the method presented in this paper is adapted to the detection anomalies of which are rectangular in shape, as discussed in Section 2.2 the rectangular assumption is not required.

The anomaly detection algorithm is structured as a sequential hypothesis testing procedure which begins by localizing anomalous behavior to relatively large areas in the overall region. Subsequent tests are used to spatially refine the coarse-scale regions and prune away structures which are deemed unlikely to contain anomalies. Finally, we have developed a decision-theoretic scheme for the easy incorporation of prior information directly relevant to the anomaly detection problem into the spatial subdivision process. A MAP hypothesis testing framework was used to capture expectations regarding issues such as the number, shapes, and sizes of structures in the region.

The algorithm has been tested on a low frequency inverse electrical conductivity problem arising in geophysical exploration. The results in Section 5 demonstrate that high detection and low false alarm rates are obtainable even in highly noisy environments where more conventional imaging-type reconstruction algorithms fail to produce useful results. The decision-theoretic detection algorithm considered in this paper was far more computationally efficient than the full scale imaging algorithm and has a structure making it well suited for implementation in a parallel or distributed processing environment where these computational gains would be significantly enhanced.

References

- [1] Richard A. Albanese, Richard L. Medina, and John W. Penn. Mathematics, medicine and microwaves. *Inverse Problems*, 10:995–1007, 1994.
- [2] Yoram Bresler, Jeffrey A. Fessler, and Albert Macovski. A Bayesian approach to reconstruction from incomplete projections of a multiple object 3D domain. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(8):840–858, November 1989.

- [3] A. J. Devaney and G. A. Tsihrintzis. Maximum likelihood estimation of object location in diffraction tomography. *IEEE Trans. ASSP*, 39(3):672–682, March 1991.
- [4] P. E. Gill, W. Murry, and M. H. Wright. *Practical Optimization*. Academic, San Diego, Calif., 1981.
- [5] Gene H. Golub, Michael Heath, and Grace Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, May 1979.
- [6] Tarek M. Habashy, Ross W. Groom, and Brian R. Spies. Beyond the Born and Rytov approximations: A nonlinear approach to electromagnetic scattering. *J. Geophys. Res.*, 98(B2):1759–1775, February 1993.
- [7] Per Christian Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580, December 1992.
- [8] R. F. Harrington. *Field Computations by Moment Methods*. Macmillan, New York, 1968.
- [9] David Isaacson. Distinguishability of conductivities by electrical current computed tomography. *IEEE Trans. on Medical Imaging*, MI-5(2):91–95, June 1986.
- [10] Mark R. Luetttgen, W. Clem Karl, and Alan S. Willsky. Efficient multiscale regularization with applications to the computation of optical flow. *IEEE Trans. Image Proc.*, 3(1):41–64, 1994.
- [11] Eric L. Miller and Alan S. Willsky. Multiscale, statistical anomaly detection analysis and algorithms for linearized inverse scattering problems.
- [12] Eric L. Miller and Alan S. Willsky. Multiscale, statistically-based inversion scheme for the linearized inverse scattering problem. *IEEE Trans. Geosc. Remote Sens.*, 34(2):346–357, March 1996.
- [13] Eric L. Miller and Alan S. Willsky. Wavelet-based methods for the nonlinear inverse scattering problem using the Extended Born Approximation. *Radio Science*, 31(1):51–67, Jan–Feb 1996.
- [14] Leon Peters Jr., Jeffrey J. Daniels, and Jonathan D. Young. Ground penetrating radar as a subsurface environmental sensing tool. *Proc. IEEE*, 82(12):1802–1822, December 1994.
- [15] David J. Rossi and Alan S. Willsky. Reconstruction from projections based on detection and estimation of objects—parts I and II: Performance analysis and robustness analysis. *IEEE Trans. on ASSP*, ASSP-32(4):886–906, August 1984.
- [16] Ken Sauer and Bede Liu. Image reconstruction from a limited number of projections using multiple object detection/estimation. In *ICASSP 90*, volume 4, pages 1861–1864, 1990.
- [17] Alon Schatzberg, Anthony J. Devaney, and Alan J. Witten. Estimating target location from scattered field data. *Signal Processing*, 40:227–237, 1994.
- [18] Carlos Torres-Verdín and Tarek M. Habashy. Rapid 2.5-D forward modeling and inversion via a new nonlinear scattering approximation. *Radio Sci.*, 29(4):1051–1079, July–August 1994.
- [19] Harry L. Van Trees. *Detection, Estimation and Modulation Theory: Part I*. John Wiley, New York, 1968.

- [20] Song Wang, Bede Liu, and Sanjeev Kulkarni. Image reconstruction from a limited number of projections: Detection/estimation of multiple discs with unknown radii. In *Proceedings ICIP-94*, volume II, pages 854–858. IEEE, IEEE Computer Society Press, 1994.

7 Figure and Table Captions

Figure 1 Inverse conductivity problem of interest in this paper.

Figure 2 General structure of anomalous regions of interest in this paper. The magnitudes, a_1 and a_2 of the two anomalies shown here are proportional to the color of the corresponding rectangles.

Figure 3 Pseudo-code for anomaly detection algorithm

Figure 4 Hypothesized regions of support used in localization-type hypothesis tests.

Figure 5 Penalty functions used in construction of prior models for anomaly detection algorithm.

Figure 6 Additional regions used in initial localization-type hypothesis test.

Figure 7 Anomaly structure to be reconstructed and associated performance curves for all examples in this paper.

Figure 8 Sample reconstructions for all anomaly configurations at 40 and 10 dB SNR. Dotted lines indicate true positions of anomalies

Figure 9 Reconstructions for anomaly in Fig. 7(a) in (a)–(e) and estimate obtained using imaging algorithm of [13]. Dotted lines indicate true positions of anomalies

Figure 10 Reconstructions for anomaly in Fig. 7(c) in (a)–(e) and estimate obtained using imaging algorithm of [13]. Dotted lines indicate true positions of anomalies.

Figure 11 Reconstructions for anomaly in Fig. 7(e) in (a)–(e) and estimate obtained using imaging algorithm of [13]. Dotted lines indicate true positions of anomalies.

Table 1 Data set definitions for observation processes of interest in the paper.

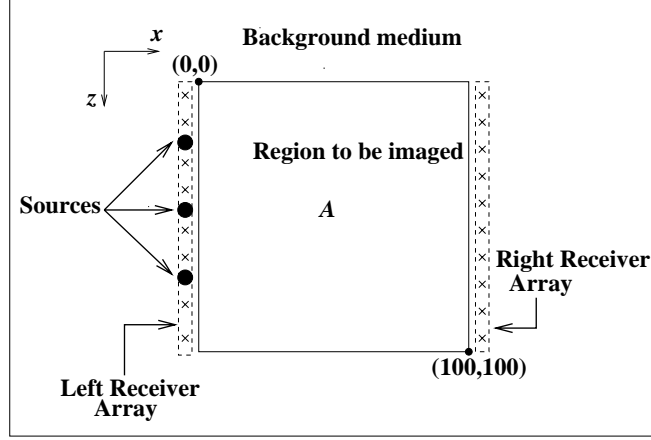


Figure 1:

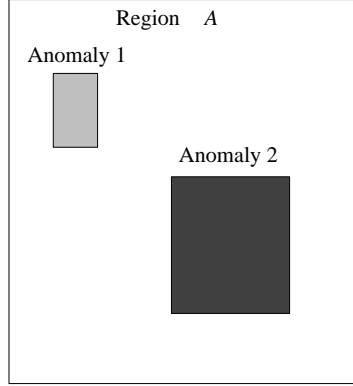


Figure 2:

Experiment number	Source Position	Frequency of source (kHz)	Receiver Array
1–6	1–6	$f_{\text{LO}} = 0.1$	Right
7–12	1–6	$f_{\text{MID}} = 1.0$	Left
13–18	1–6	$f_{\text{HI}} = 10.0$	Left

Table 1:


```

INITIALIZE ANOMALY ESTIMATES
 $\mathbf{B}_{new} := \mathbf{1}$ ;  $\mathbf{B}_{old} := \mathbf{B}_{new}$ ;
 $\mathbf{a}_{new} := 0$ ;  $\mathbf{a}_{old} := 0$ ;

PERFORM INITIAL LOCALIZATION
 $[\mathbf{B}_{new}, \mathbf{a}_{new}] := \text{extended\_localization\_test}(\mathbf{B}_{old}, \mathbf{a}_{old})$ ;
 $\mathbf{B}_{old} := \mathbf{B}_{new}$ ;  $\mathbf{a}_{old} := \mathbf{a}_{new}$ ;

START PROCESSING LOOP
done_all := false;
while (done_all  $\neq$  true) do
    PERFORM A LOCALIZATION TEST.
     $[\mathbf{B}_{new}, \mathbf{a}_{new}] := \text{localization\_test}(\mathbf{B}_{old}, \mathbf{a}_{old})$ ;

    IF NO MORE LOCALIZATION, THEN SET FLAG.
    OTHERWISE UPDATE  $B$  AND CONTINUE.
    if (identical_anomalies( $\mathbf{B}_{new}, \mathbf{B}_{old}$ ))
        no_localize := true;
    else
         $\mathbf{B}_{old} := \mathbf{B}_{new}$ ;  $\mathbf{a}_{old} := \mathbf{a}_{new}$ ;
        no_localize := false;
    endif

    PRUNE AWAY FALSE ALARMS ONE STRUCTURE AT A TIME
    done_prune := false;
    while (done_prune  $\neq$  true) do
         $[\mathbf{B}_{new}, \mathbf{a}_{new}] := \text{prune\_test}(\mathbf{B}_{old}, \mathbf{a}_{old})$ ;
        if (identical_anomalies( $\mathbf{B}_{new}, \mathbf{B}_{old}$ )  $\vee$  (number_anomalies( $\mathbf{B}_{new}$ ) == 0))
            done_prune := true;
            no_prune := true;
        else
             $\mathbf{B}_{old} := \mathbf{B}_{new}$ ;  $\mathbf{a}_{old} := \mathbf{a}_{new}$ ;
        endif
    end

    IF NO NEW LOCALIZATION AND NO PRUNING THEN TERMINATE
    if ((no_localize == true)  $\wedge$  (done_prune == true))
        all_done == true;
    endif
end

```

Figure 3:

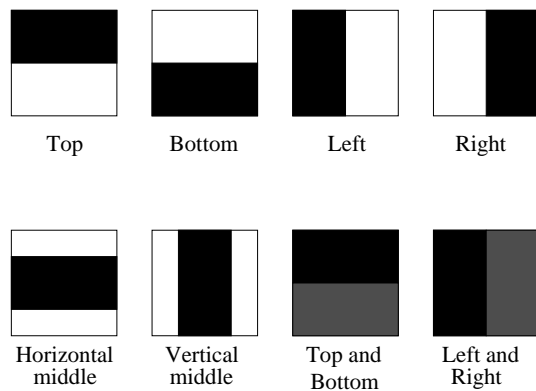


Figure 4:

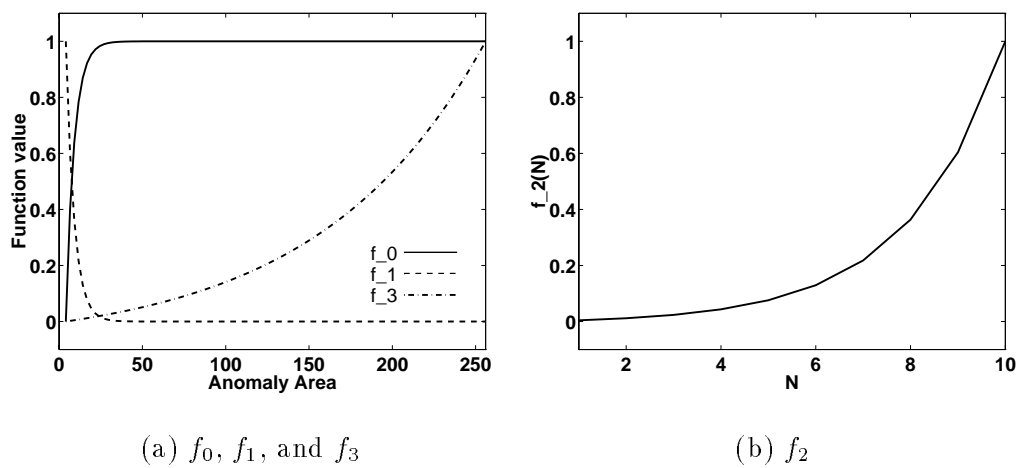


Figure 5:

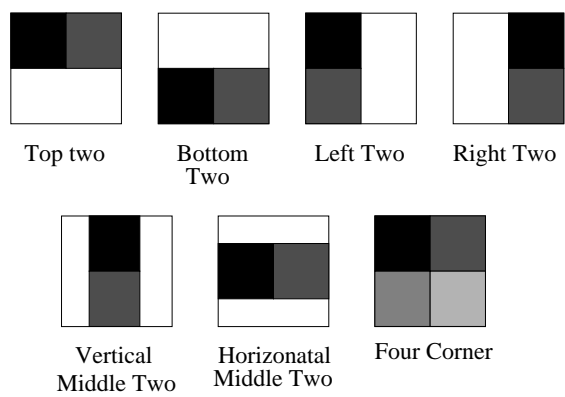
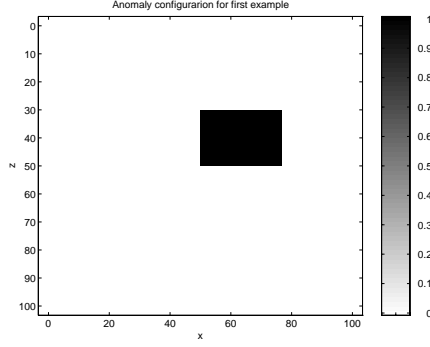
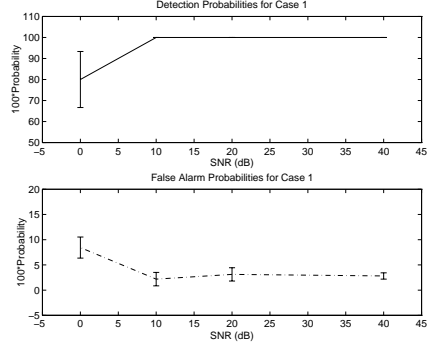


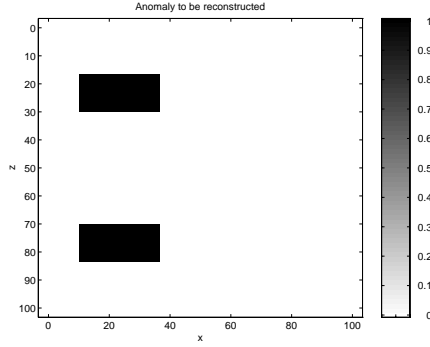
Figure 6:



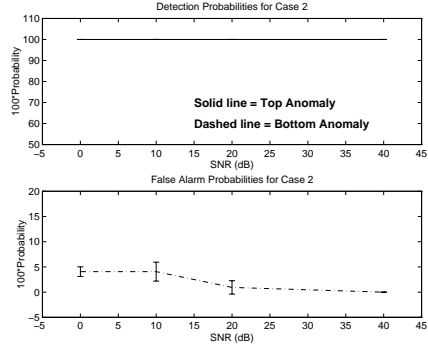
(a) Anomaly distribution for first example.



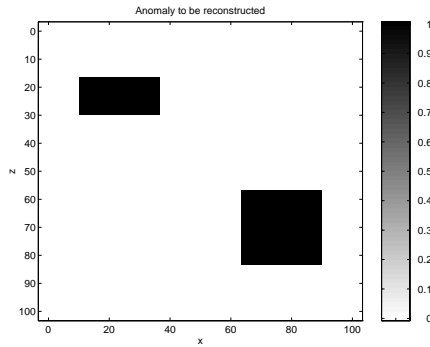
(b) Detection (top) and false alarm (bottom) probabilities for first example



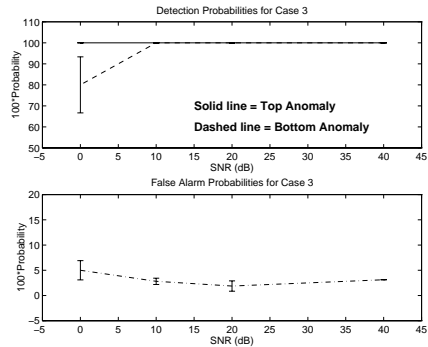
(c) Anomaly distribution for second example.



(d) Detection (top) and false alarm (bottom) probabilities for second example.

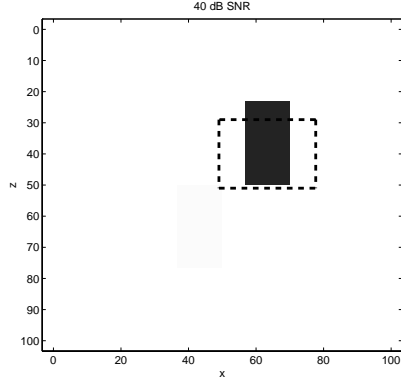


(e) Anomaly distribution for third example.

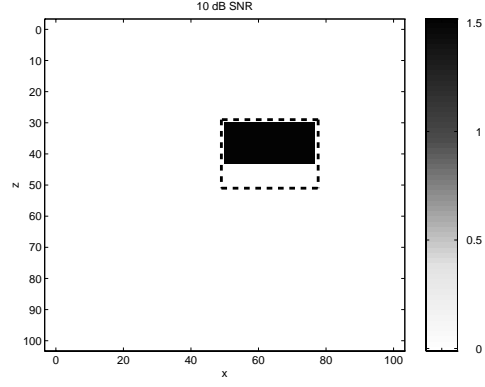


(f) Detection (top) and false alarm (bottom) probabilities for third example

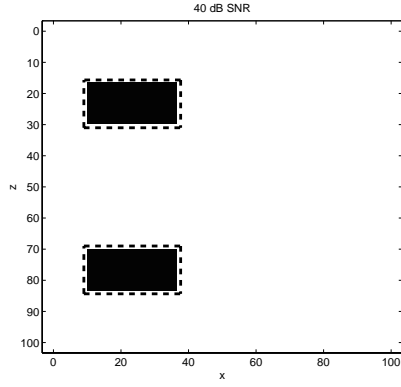
Figure 7:



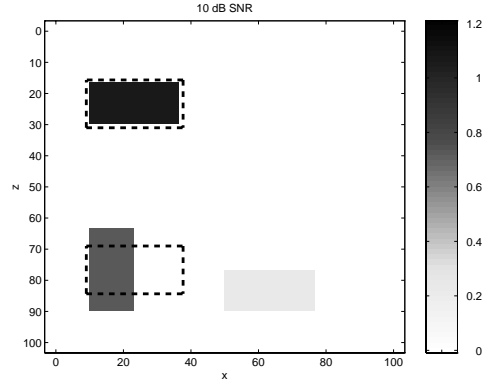
(a) Sample reconstruction at 40 dB for anomaly in Fig. 7(a)



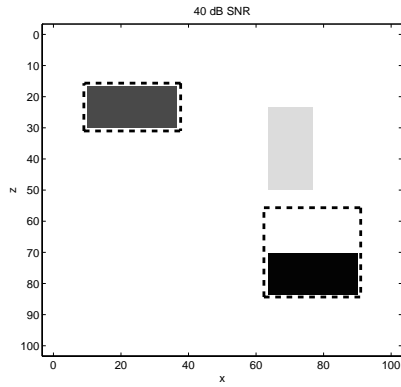
(b) Sample reconstruction at 10 dB for anomaly in Fig. 7(a)



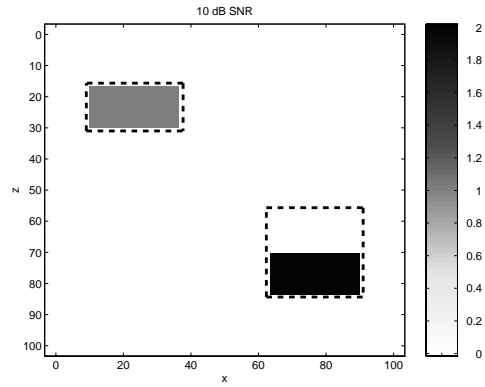
(c) Sample reconstruction at 40 dB for anomaly in Fig. 7(c)



(d) Sample reconstruction at 10 dB for anomaly in Fig. 7(c)

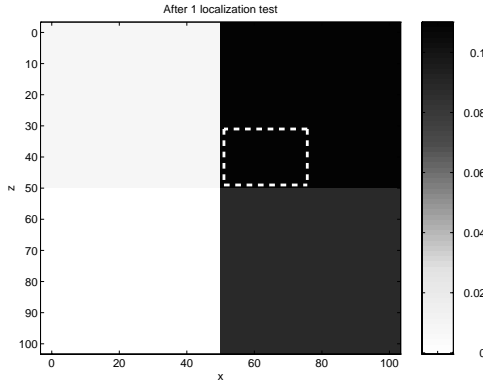


(e) Sample reconstruction at 40 dB for anomaly in Fig. 7(e)

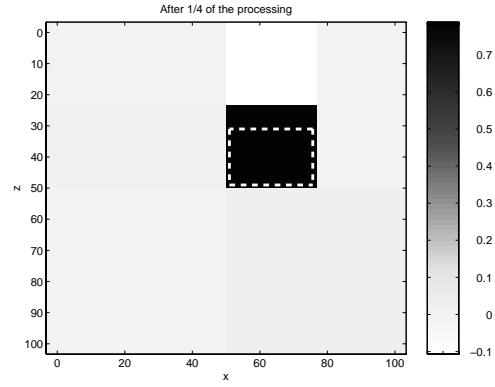


(f) Sample reconstruction at 10 dB for anomaly in Fig. 7(e)

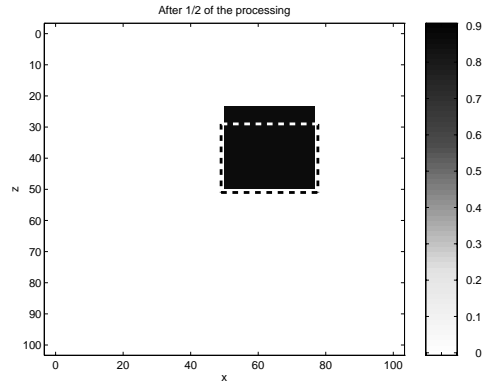
Figure 8:



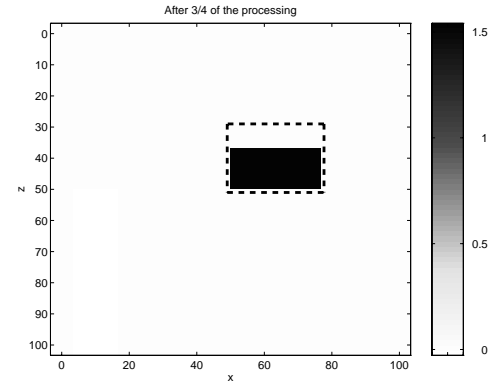
(a) Estimate after initial localization test for anomaly in Fig. 7(a)



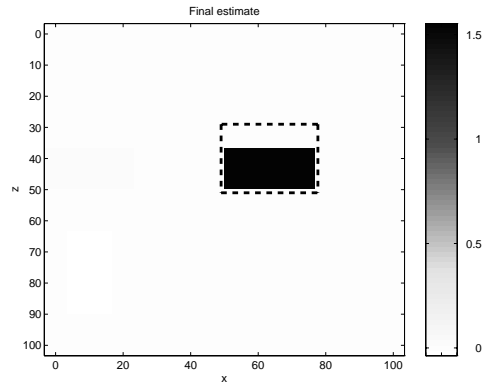
(b) Estimate after 1/4 of detection algorithm for anomaly in Fig. 7(a)



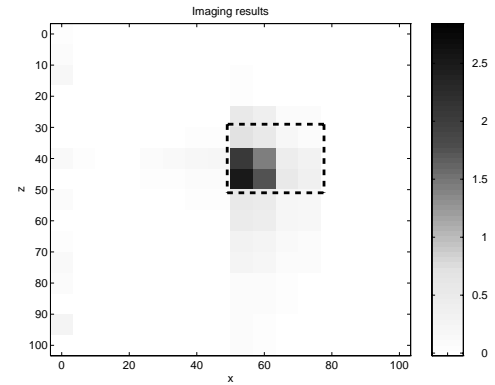
(c) Estimate after 1/2 of detection algorithm for anomaly in Fig. 7(a)



(d) Estimate after 3/4 of detection algorithm for anomaly in Fig. 7(a)

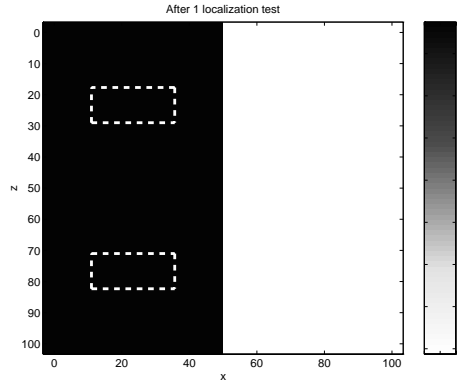


(e) Final estimate for anomaly in Fig. 7(a)

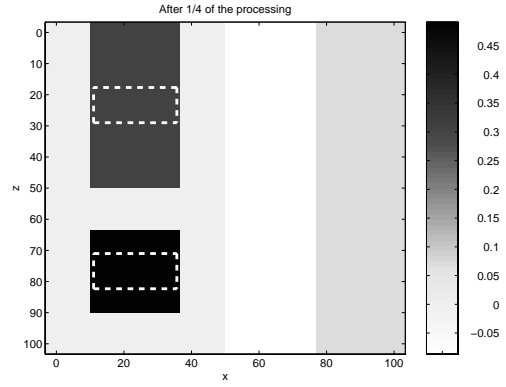


(f) Estimate after initial localization test for anomaly in Fig. 7(a)

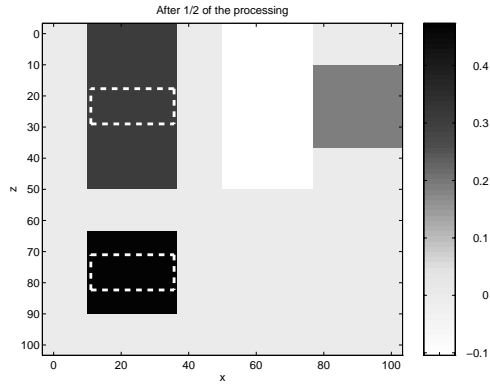
Figure 9:



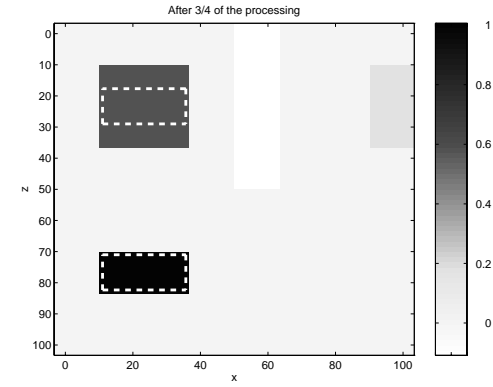
(a) Estimate after initial localization test for anomaly in Fig. 7(c)



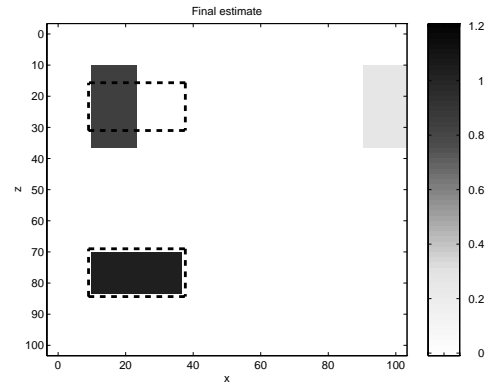
(b) Estimate after 1/4 of detection algorithm for anomaly in Fig. 7(c)



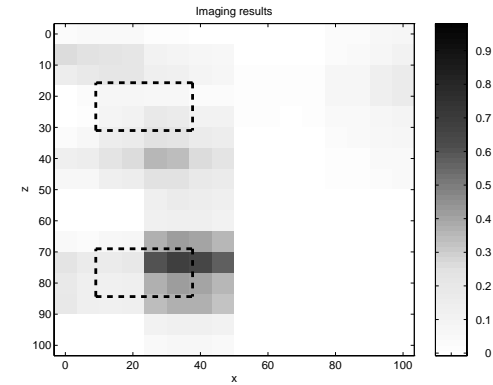
(c) Estimate after 1/2 of detection algorithm for anomaly in Fig. 7(c)



(d) Estimate after 3/4 of detection algorithm for anomaly in Fig. 7(c)

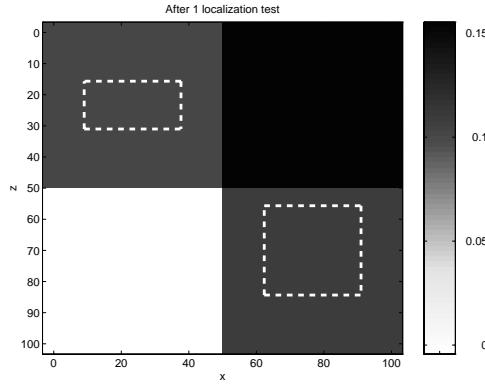


(e) Final estimate for anomaly in Fig. 7(c)

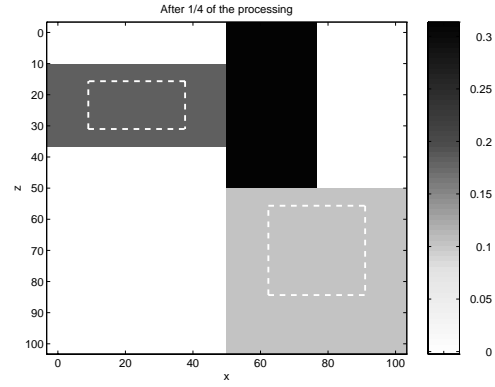


(f) Estimate after initial localization test for anomaly in Fig. 7(c)

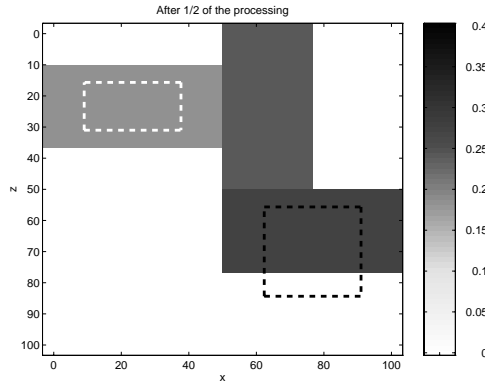
Figure 10:



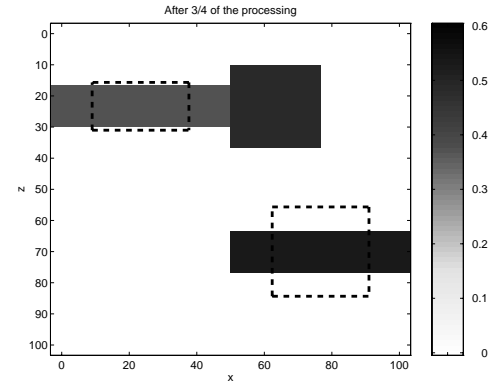
(a) Estimate after initial localization test for anomaly in Fig. 7(e)



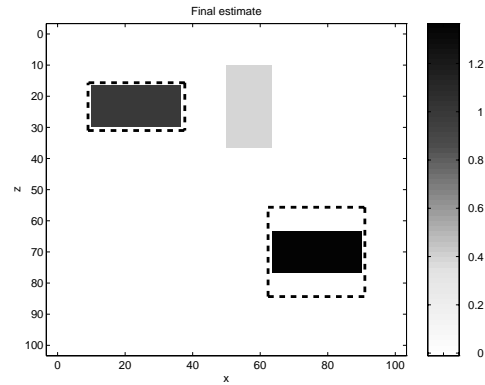
(b) Estimate after 1/4 of detection algorithm for anomaly in Fig. 7(e)



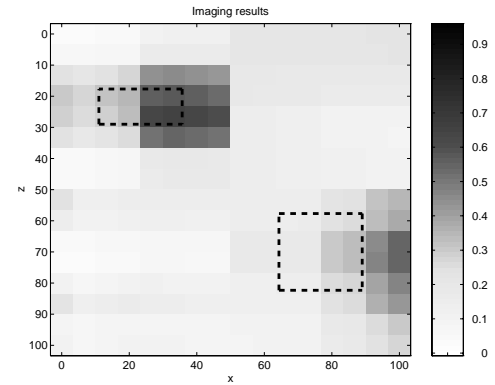
(c) Estimate after 1/2 of detection algorithm for anomaly in Fig. 7(e)



(d) Estimate after 3/4 of detection algorithm for anomaly in Fig. 7(e)



(e) Final estimate for anomaly in Fig. 7(e)



(f) Estimate after initial localization test for anomaly in Fig. 7(e)

Figure 11:

Recursive T-Matrix Algorithm for Multiple Metallic Cylinders *

Adnan Şahin and Eric L. Miller
Center for Electromagnetics Research,
235 Forsyth Building, 360 Huntington Ave.
Northeastern University, Boston, MA 02115
Telephone: (617) 373-8386
Telefax : (617) 373-8627
email: adnan@cdsp.neu.edu

February 14, 1997

Abstract

We present a new application of the recursive T-matrix algorithm to calculate the scattered field from a single or multiple metallic cylinders of arbitrary shapes. Using the equivalence theorem each metallic object is replaced with small metallic cylinders along its perimeter, then scattered fields are calculated using the recursive T-matrix algorithm. Results are verified with those in the literature and analytical calculations.

Keywords: Scattering, recursive T-matrix algorithm

*This work was supported in part DOE contract DE-FC07-95ID13395, NSF Grant MIP-9623721, and by sub-contract GC123920NDG from Boston University under the AFOSR MURI Program on Reduced Signature Target Recognition

1 Introduction

The transition matrix (T-matrix) technique is known to be an efficient electromagnetic forward solver for scattering problems involving objects with simple shapes. Finite difference and method of moments (MoM) techniques both discretize the entire region of interest resulting in a large number of unknowns, whereas T-matrix methods replace this discretization with harmonic expansions of the fields thereby reducing the number of unknowns for numerous problems. Waterman [1] developed the T-matrix technique for single metallic or dielectric scatterers. Peterson et.al. [2] introduced an iterative algorithm which finds the scattering due to multiple scatterers. Recently Chew and co-workers have developed a number of fast, recursive T-matrix algorithms for determining the scattered fields in a variety of scenarios [3–9]. Among this work, in [4], problems involving electrically large dielectric objects are considered. By tessellating the objects into many small cylindrical sub-scatterers and using multipole expansions of the fields for each sub-scatterer the authors arrive at a highly efficient, T-matrix based algorithm for computing the scattered fields. In [7], Chew et.al. consider a scattering problem involving a group of metallic strips. Here the method of moments is used to compute the T-matrices for each, individual strip and the same recursion as in [6] is employed to solve the overall, multi-object scattering problem. In [8] the scattered field from an ogive shaped scatterer with metallic and dielectric parts is found using the recursive T-matrix algorithm. In that paper, as in [7], the metallic scatterer is decomposed into a collection of strips arrayed about the boundary and the T-matrices for the individual strips are found using the method of moments.

In this letter we consider an alternative sub-scatterer discretization for metallic objects from that in [7,8]. Instead of using metallic strips to model the perimeter of scatterers, we use metallic cylinders (similar to the concept in [4]) placed about the perimeter and employ the same recursive

algorithm given in [4] to calculate the scattered field. As a result, we obtained an accurate, efficient forward solver which does not require the use of method of moments to form the single scatterer T-matrices. Rather we obtained these quantities by using closed form, low order harmonic expansions associated with the small metallic cylinders. We apply this method to single electrically large metallic objects and verify the results with those in literature and analytical results. Additionally we demonstrate the usefulness of the method for the multi-object case by verifying against previously published results.

2 Recursive T-matrix Algorithm

The algorithms in [3,4,6,7,9] are order recursive methods for constructing the T-matrix for a multi-object scattering problem given the T-matrices for each individual object. The algorithm uses the basic principle of the single scatterer T-matrix formulas in that for each object, the scattered fields from others are assumed a part of a total incident field. The recursion starts with the T-matrices of individual scatterers, then one by one scatterers are incorporated into the equation and the T-matrices are updated until the final form of the T-matrix, including all multiple scattering effects, is obtained.

Formally, for L scatterers, the harmonic expansion of scattered field can be written as [4]:

$$\psi^{sc}(\underline{r}) = \sum_{i=1}^L \underline{\psi}^T(\underline{r}_i) \mathbf{T}_{i(L)} \beta_{i,0} \underline{a} \quad (1)$$

where $\mathbf{T}_{i(L)}$ is the T-matrix for i th object in the presence of L scatterers, \underline{a} is the coefficient vector used in the expansion of the incident plane wave in terms of cylindrical basis functions and $\beta_{i,0}$ is the translation matrix used to translate same type basis functions between reference coordinate systems. (The translation matrices $\beta_{i,0}$ contain Bessel functions and complex exponentials. For details about these matrices see [2,3].) Expansion of the scattered field in (1) is valid if all observation points

are outside the smallest circle enclosing all scatterers. Following Chew's derivation, the recursive construction of $\mathbf{T}_{i(L)}$ can be written as [4, eq.10-11] :

$$\mathbf{T}_{n+1(n+1)}\boldsymbol{\beta}_{n+1,0} = \left[\mathbf{I} - \mathbf{T}_{n+1(1)} \sum_{i=1}^n \boldsymbol{\alpha}_{n+1,i} \mathbf{T}_{i(n)} \boldsymbol{\beta}_{i,0} \boldsymbol{\alpha}_{0,n+1} \right]^{-1} \mathbf{T}_{n+1(1)} \left[\boldsymbol{\beta}_{n+1,0} + \sum_{i=1}^n \boldsymbol{\alpha}_{n+1,i} \mathbf{T}_{i(n)} \boldsymbol{\beta}_{i,0} \right] \quad (2)$$

and

$$\mathbf{T}_{i(n+1)}\boldsymbol{\beta}_{i,0} = \mathbf{T}_{i(n)}\boldsymbol{\beta}_{i,0} + \mathbf{T}_{i(n)}\boldsymbol{\beta}_{i,0} \boldsymbol{\alpha}_{0,n+1} \mathbf{T}_{n+1(n+1)}\boldsymbol{\beta}_{n+1,0} \quad (3)$$

where $n = 1, 2, \dots, L$, $i = 1, 2, \dots, n$ and $\boldsymbol{\alpha}_{n,i}$ is another translation matrix [3]. The recursion starts with the individual T-matrices, $\mathbf{T}_{i(1)}$, of the scatterers, i.e. the T-matrix of the i th scatterer when there are no other scatterers in the medium.

Theoretically the matrices $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, \mathbf{T} are of infinite dimension. The T-matrix algorithms truncate these matrices with finite values N and M such that the residual error is below the machine precision or acceptable levels. Here N represents the number of harmonics used to expand the fields at the scattering origin and M represents the number of harmonics used to expand the fields in the objects' local coordinate systems. Thus, the T-matrix is of size $M \times M$, $\boldsymbol{\beta}_{i,0}$ is of size $M \times N$ and $\boldsymbol{\alpha}_{i,n+1}$ is of size $M \times M$. It has been shown that computational complexity of (2)-(3) is $O(L^2 M^2 N)$ for L scatterers [4].

The contribution of this letter is to show that, based on the equivalence theorem, recursive T-matrix algorithms can be used to calculate the scattered fields from metallic objects by placing small metallic cylinders on their perimeter. Traditionally, the recursive T-matrix algorithm has been applied in one of two manners. In the case of dielectric scatterers, the whole object was decomposed into small cylinders, low order harmonic expansions were used to represent the fields from each object, and the recursive algorithm was used to solve the scattering problem. For metallic objects, the equivalence theorem was used to decompose the surface of the scatterer into small

strips, moment methods were then employed to find the individual T-matrices for each strip, and the same T-matrix recursions were used to solve the overall scattering problem. The objective of this letter is to show that one may make use of the cylinder approach (Fig.1) as well for the metallic scattering problem and still obtain highly accurate solutions. In particular, by using cylinders, one may employ the closed-form harmonic expansion method to find the individual scatterers thereby avoiding the moment method computation. In the next section, we will give examples of scattering from circular and rectangular cylinders and the results are verified with those in the literature or analytical calculations.

3 Examples and Discussions

In this section we verify that replacing metallic objects with small metallic cylinders along their perimeters actually produces the results reported in the literature or results obtained analytically. First we define the terms used in this section. The normalized scattering field pattern is defined as:

$$F(\phi) = 10 \log_{10} \left\{ \lim_{r \rightarrow \infty} 2\pi r \frac{|\psi^{sca}(\underline{r})|^2}{\max\{|\psi^{sca}(\underline{r})|^2\}} \right\}. \quad (4)$$

Normalized power density, or the “gain”, is defined as:

$$G_E(\phi) = \lim_{r \rightarrow \infty} \frac{|\psi^{sca}(\underline{r})|^2}{\frac{1}{2\pi} \int_0^{2\pi} |\psi^{sca}(\underline{r})|^2 d\phi}. \quad (5)$$

In all examples the cylinders are embedded in free space with an E_z polarized plane wave incident on them. The first example is the scattering from a single circular cylinder of radius 0.8λ ($ka = 5$). As seen in Fig. 2(a), the cylinder is approximated by 60 smaller cylinders along its circumference which corresponds to a sampling 12 cylinders per wavelength. The normalized scattering field pattern, $F(\phi)$, obtained from the recursive algorithm is plotted against the analytical solution in Fig. 2(b). The second example is the scattering from two circular cylinders with radii of 0.8λ ($ka = 5$) and separated by a distance 2.55λ ($kd = 16$). As in previous example, each cylinder is

approximated by 60 small metallic cylinders with 12 cylinders per wavelength. Fig. 3(a) shows the scattering geometry and Fig. 3(b) compares the normalized scattering patterns obtained using the recursive T-matrix algorithm with those given in [10]. The last example shows the normalized power densities for a slanted rectangular cylinder for two different sizes. The geometry is shown in Fig. 4(a) and the far field power densities, $G_E(\phi)$, for $ka = 3$ and $ka = 5$ ($a = 0.48\lambda$ and $a = 0.8\lambda$, both with $a = 2b$) are depicted in Fig. 4(b). For both cases, the perimeter is sampled at approximately 13 cylinders per wavelength. In this figure, the scattering patterns are compared with the results reported in [11]. As these plots have shown, the scattered fields from metallic objects can be found by replacing these objects with smaller cylinders along the perimeter and then using the recursive T-matrix algorithm of [4].

4 Conclusion

In this letter an alternative discretization along the perimeter of metallic scatterers is used with recursive T-matrix algorithm to calculate the scattered field. The results are verified with previous work.

References

- [1] P.C. Waterman. New formulation of acoustic scattering. *J. Acoust. Soc. Am.*, 45(6):1417–1429, 1969.
- [2] B. Peterson S. Ström. Matrix formulation of acoustic scattering from an arbitrary number of scatterers. *J. Acoust. Soc. Am.*, 56(3):771–780, September 1974.
- [3] W.C. Chew. *Waves and Fields in Inhomogeneous Media*. Van Nostrand Reinhold, 1990.
- [4] Y.M. Wang W.C. Chew. An efficient algorithm for solution of a scattering problem. *Microwave and Optical Technology Letters*, 3(3):102–106, March 1990.
- [5] W.C. Chew J.A. Friedrich R. Geiger. A multiple scattering solution for effective permittivity of a sphere mixture. *IEEE Trans. Geoscience and Remote Sensing*, 28(2):207–214, March 1990.
- [6] W.C. Chew Y.M. Wang. A fast algorithm for solution of a scattering problem using a recursive aggregate τ matrix method. *Microwave and Optical Technology Letters*, 3(5):164–169, May 1990.
- [7] W.C. Chew Y.M. Wang L. Gürel. Recursive algorithm for wave-scattering solutions using windowed addition theorem. *Journal of Electromagnetic Waves and Applications*, 6(11):1537–1560, 1992.
- [8] J.-H. Lin W.C. Chew. Scattering solution of oblong, dielectric coated, metallic scatterers by recursive algorithm. *Electronics Letters*, 28(2):185–187, January 1992.
- [9] W.C. Chew L. Gürel Y.M. Wang G. Otto R.L. Wagner Q.H. Liu. A generalized recursive algorithm for wave-scattering solutions in two dimensions. *IEEE Trans. on Microwave Theory and Techniques*, 40(4):716–722, April 1992.
- [10] A.Z. Elsherbeni M. Hamid. Scattering by parallel conducting circular cylinders. *IEEE Trans. on Antennas and Propagation*, 35(3):355–358, March 1987.
- [11] K.K. Mei J.G. Van Bladel. Scattering by perfectly conducting rectangular cylinders. *IEEE Trans. on Antennas and Propagation*, 11:185–192, 1963.

Figure Captions

Figure 1: Tessellation of metallic cylinders along their perimeters

Figure 2: Comparison of normalized scattering field pattern calculated using the recursive T-matrix algorithm with the analytically calculated one.

Figure 3: Comparison of normalized scattering field pattern calculated using the recursive T-matrix algorithm with [10].

Figure 4: Comparison of normalized power density calculated using the recursive T-matrix algorithm with [11].

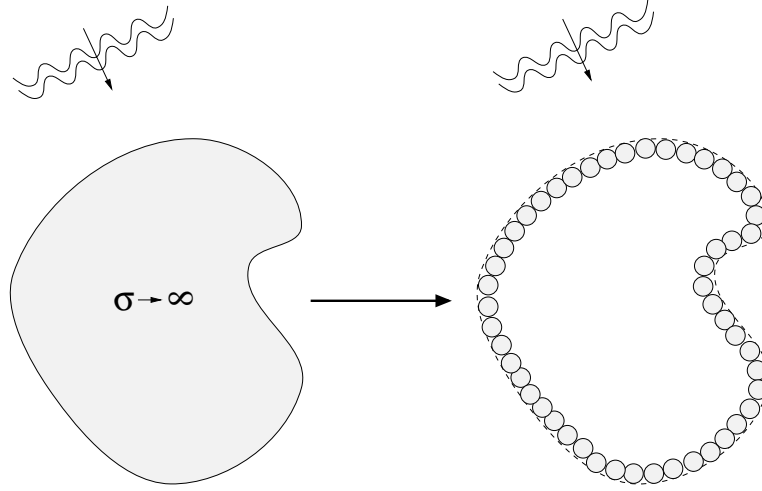


Figure 1:

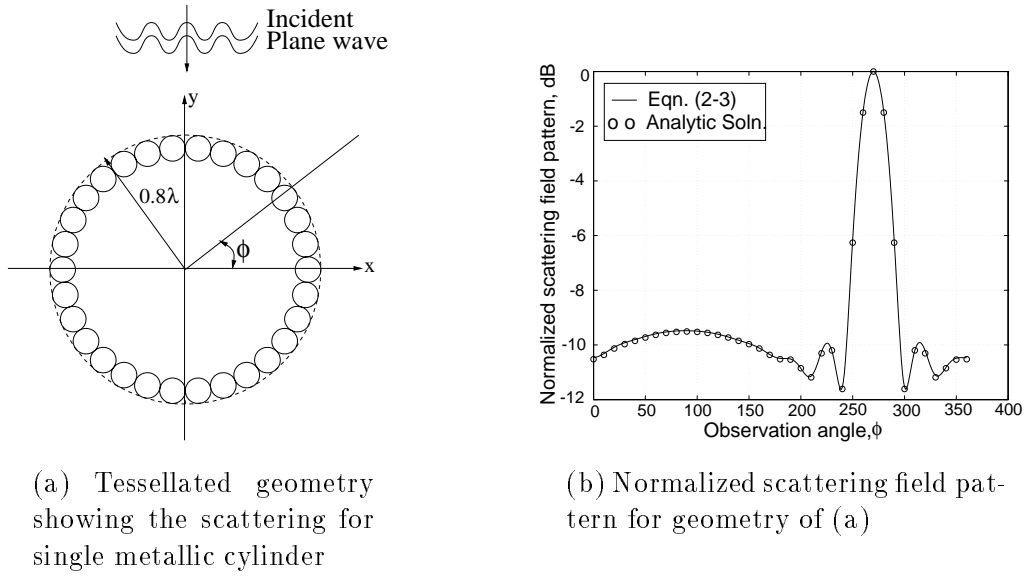
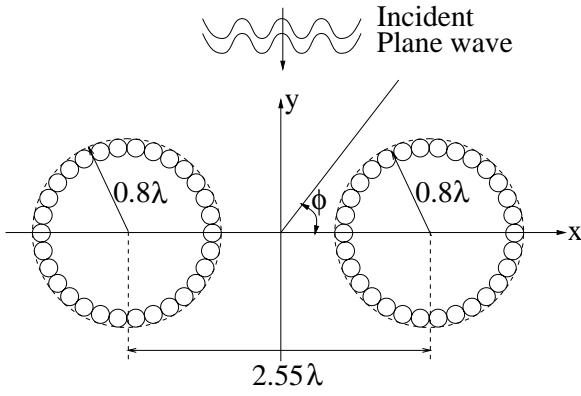
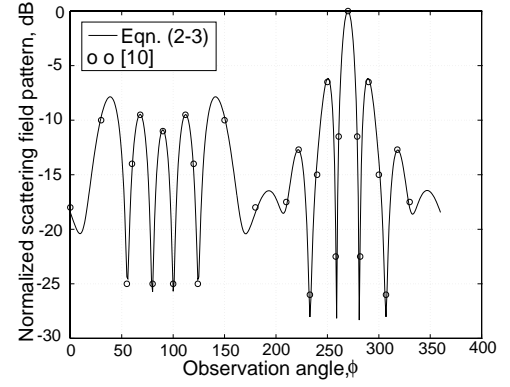


Figure 2:

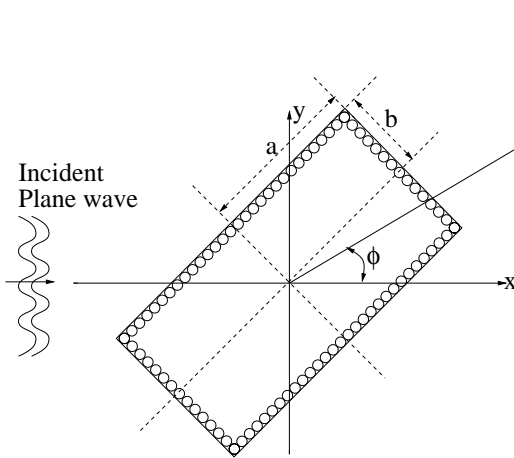


(a) Tessellated geometry showing the scattering for two metallic cylinders

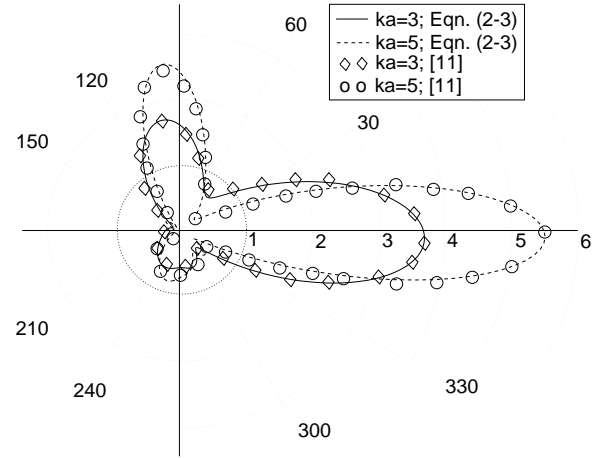


(b) Normalized scattering field pattern for geometry of (a)

Figure 3:



(a) Tessellated geometry showing the scattering for 45° slanted metallic rectangle



(b) Polar plot showing the normalized power density vs. ϕ for geometry of (a)

Figure 4:

Parallel Implementation of the Steepest Descent Fast Multipole Method (SDFMM) On a Beowulf Cluster for Subsurface Sensing Applications

D. Jiang¹, W. Meleis¹, M. El-Shenawee², E. Mizan¹, M. Ashouei¹ and C. Rappaport¹.

¹Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
meleis@ece.neu.edu

²Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701
magda@uark.edu

Abstract—We present the parallel, MPI-based implementation of the SDFMM computer code using a thirty two-node Intel Pentium-based Beowulf cluster. The SDFMM is a fast algorithm that is a hybridization of the Method of Moment (MoM), the Fast Multipole Method (FMM) and the Steepest Descent Integration rule (SDP), which is used to solve large-scale linear systems of equations produced electromagnetic scattering problems. An overall speedup of 7.2 has been achieved on the 32-processor Beowulf cluster and a significant reduced runtime is achieved on the 4-processor 667MHz Alpha workstation.

I. INTRODUCTION

The SDFMM was originally developed at the University of Illinois at Urbana Champaign to analyze large-scale three dimension (3-D) scattering problems [1]–[3]. Recently its computer code has been successfully modified to handle subsurface sensing applications, in particular, the scattering from a PEC and/or penetrable spheroid buried under a two dimensional randomly rough ground surface [4]–[5]. The SDFMM has computational complexity for CPU time and memory equal to only $O(N)$ per iteration versus $O(N^2)$ for the MoM, where N is the total number of the

unknowns [1]. The significantly reduced complexity of the SDFMM over several other computational electromagnetics techniques has enabled efficient Monte Carlo simulation studies [5]. Additional speedup is desirable for increased Monte Carlo sample size or for inverse scattering applications. In this work, we used the MPI library for the parallel implementation of the SDFMM code [6]–[8].

II. PARALLELIZATION

The SDFMM is used to solve the linear system of equations given by [1]–[5]:

$$\bar{\bar{Z}} \bar{I} = \bar{V} \quad (1a)$$

where $\bar{\bar{Z}}$ is the impedance matrix, \bar{I} is the vector of unknown coefficients of the electric and magnetic surface currents and \bar{V} is associated with the incident waves on the rough ground surface.

The matrix $\bar{\bar{Z}}$, which is filled in MoM formulations, becomes sparse with SDFMM and the system of equations in (1a) can be written as:

$$\bar{\bar{Z}}' \bar{I} + \bar{\bar{Z}}'' \bar{I} = \bar{V} \quad (1b)$$

The sparse matrix $\bar{\bar{Z}}'$ has its non-zero elements calculated and stored using the conventional MoM, which are then multiplied by the vector \bar{I} (near field interactions) while the matrix–vector multiply $\bar{\bar{Z}}'' \bar{I}$ is computed in one step without calculating or storing any elements of the matrix $\bar{\bar{Z}}''$. This is achieved by using the FMM hybridized with the SDP integration rule.

The following three bottlenecks in the SDFMM code can benefit from being parallelized: (i) the subroutines that calculates the elements of the sparse matrix $\bar{\bar{Z}}'$; (ii) the subroutines that execute the

matrix vector multiplication $\bar{\bar{Z}}'\bar{I}$ in each iteration of the solver; and (iii) the subroutines that execute the fast multipole method for $\bar{\bar{Z}}''\bar{I}$ (far field interactions).

The computer code has been parallelized by exploiting the underlying available data parallelism. The key data structure in subroutine (i) is the sparse matrix $\bar{\bar{Z}}'$, which is stored as blocks of nonzero elements. These blocks are distributed among all processors, and no additional communication is needed. When this routine is parallelized we achieved near-linear speedups on 32 processors. In the matrix-vector multiplication $\bar{\bar{Z}}'\bar{I}$, the computation is parallelized by distributing \bar{I} to all processors in each iteration. The resulting vector components produced by the multiplication are then distributed to all processors. For bottleneck (iii), there are two involved subroutines to compute the far field interactions consisting of a series of loops with complex interdependences. Each loop is separately parallelized, with collective communication used to distribute the results to all processors after executing each subroutine. In addition these two subroutines are executed in parallel, followed by subsequent distribution of the results to all processors. Load balance between these two subroutines is achieved using a detailed performance model based on the serial execution time of each routine, the time required for collective communication operations, and the amount of communication overhead needed. The structure of the parallelized SDFMM application is shown in Fig. 1.

We evaluated the parallel implementation of the SDFMM computer code on a 32-node Intel Pentium-based Beowulf cluster. Thirty one nodes of the Beowulf cluster are 350MHz Intel Pentium IIs with 256 MB of RAM and one node is a 4x450MHz Intel Pentium II Xeon shared memory processor with 2GB of RAM. The nodes are connected to a 100 BaseTX Ethernet network and they use the SuSE 6.1 operating system with Linux kernel 2.2.13, and the MPICH 1.2.1 implementation of the MPI library. We also tested the parallelized code on a 4-node shared

memory Compaq Alpha-based workstation (667Mhz Alpha 21264) of 16GB total RAM. The processor uses the UNIX OSF/1 V5.1 operating system with the MPICH 1.1.2 MPI library.

Our benchmark includes three small-scale cases executed on the 256MB Intel cluster, and in addition one moderate-scale case that is executed on the Alpha workstation. All results obtained by executing the parallel version of the code are validated with those computed by the serial version of the code [4]–[5]. The scattering problem configurations used in [5] are employed here, but for only one rough surface realization. The rough ground (characterized by Gaussian statistics with zero mean for the height), is described by the rms height (σ) and the correlation length (l_c). In all cases, the relative dielectric constant of the ground soil (dry sand) and the penetrable buried object (TNT in a land mine) are $\epsilon_r = 2.5 - j0.18$ and $\epsilon_r = 2.9 - j0.0092$, respectively, and the ground correlation length is $l_c = 0.5\lambda_0$. A Gaussian beam with horizontal polarization is employed for the incident waves [5]. In Case 2, the buried sphere has radius of $a = 0.16\lambda_0$ with burial depth equal to $z = -0.32\lambda_0$ measured from its center to the mean plane of the ground while in Case 3 and 4 the buried spheroid has dimensions $a = 0.3\lambda_0$ and $b = 0.15\lambda_0$, and is buried at $z = -0.3\lambda_0$. The ground dimensions are $3\lambda_0 \times 3\lambda_0$ in Cases 1–3 and $8\lambda_0 \times 8\lambda_0$ in Case 4. Table I summarizes the parameters and output results for Cases 1–4.

Table I

Case #	# of Unknowns	σ	Object	System	# of Processors	Serial/Par. Time (min.)	Speedup (overall)
1	8,800	$0.3\lambda_0$	None	Cluster	32	99/14	7.1
2	8,800	$0.1\lambda_0$	Sphere	Cluster	32	90/14	6.2
3	8,800	$0.04\lambda_0$	Spheroid	Cluster	32	88/12	7.2

4	60,320	$0.04\lambda_o$	Spheroid	Alpha Server	4	96/37	2.5
---	--------	-----------------	----------	-----------------	---	-------	-----

The speedup of a parallelized application is defined as the ratio of the serial runtime to the parallel runtime. In Fig. 2 the overall speedup and the speedup for the initialization routine (filling matrix $\bar{\bar{Z}}'$) are plotted versus the number of processors for Case 1. The speedup curves for Cases 2 and 3 are similar, with slightly different peak values of 6.2 and 7.2, respectively. The results show the significant speedup in the initialization time that is needed to fill the sparse matrix $\bar{\bar{Z}}'$. This initialization speedup dramatically affects the overall speedup of the code as shown in Fig. 2. In each case the peak overall speedup is observed when running on 32 processors, but most of this speedup is achieved using only 12 processors.

The efficiency of an application for a given number of processors is defined as the ratio of the speedup to the number of processors. Over Cases 1–3, the average speedup on 32 processors is 6.8, giving an efficiency of 0.21. Based on the serial runtimes, 88% of the code is executed in parallel. Therefore by Amdahl's Law [9], the peak speedup achievable for the current parallelization of the code is 8.3. We conclude that communication overhead and load imbalance among the processors accounts for the reduction in speedup from 8.3 to 6.8.

A comparison between the speedups achieved in the other bottlenecks (i)–(iii) mentioned in Section II is also shown in Fig 2. These results demonstrate that the overall speedup is almost the same as that achieved in the matrix–vector multiplication $\bar{\bar{Z}}\bar{\bar{T}}$ which is the bottleneck in (ii).

In the second set of experiments, we solved the moderate–scale problem of Case 4 (60,320 unknowns) on the Alpha SMP using all four available processors. The overall speedup in this case is 2.5 which is close to the predicted peak speedup of 2.9. This implies that executing the parallel

code on the 4-Alpha 667 MHz processor gives an impressive reduced absolute runtime for this moderate-scale case. The serial version of the code requires 950MB of memory, while the parallel version requires 1154MB of memory distributed over four processors (288, 290, 289 and 287MB each). The results of the parallel solution were identical to those of the serial implementation presented in [5].

The results described in this section demonstrate that by exploiting fine-grained parallelism within a single surface realization (one run of the code), we have achieved significant speedups. However, when the number of rough surface realizations is much larger than the number of available processors, as with Monte Carlo simulations, larger speedups are possible. This situation occurs when we need to run Monte Carlo simulations [5]. In this case we assign a group of these realizations (runs of the code) to be executed in parallel on each processor. Since the computations are independent and little communication is needed, this coarse-grained parallelism gives a perfect speedup that is only limited by the number of available processors. In other subsurface scattering configurations, we may need to obtain multiple views of a target buried under the same rough surface realization [4], which requires running the code only few times. A combination of fine and coarse-grained parallelism can make efficient use of all available processors.

III. CONCLUSIONS

MPI is successfully employed for the parallel implementation of the SDFMM. A significant overall speedup of 7.2 has been achieved on the 32-processor Beowulf cluster and a dramatic reduced runtime is gained using the 4-processor Alpha workstation. The greatest potential for speedup occurs in the sparse matrix filling and far field interaction steps.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. W. Chew and Prof. E. Michielssen at UIUC for allowing them the use and modification of the SDFMM computer code for the current application. This research was sponsored by the ERC Program of the NSF under award number EEC-9986821, in

part by the ARO Demining MURI grant # DAA 0-55-97-0013, and in part by the College of Engineering at the University of Arkansas.

REFERENCES

- [1] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [2] V. Jandhyala, B. Shanker, E. Michielssen, and W. C. Chew, "A fast algorithm for the analysis of scattering by dielectric rough surfaces," *J. Opt. Soc. Am. A*, vol. 15, no. 7, pp. 1877–1885, July 1998.
- [3] M. El-Shenawee, V. Jandhyala, E. Michielssen and W. C. Chew, "The steepest descent fast multipole method (SDFMM) for solving combined field integral equation pertinent to rough surface scattering," *Proc. of the IEEE APS/URSI '99 conf.*, Orlando, Florida, pp. 534–537, July 1999.
- [4] M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, "3-D subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces: Use of the steepest descent fast multipole method (SDFMM)," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 6, pp. 1174–1182, June 2001.
- [5] M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo simulations of electromagnetic wave scattering from random rough surface with 3-D penetrable buried object: Mine detection application using the SDFMM," *J. Opt. Soc. Am. A*, to appear in August 2001.
- [6] Message Passing Interface Forum. MPI: A Message-Passing Interface standard. *The International Journal of Supercomputer Applications and High Performance Computing*, no. 8, 1994.
- [7] S. V. Velamparambil, J. E. Schutt-Aine, J. G. Nickel, J. M. Song, and W. C. Chew, "Solving large scale electromagnetic problems using a linux cluster and parallel MLFMA," *IEEE Antennas Propagat. Symp.*, 1:636–639, July 1999.
- [8] S. Li, C. H. Chan, L. Tsang, Q. Li, and L. Zhou, "Parallel Implementation of the sparse matrix/canonical grid method for the analysis of two-dimensional random rough surfaces (three-dimensional scattering problem) on a Beowulf system," *IEEE Trans. Geoscience Rem. Sensing*, vol. 38, no. 4, pp. 1600–1608, July 2000.

- [9] G. M. Amdahl, "Validity of the single processor approach to achieving large scale computing capabilities", *Proc. AFIPS Spring Joint computer Conference* 30, Atlantic City, N. J., pp. 483–485, April 1967.

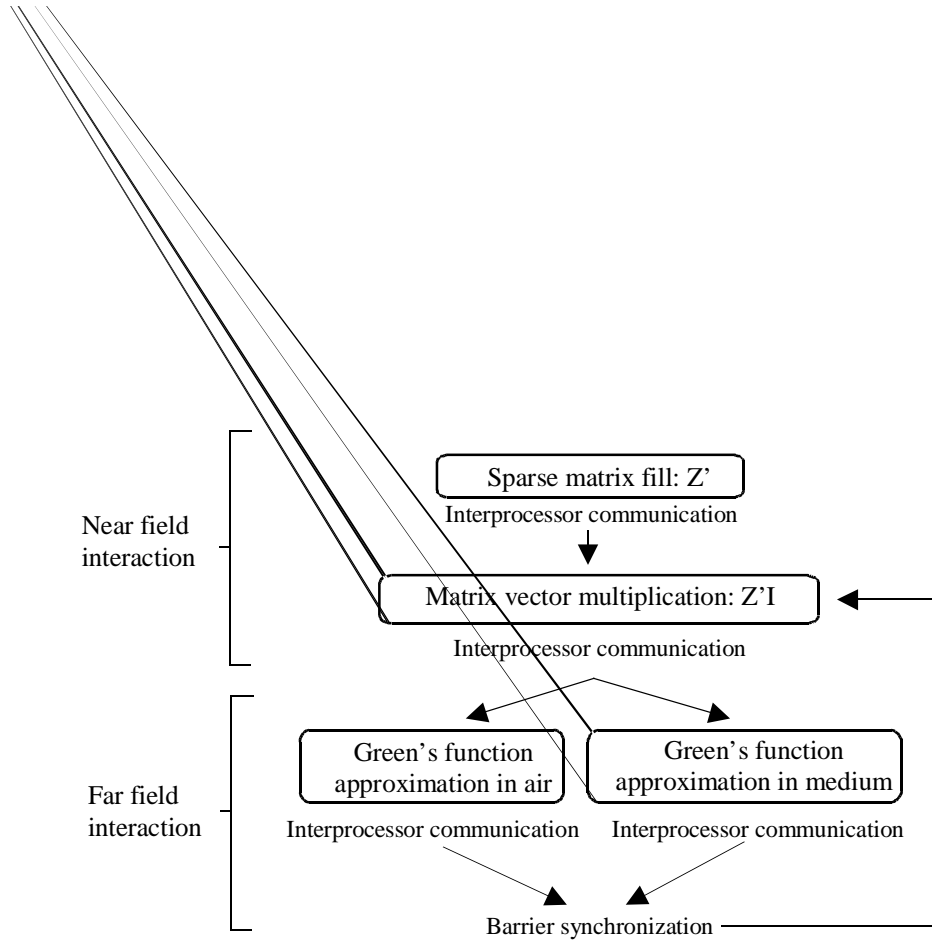


Figure 1. Structure of the parallelized SDFMM application.

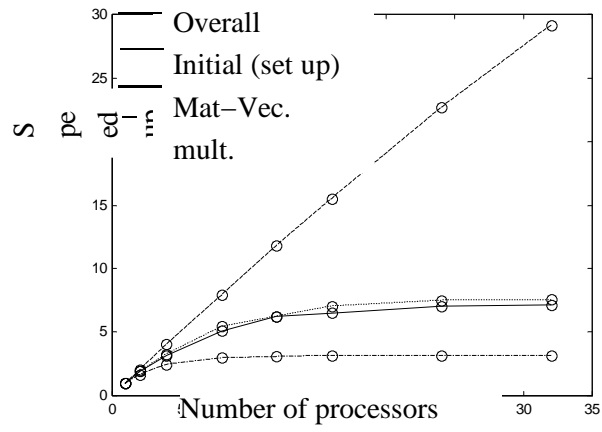


Figure 2. Speedup of parallel code versus the number of processors, and speedup for the three separate bottlenecks in the code versus the number of processors on the Beowulf cluster for Case 1.

COMPLEX-SOURCE-POINT NARROW-WAISTED RAY-LIKE GAUSSIAN BEAMS FOR FREQUENCY AND TIME DOMAIN RADIATION AND SCATTERING

Leopold B. Felsen^{1,2} and Vincenzo Galdi^{2,3}

¹ Department of Aerospace and Mechanical Engineering
Boston University, Boston, MA 02215, USA

Also, University Professor Emeritus,
Polytechnic University, Brooklyn, NY 11201, USA

² Department of Electrical and Computer Engineering
Boston University, Boston, MA 02215, USA

³ Waves Group, University of Sannio, Benevento, Italy

I. INTRODUCTION

Through replacement of the real spatial or spatial-temporal source locations in the frequency domain (FD) or time domain (TD) Maxwell field equations by locations in complex space or space-time, respectively, one may generate a new class of exact field solutions which convert point-source-excited fields in any environment into fields excited by Gaussian-beam-like wave objects in that environment (Deschamps, 1971; Felsen, 1976). While most applications of this elegant and physically appealing complex-source-point (CSP) technique have been concerned with "high frequency" tracking of well-collimated "wide-waisted" beams, we shall be concerned here with utilizing a tight superposition of narrow-waisted ray-like CSP beams (centered on a Gabor lattice) for synthesis of FD and TD distributed aperture radiation, and the interaction of these radiated fields with complex propagation and scattering environments. Previous studies have employed this algorithm for FD distributed phased apertures (Maciel and Felsen, 1989) and for transmission of these fields through focal regions and through plane or cylindrically stratified dielectrics (Maciel and Felsen, 1990a,b). Here, we extend this algorithm to FD scattering by a moderately rough perfectly conducting boundary. We also extend the FD aperture radiation algorithm to the short-pulse TD, utilizing CSP pulsed beam wavepackets. The FD problem is considered first.

II. RADIATION AND SCATTERING IN THE FREQUENCY DOMAIN

The problem geometry is shown in Fig. 1a. A y -directed two-dimensional y -independent time-harmonic electric field with suppressed $\exp(-i\omega t)$ dependence and spatial profile $f(x)$ is assumed to occupy the region $|x| \leq d/2$ in the aperture plane $z = z_A$,

$$E_y(x, z_A) = f(x), \quad |x| \leq d/2. \quad (1)$$

This field irradiates a perfectly conducting boundary with sinusoidal height profile

$$h(x) = A[1 - \cos(2\pi x/\Lambda)] \quad (2)$$

Figure 1. Problem geometry and Gabor phase space lattice. (a) Boundary profile: $h_{max} = 2A$; Λ =period. Aperture profile (inset): $f(x) = \cos(\pi x/d)$, $|x| \leq d/2$; $f(x) = 0$, $|x| > d/2$. (b) Gabor lattice: L_x =spatial period; β_x =spectral period.

measured from the $z = 0$ plane. We first consider the aperture problem.

II.a Radiation from aperture distributions

We summarize here essential results from previous publications (Bastiaans, 1980; Einziger and Shapira, 1986; Maciel and Felsen, 1989; Steinberg *et al*, 1991). The field radiated into the half-space $z < z_A$ from the aperture in (1) can be expressed as a line-source superposition (Kirchhoff integration)

$$E_y(x, z) = \frac{\partial}{\partial z} u(x, z), \quad u(x, z) = \frac{i}{2} \int_{-d/2}^{d/2} f(x') H_0^{(1)}(kR) dx', \quad (3)$$

where $k = \omega \sqrt{\epsilon_0 \mu_0} = 2\pi/\lambda$ is the free space wavenumber; λ is the free space wavelength; $H_0^{(1)}(\cdot)$ is the zeroth order Hankel function of the first kind; u is a scalar Debye potential; and

$$R = \sqrt{(x - x')^2 + (z - z_A)^2}. \quad (4)$$

II.b Beam discretization

The aperture field $f(x)$ is to be parameterized in terms of Gaussian beam basis functions via the rigorous self-consistent Gabor series representation

$$f(x) = \sum_{m,n=-\infty}^{\infty} A_{mn} w(x - mL_x) \exp(in\beta_x x), \quad (5)$$

where $w(x)$ represents the normalized Gaussian window

$$w(x) = \left(\frac{\sqrt{2}}{L_x} \right)^{1/2} \exp[-\pi(x/L_x)^2], \quad \int_{-\infty}^{\infty} w^2(x) dx = 1. \quad (6)$$

With k_x representing the x -domain wavenumber, this representation places the beams on a discretized (x, k_x) phase-space lattice (Fig. 1b), on which each lattice point gives rise to a Gaussian beam whose spatial and spectral (tilting) shifts are tagged by the indexes m and n , respectively. Spatial and spectral periods are related by the self-consistency relation (configuration-spectrum tradeoff) $L_x \beta_x = 2\pi$ (Bastiaans, 1980). The expansion coefficients can be computed by introducing an auxiliary function $\gamma(x)$ defined through the *biorthogonality* condition (Bastiaans, 1980),

$$\int_{-\infty}^{\infty} w(x) \gamma^*(x - mL_x) \exp(-in\beta_x x) dx = \delta_m \delta_n, \quad (7)$$

where $*$ denotes the complex conjugate, while $\delta_q = 1$ for $q = 0$ and $\delta_q = 0$ for $q \neq 0$. Accordingly (Bastiaans, 1980),

$$A_{mn} = \int_{-\infty}^{\infty} f(x) \gamma^*(x - mL_x) \exp(-in\beta_x x) dx. \quad (8)$$

For Gaussian windows, the biorthogonal function is given in (Bastiaans, 1980). For numerical computation of the Gabor coefficients, see (Einziger and Shapira, 1986). The radiated potential field in the half-space $z < z_A$ (see (3)) can be represented as (Maciel and Felsen, 1989)

$$u(x, z) = \sum_{m, n=-\infty}^{\infty} A_{mn} B_{mn}(x, z), \quad (9)$$

where the beam functions $B_{mn}(x, z)$ are synthesized by Gabor-weighted line-source superposition

$$B_{mn}(x, z) = \frac{i}{2} \int_{-d/2}^{d/2} w(x' - mL_x) \exp(in\beta_x x') H_0^{(1)}(kR) dx', \quad (10)$$

R being defined in (4). The integral in (10) (or its spectral counterpart) can be evaluated asymptotically in the beam *paraxial far zone*, yielding the following *complex source point* (CSP) approximation (Maciel and Felsen, 1990a)¹,

$$B_{mn}(x, z) \sim 2^{5/4} \left(\frac{L_x}{8\pi k \tilde{R}_{mn}} \right)^{1/2} \exp \left\{ i \left[k \left(\tilde{R}_{mn} + ib \right) + \pi/4 \right] \right\}, \quad (11)$$

with \tilde{R}_{mn} representing the *complex distance* between the observer at $P = (x, z)$ and the *complex source point* (here and henceforth, the tilde \sim denotes a complex quantity),

$$\tilde{P}'_{mn} = (\tilde{x}'_{mn}, \tilde{z}'_{mn}), \quad \tilde{x}'_{mn} = mL_x + ib \sin \theta_n, \quad \tilde{z}'_{mn} = z_A - ib \cos \theta_n, \quad (12)$$

$$\tilde{R}_{mn} = \overline{P \tilde{P}'_{mn}} = \sqrt{(x - \tilde{x}'_{mn})^2 + (z - \tilde{z}'_{mn})^2}. \quad (13)$$

In accord with the radiation condition, the square root is defined by $\text{Re}(\tilde{R}_{mn}) \geq 0$. The displacement parameter (Fresnel length) b is related to the beam lattice period L_x and the beam axis angle $\theta_n = \sin^{-1}(n\lambda/L_x)$ via $b = (L_x \cos \theta_n)^2/\lambda$ (Maciel and Felsen, 1990a), whence (11) is valid in the paraxial far-zone of each beam, $|\tilde{R}_{mn}| \gg b$. For large tilt angles with $n > L_x/\lambda$, the beam tilt angle θ_n is complex and the corresponding beams become *evanescent*.

II.c Narrow-waisted beams

For *narrow-waisted* beams ($L_x \ll d$), the Gabor coefficients can be effectively estimated by *sampling* the aperture field distribution, without performing the integration in (8) (Maciel and Felsen, 1989),

$$A_{mn} \approx \begin{cases} (L_x/\sqrt{2})^{1/2} f(mL_x), & n = 0, \\ 0, & n \neq 0, \end{cases} \quad (14)$$

so that from (9) and (11)

$$u(x, z) \approx \sum_{|m| \leq (d/2L_x)} A_{m0} B_{m0}(x, z), \quad (15)$$

where B_{m0} is obtained from (11) with $n = 0$. The tilted ($n \neq 0$) beams in the Gabor expansion, which generate evanescent "far fields", are ignored in this approximation.

II.d Linearly-phased aperture

Narrow-waisted beams work very well for nonphased apertures, but usually require finer aperture sampling in the presence of phasing (Maciel and Felsen, 1990a). Here we consider a linearly-phased cosine aperture field,

$$f(x) = g(x) \exp(ikx \sin \theta_A), \quad g(x) = \cos(\pi x/d), \quad |x| \leq d/2, \quad (16)$$

where θ_A is the tilt angle of the main radiation lobe with respect to the z axis. In this case, a more effective discretization can be obtained by Gabor-expanding the real function $g(x)$ only, and including the linear phasing in the beam integral (10) for the B_{m0} beam propagator. Accordingly, the narrow-waisted ($L_x \ll d$) beam expansion can be recast as

$$u(x, z) \approx \sum_{|m| \leq (d/2L_x)} C_m B_{m0}(x, z), \quad C_m = (L_x/\sqrt{2})^{1/2} g(mL_x), \quad (17)$$

$$B_{m0}(x, z) \sim 2^{5/4} \left(\frac{L_x}{8\pi k \tilde{R}_{m0}} \right)^{1/2} \exp \left\{ i \left[k \left(\tilde{R}_{m0} + mL_x \sin \theta_A + ib \right) + \pi/4 \right] \right\}, \quad (18)$$

¹ Note that there are some sign changes with respect to (Maciel and Felsen, 1990a), since here we assume propagation into the half-space $z < z_A$.

Figure 2. Linearly phased cosine aperture distribution $f(x) = \cos(\pi x/d) \exp(ikx \sin \theta_A)$ ($d = 10\lambda$, $\theta_A = 30^\circ$). Near-zone radiated potential field $u(x, z)$ at $z_A - z = 5\lambda$ synthesized using narrow-waisted nontilted and tilted beams (with $L_x = 0.02d$) is compared with the Kirchhoff integration reference solution.
— Reference solution; - - - Nontilted beams; ····· Tilted beams.

$$\tilde{R}_{m0} = \sqrt{(x - mL_x - ib \sin \theta_A)^2 + (z - z_A + ib \cos \theta_A)^2}, \quad b = (L_x \cos \theta_A)^2 / \lambda. \quad (19)$$

The beam propagator (18) differs from B_{m0} in (15) by the phase shift $(ikmL_x \sin \theta_A)$ which produces the propagation-matched tilt θ_A in the beam direction. In Fig. 2, the near-zone potential field synthesized using the narrow-waisted tilted beam decomposition in (17) is compared with a calibrated computation-intensive Kirchhoff integration reference solution based on (3), and with the nontilted formulation in (15) when applied to the *entire* aperture field in (16). The tilted beam synthesis is hardly distinguishable from the reference solution, whereas the nontilted synthesis is less accurate in the magnitude.

II.e Reflection from a periodic perfectly conducting boundary

The field radiated by the aperture distribution in (16) is now assumed to impinge on a perfectly conducting moderately rough periodic boundary described by the continuous function $h(x)$ which is assumed to vary slowly over a wavelength scale (Fig. 1a). Moderately rough irregular dielectric interfaces separating two dielectric half-spaces are treated elsewhere (Galdi *et al*, 2000a).

The reflected field can be constructed rigorously by *complex ray tracing* applied to each beam in the aperture decomposition; this requires the analytic continuation, into a complex configuration space, of *all* geometrical parameters involved (with the exception of the observation point). However, *narrow-waisted* beams can be tracked accurately and much more efficiently via a beam-tracing paraxial *almost real ray-tracing* scheme (Ruan and Felsen, 1986; Maciel and Felsen, 1990b)), valid in appropriately calibrated observation ranges. For the new application to a periodic boundary, we first treat the canonical problem of CSP Gaussian beam reflection from a curved segment on a conducting boundary. The problem geometry is illustrated in Fig. 3.

An incident Gaussian beam is generated by a CSP at

$$\tilde{P}_s = (\tilde{x}_s, \tilde{z}_s), \quad \tilde{x}_s = x_s + ib \cos \alpha_s, \quad \tilde{z}_s = z_s + ib \sin \alpha_s, \quad (20)$$

α_s being the beam axis *real* departure angle with respect to the x axis. For electrically large and smooth scatterers, and when the observation point $P = (x, z)$ lies in the paraxial region of a reflected beam ($\Delta \ll (|\tilde{F}_{r0} P_0|^2 + b^2)^{1/2}$ in Fig. 3), the field can be approximated in terms of the on-axis field of that beam (at P_0) and a complex phase correction. Denoting the on-axis parameters by the subscript zero, one finds for the potential field (Ruan and Felsen, 1986)

$$u^{refl}(P) \approx u^{refl}(P_0) \big|_{\tilde{P}_i=P_{i0}} \exp(ik\tilde{\delta}_p) = \Gamma u^{inc}(P_{i0}) \left(\frac{\tilde{f}_{r0}}{\tilde{R}_0} \right)^{1/2} \exp \left[ik(L_{r0} + \tilde{\delta}_p) \right], \quad (21)$$

$$\tilde{f}_{r0} = \frac{\tilde{L}_{i0} r_c \cos \theta_i}{2\tilde{L}_{i0} + r_c \cos \theta_i}, \quad r_c = - \frac{[1 + (h'(x_{i0}))^2]^{3/2}}{h''(x_{i0})}, \quad ' \equiv d/dx. \quad (22)$$

Figure 3. CSP beam reflection at a curved conducting boundary described by the contour $h(x)$.

Here (see Fig. 3): $\tilde{R}_0 = \overline{\tilde{F}_{r0}P_0}$; $\tilde{R} = \overline{\tilde{F}_{r0}P}$; $\tilde{L}_{i0} = \overline{\tilde{P}_sP_{i0}}$; $L_{r0} = \overline{P_{i0}P_0}$; $\delta_p = \tilde{R} - \tilde{R}_0$ is the complex phase correction; $\tilde{F}_{r0} = (x_{i0} - \tilde{f}_{r0} \cos \alpha_r, z_{i0} - \tilde{f}_{r0} \sin \alpha_r)$ is the *complex* (virtual) focus obtained by analytic continuation of the standard ray-optical formulas (Felsen and Marcuvitz, 1973, p. 168); the *complex* incidence point \tilde{P}_i is approximated by the *real* beam-axis incidence point $P_{i0} = (x_{i0}, z_{i0})$; α_r is the reflected-beam-axis real departure angle with respect to the x axis; r_c is the curvature radius at P_{i0} ; and $\Gamma = 1$ is the plane-wave potential field reflection coefficient. As shown in (Ruan and Felsen, 1986), this scheme corresponds to tracing a ray along a *complex* trajectory from the CSP at \tilde{P}_s to the intersection P_{i0} of the *real* beam axis with the *real* surface; from there, the path to the observer proceeds entirely in *real* configuration space, along the beam axis. Multiple reflections, which may occur in the configuration in Fig. 1a, can be incorporated by iterating (21), whereby the complex focus (either virtual or real) associated with each iteration becomes the phase reference for the next iteration (Galdi *et al.*, 2000a). Apart from the complex ray connecting the CSP to the first real incidence point, the multi-hop path to the observer proceeds entirely in *real* configuration space along the beam axes, and the phase correction is applied only on the last path segment leading to the observer.

II.f Application: infinite sinusoidal boundary

For a first check on the applicability of the narrow-beam algorithm to surface scattering problems, we have considered the sinusoidal boundary in Fig. 1a illuminated by a nonphased cosine aperture distribution with wavelength λ . Extensive numerical experiments have been performed for various observation heights z and aperture heights z_A ; profiles $h(x)$ with various minimum curvature radii r_c ; and various beam lattice spacings L_x . All of these numerical implementations for the scattered potential field have been compared with a numerically integrated, computation intensive Physical Optics-Kirchhoff reference solution based on (3); by previous calibration, Physical Optics has been confirmed to apply to the profile parameters under consideration here. Typical results for the potential u^{refl} are displayed in Fig. 4. In general, we have found that the accuracy of the narrow-waisted beam algorithm *improves* for *greater observation distance* (because of the far-zone paraxial approximation), but even at moderate distances the agreement is satisfactory. We also found that the beam algorithm is quite robust with respect to *scramblings* of the Gabor lattice. As discussed in (Maciel and Felsen, 1990b), this feature can be exploited to obtain *a priori* accuracy assessments when reference solutions are not available. For the present nonphased aperture example, we found the best tradeoff between accuracy and computational cost to occur for $L_x \sim \lambda$. A *finer* sampling may, however, be required for *phased* aperture field distributions. On the other hand, the accuracy gets *worse* as the *distance of the aperture from the surface increases*. We found that in order to get robust and accurate predictions, we should have $z_A - h_{max} \leq \lambda$, where h_{max} is the maximum of the boundary profile. However, this is not a very restrictive limitation since it is always possible (and computationally cheap because of the simple determination of the Gabor coefficients via (14)) to perform a multi-step Gabor decomposition for greater aperture-to-boundary distances, i.e., project the beam-computed radiated potential field onto a virtual aperture suitably close to the surface and then again apply the narrow-waisted beam algorithm. Concerning degradation of accuracy with *increase* in wavelength, we found that even for relatively "low-frequency" geometries, i.e. moderate (minimum r_c)/ λ as in Fig. 4b, the beam algorithm, though no longer *highly* accurate, still provides reasonably good predictions (for details, see (Galdi *et al.*, 2000a)).

Figure 4. Beam-computed reflected potential $u(x, z)$ and Physical Optics (PO) reference solution. (a) Aperture: width $d = 100\lambda$; height $z_A = 11\lambda$. Boundary: period $\Lambda = 50\lambda$; height $h_{max} = 2A = 10\lambda$; minimum $r_c = 12\lambda$; maximum slope $\phi_{max} = 32^\circ$. Beam lattice period: $L_x = 0.01d = \lambda$. Observation plane at $z_{obs} = 50\lambda$. (b) $d = 20\lambda$; $z_A = 3\lambda$; $\Lambda = 10\lambda$; $h_{max} = 2A = 2\lambda$; minimum $r_c = 2.5\lambda$; $\phi_{max} = 32^\circ$; $L_x = 0.05d = \lambda$; $z_{obs} = 10\lambda$. Because the reflected field is symmetric with respect to $x = 0$, only the $x > 0$ portion is shown over the range $0 \leq x \leq 2\Lambda$. — PO reference solution; - - - Beams.

III. TIME DOMAIN RADIATION FROM DISTRIBUTED APERTURES

We shall now explore the extension of the FD results for aperture radiation in Sec. II.a to time-dependent excitation, in particular to short pulses. To this end, we consider a space-time aperture field distribution at $z = 0$ with separable space-time dependence and linear time delay

$$e_y(x, 0, t) = f(x, t), \quad |x| \leq d/2, \quad f(x, t) = g(x)p(t - c^{-1}x \sin \theta_A), \quad (23)$$

where c is the speed of light and $p(t)$ is a pulse with characteristic width T_p . This distribution represents the TD counterpart of the linearly-phased time-harmonic aperture in (16). The present TD formulation is restricted to the radiated field only, and we analyze the propagation into the $z > 0$ halfspace, wherein (23) gives rise to a space-time pulse propagating in the θ_A direction (Fig. 5). It is assumed that the normalized width of the pulse is much shorter than the aperture dimension d , i.e., $cT_p \ll d$.

III.a Reference solution

Using the two-dimensional TD Green's function

$$G(x, z, t; x', z', t') = \frac{1}{2\pi\sqrt{(t-t')^2 - (s/c)^2}} H[(t-t') - s/c], \quad s(z') = \sqrt{(x-x')^2 + (z-z')^2}, \quad (24)$$

where $H(\cdot)$ is the Heaviside step function, the field radiated into the half-space $z > 0$ can be represented as a space-time Kirchhoff integration (Morse and Feshbach, 1973, Sec. 7.3; Kragalott *et al.*, 1997),

$$e_y(x, z, t) = \frac{z}{\pi} \int_{-d/2}^{d/2} dx' \frac{1}{s(0)^2} \int_{-\infty}^{t-s(0)/c} \frac{(t-t') \frac{\partial f}{\partial t'}(x', t')}{\sqrt{(t-t')^2 - (s(0)/c)^2}} dt'. \quad (25)$$

Direct numerical integration of (25) is complicated by the square-root (integrable) singularity at the upper limit $t' = t - s(0)/c$, and requires special care. We have used the Newton-Cotes scheme (Kragalott *et al.*, 1997) for the numerical integration of (25), which represents our reference solution.

III.b Beam discretization

The formal extension of the Gabor-based time-harmonic aperture radiation to time-dependent excitation involves a four-index Gabor series set on a discretized lattice in an eight-dimensional phase space (space-wavenumber, time-frequency). For a rigorous treatment and computational issues, see (Steinberg and Heyman, 1991). We shall explore to what extent the narrow-waisted beam approach,

Figure 5. Linear-delay space-time aperture field distribution and pulsed beam coordinates.

effective for time-harmonic excitation, can be generalized to TD (short-pulse) excitation. The linear-delay aperture field distribution (23) admits the equivalent spectral representation

$$\hat{f}(x, \omega) \equiv \int_{-\infty}^{\infty} f(x, t) \exp(i\omega t) dt = g(x) P(\omega) \exp(ikx \sin \theta_A), \quad P(\omega) = \int_{-\infty}^{\infty} p(t) \exp(i\omega t) dt. \quad (26)$$

A TD beam discretization can be obtained by Fourier-inverting the narrow-waisted tilted beam expansion (17) for the FD linearly-phased aperture presented in Sec. II.d. In order to accommodate the evanescent spectra in the FD beam propagators (18), we use the *analytic signal* formulation instead of the standard Fourier transform (Heyman and Melamed, 1998). Concerning the beam lattice discretization, one can choose a frequency-independent beam lattice period L_x (resulting in a frequency-dependent beam parameter b), or a frequency-independent beam parameter b (resulting in a frequency-dependent L_x). We choose L_x frequency-independent because it yields frequency-independent Gabor coefficients (see (17)). The TD counterpart of the narrow-waisted FD beam expansion (17) for the aperture field distribution (26), with reference to the y -directed electric field, can be thus written as

$$e_y(x, z, t) \approx \sum_{|m| \leq (d/2L_x)} c_m \mathbf{b}_m(x, z, t), \quad c_m = (L_x/\sqrt{2})^{1/2} g(mL_x). \quad (27)$$

The pulsed beam propagator \mathbf{b}_m is the TD counterpart of the FD paraxial, far-zone beam propagator in (18), with (19)² (for simplicity, the subscript " m_0 " is henceforth replaced by " m ")

$$\mathcal{B}_m(x, z, \omega) = -ik2^{5/4} \left(\frac{L_x}{8\pi k} \right)^{1/2} \exp \left\{ i \left[k \left(\tilde{R}_m + mL_x \sin \theta_A + ib \right) + \pi/4 \right] \right\} \frac{(z - ib \cos \theta_A)}{\tilde{R}_m^{3/2}}, \quad (28)$$

$$\tilde{R}_m = \sqrt{(x - mL_x - ib \sin \theta_A)^2 + (z - ib \cos \theta_A)^2}, \quad b = (L_x \cos \theta_A)^2 / \lambda. \quad (29)$$

Via the analytic signal formulation, one has (Re=real part)

$$\mathbf{b}_m(x, z, t) = \text{Re} \left[\frac{1}{\pi} \int_0^\infty \mathcal{B}_m(x, z, \omega) P(\omega) \exp(-i\omega t) d\omega \right]. \quad (30)$$

While the integral in (30) cannot be evaluated explicitly in general, we have found useful closed-form approximations for the important class of Gaussian pulses. In particular, we consider a Rayleigh (four-times-differentiated Gaussian) pulse

$$p(t) = \exp \left[-\frac{50(t - T_p/2)^2}{T_p^2} \right] \left[1 + \frac{10000(t - T_p/2)^4 - 600(t - T_p/2)^2 T_p^2}{3T_p^4} \right], \quad (31)$$

²Note that (28), (29) are slightly different from (18), (19), since here the electric field is considered instead of the potential, the aperture plane is located at $z_A = 0$, and propagation is into the $z > 0$ halfspace.

but the procedure presented below can be applied to any kind of modulated or differentiated Gaussian.

Since the beam lattice period L_x has been chosen frequency independent, the beam parameter b and hence the complex distance \tilde{R}_m in (29) are *frequency dependent*. For $z \gg |b \cos \theta_A|$, we can approximate in the amplitude factor of (28)

$$\frac{(z - ib \cos \theta_A)}{\tilde{R}_m^{3/2}} \approx \frac{z}{R_m^{3/2}}, \quad R_m = \sqrt{(x - mL_x)^2 + z^2}, \quad (32)$$

rendering the distance R_m *real*. In the phase, we retain the first order paraxial correction

$$\tilde{R}_m \approx z_{bm} - ib + \frac{x_{bm}^2(z_{bm} + ib)}{2z_{bm}^2}, \quad (33)$$

valid for $|z_{bm} - ib| \gg |x_{bm}|$ and $z_{bm} \gg b$; here x_{bm}, z_{bm} are the beam coordinates (see Fig. 5)

$$\begin{bmatrix} x_{bm} \\ z_{bm} \end{bmatrix} = \begin{bmatrix} \cos \theta_A & -\sin \theta_A \\ \sin \theta_A & \cos \theta_A \end{bmatrix} \begin{bmatrix} x - mL_x \\ z \end{bmatrix}. \quad (34)$$

In the TD, the beam parameter b must be small over the entire bandwidth Ω_p of $p(t)$ in (31),

$$b = \frac{(L_x \cos \theta_A)^2 \omega}{2\pi c} \ll z_{bm}, \quad \omega \leq \Omega_p. \quad (35)$$

With these assumptions, the integral in (30) can be reduced to the generic form (the spectrum $P(\omega)$ of $p(t)$ in (31) is evaluated readily),

$$\int_0^\infty \omega^{9/2} \exp(\xi\omega + \zeta\omega^2) d\omega, \quad \text{Re}(\zeta) < 0, \quad (36)$$

which can be expressed in terms of confluent hypergeometric functions (Abramowitz and Stegun, 1964, Sec. 13). Accordingly, the TD beam propagator can be written explicitly as

$$b_m(x, z, t) = \text{Re} \left\{ \eta \left[13\alpha \Gamma\left(\frac{11}{4}\right) G_1\left(5\sqrt{\frac{\pi}{2}}\frac{\beta}{\alpha}\right) + 20i\sqrt{2\pi}\beta \Gamma\left(\frac{17}{4}\right) G_2\left(5\sqrt{\frac{\pi}{2}}\frac{\beta}{\alpha}\right) \right] \right\}, \quad (37)$$

$$\alpha = \sqrt{\pi(cT_p z_{bm})^2 + 50(L_x x_{bm} \cos \theta_A)^2}, \quad \beta = x_{bm}^2 + z_{bm}(2z_{bm} + cT_p - 2ct) + 2L_x m z_{bm} \sin \theta_A, \quad (38)$$

$$\eta = \frac{4z\sqrt{10L_x}\pi^{7/4}T_p^5(cz_{bm})^{11/2}\exp(-i\pi/4)}{39\sqrt{c}R_m^{3/2}\alpha^{13/2}}, \quad G_1(v) = M\left(\frac{11}{4}, \frac{1}{2}, -v^2\right), \quad G_2(v) = M\left(\frac{13}{4}, \frac{3}{2}, -v^2\right), \quad (39)$$

where R_m, x_{bm} and z_{bm} are defined in (32) and (34); $\Gamma(\cdot)$ is the gamma function (Abramowitz and Stegun, 1964, Sec. 6); and $M(q, s, v)$ is the confluent hypergeometric function (Abramowitz and Stegun, 1964, Sec. 13). The above procedure can be applied to *any* Gaussian pulse; modulation or differentiation only affects the q, s arguments of M . We found simple rapidly converging approximations for the functions G_1 and G_2 in the form $G_{1,2} \sim \exp(-v^2)$ (power series in v^2) (Galdi *et al*, 2000b). These functions resemble the functional form of the time pulse excitation in (31). Using these approximations the TD beam propagator in (37) can be computed efficiently.

III.c Assessment of accuracy

The restriction in (35) is the most serious because, for specified Ω_p, θ_A , and observation point, it determines the maximum allowable lattice period (i.e., the minimum number of beams). The overall constraint can be expressed as (Galdi *et al*, 2000b)

$$Q \equiv \frac{L_x}{d} \sqrt{\frac{\kappa \cos^3 \theta_A}{\chi}} \ll 1, \quad (40)$$

where $\chi = z_{obs}/F$ is the distance of the observation plane scaled by the Fresnel distance of the aperture, $F = d^2/(cT_p)$; $\kappa = \Omega_p T_p/2\pi$ is the normalized bandwidth of the pulse $p(t)$; and L_x/d determines the number of beams in the expansion (27). The nondimensional estimator $Q \ll 1$ expresses the range of validity of the algorithm in terms of all relevant parameters of the problem. For example, increasing the lattice period L_x (i.e., decreasing the number of beams) can be compensated by a corresponding

Figure 6. Nonphased ($\theta_A = 0$) sine aperture distribution with $cT_p = 0.5$, $d = 5 = 10cT_p$ (arbitrary units). TD narrow-waisted beam synthesis (for various values of the beam-lattice period) is compared with space-time Kirchhoff integration (reference solution). (a) Time evolution at a fixed observation point $x = 1.25$, $z = 5 = 0.1F$ (arbitrary units); (b) transverse cut (spatial profile) at $z = 5$, $ct = 5.25$.
— Reference solution; - - - Beams ($L_x = d/20$, $Q = 0.45$); $\cdots\cdots$ Beams ($L_x = d/50$, $Q = 0.18$); - - - - Beams ($L_x = d/100$, $Q = 0.09$).

Figure 7. Parameters as in Fig. 6, but with linear delay ($\theta_A = 30^\circ$). (a) Time evolution at a fixed observation point $x = 4.14$, $z = 5 = 0.1F$; (b) transverse cut at $z = 5$, $ct = 6.7$.
— Reference solution; - - - Beams ($L_x = d/20$, $Q = 0.36$); $\cdots\cdots$ Beams ($L_x = d/50$, $Q = 0.14$); - - - - Beams ($L_x = d/100$, $Q = 0.07$).

increase of $(z_{obs})^{1/2}$. In order to assess the accuracy of the proposed TD beam expansion, we have performed computations for the linear-delay space-time aperture distribution (23) with a sine spatial tapering, $g(x) = \sin(2\pi x/d)$, $|x| \leq d/2$, excited by the Rayleigh time pulse (31). Figure 6a shows the time evolution of the y -directed electric field at a fixed observation point (x, z) in the near zone ($z = 0.1F$) of a large aperture ($d = 10cT_p$) without phase delay ($\theta_A = 0$), computed via the space-time Kirchhoff integration (25), and via the TD beam synthesis (27) with various beam lattice periods. As expected, the agreement improves as the beam lattice period L_x decreases, and satisfactory accuracy is achieved for $Q \leq 0.1$ (for this example $\kappa \sim 8$). It is observed from the transverse cut in Fig. 6b that, despite the use of the paraxial paraxial far-zone approximation, the TD beam synthesis is quite accurate even in the near zone of the aperture ($z = 0.1F$) and not only around the main radiation lobes. We found that, as the observation distance increases, a coarser discretization can be used according to (40); even at moderate distances, quite accurate syntheses can be achieved with a relatively small number of beams (~ 50) (Galdi *et al.*, 2000b). The corresponding results for linear-delay ($\theta_A = 30^\circ$) are shown in Fig. 7 and the same considerations apply.

IV. CONCLUSIONS

In many current forward and inverse scattering scenarios, there is a need for numerically efficient robust

forward solvers for fields excited by distributed sources in the presence of complex environments. This motivation has led us to re-visit the previously developed FD narrow-waisted Gaussian beam algorithms (Maciel and Felsen, 1989; 1990a,b) and to extend them to new FD applications as well as to the short-pulse TD. The outcomes from the rough sinusoidal scattering example here, as well as the first results in the TD, are encouraging. Further FD extensions to irregularly rough interfaces between dielectrics are already in progress, as are TD interactions with canonical scatterers to learn the new rules.

ACKNOWLEDGEMENTS

The authors acknowledge fruitful discussions with Professor D.A. Castañon (Boston University) on forward and inverse scattering and imaging which motivated this investigation. We also acknowledge partial support by ODDR&E under MURI grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028. In addition, VG acknowledges a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy.

REFERENCES

- Abramowitz, M., and Stegun, I.A., 1964, *Handbook of Mathematical Functions*, Dover, New York.
- Bastiaans, M.J., 1980, Gabor's expansion of a signal into Gaussian elementary signals, *Proc. IEEE*, 68:538.
- Deschamps, G.A., 1971, Gaussian beams as a bundle of complex rays, *Electron. Lett.*, 7:684.
- Einziger, P.D., and Shapira, M., 1986, Gabor representation and aperture theory, *J. Opt. Soc. Am. A*, 3:508.
- Felsen, L.B., Complex-source-point solutions of the field equations and their relation to the propagation and scattering of Gaussian beams, *Symp. Matemat., Istituto Nazionale di Alta Matematica*, "Academic, London, XXVIII:40.
- Felsen, L.B., and Marcuvitz, N., 1973, *Radiation and Scattering of Waves*, Prentice Hall, Englewood Cliffs, NJ. Classic reissue, IEEE Press, Piscataway, NJ, 1994.
- Galdi, V., Felsen, L.B., and Castañon, D.A., 2000a, Quasi-ray Gaussian beam algorithm for scattering by, and reconstruction of, moderately rough interfaces - Part I: forward scattering (internal memorandum, in preparation for publication).
- Galdi, V., Felsen, L.B., and Castañon, D.A., 2000b, Narrow-waisted Gaussian beam discretization for time-dependent radiation from large apertures (internal memorandum, in preparation for publication).
- Heyman, E., and Melamed, T., 1998, Space-time representation of ultra wideband signals, *Advances in Imaging and Electron Physics*, 103:1.
- Kragalott, M., Kluskens, M.S., and Pala, W.P., 1997, Time-domain fields exterior to a two-dimensional FDTD space, *IEEE Trans. Antennas Propagat.*, 45:1655.
- Maciel, J.J., and Felsen, L.B., 1989, Systematic study of fields due to extended apertures by Gaussian beam discretization, *IEEE Trans. Antennas Propagat.*, 37:884.
- Maciel, J.J., and Felsen, L.B., 1990a, Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part I - plane layer, *IEEE Trans. Antennas Propagat.*, 38:1607.
- Maciel, J.J., and Felsen, L.B., 1990b, Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part I - circular cylindrical layer, *IEEE Trans. Antennas Propagat.*, 38:1618.
- Morse, P.M., and Feshbach, H., 1953, *Methods of Theoretical Physics*, McGraw-Hill, New York.
- Ruan Y.Z., and Felsen, L.B., 1986, Reflection and transmission of beams at a curved interface, *J. Opt. Soc. Am. A*, 3:566.
- Steinberg, B.Z., Heyman, E., and Felsen, L.B., 1991, Phase-space beam summation for time-harmonic radiation from large apertures, *J. Opt. Soc. Am. A*, 8:41.
- Steinberg, B.Z., and Heyman, E., 1991, Phase-space beam summation for time-dependent radiation from large apertures: discretized parameterization, *J. Opt. Soc. Am. A*, 8:959.

Multifrequency Reconstruction of Moderately Rough Interfaces via Quasi-Ray Gaussian Beams

Vincenzo Galdi, *Member, IEEE*,
David A. Castañon, *Senior Member, IEEE*,
and Leopold B. Felsen, *Life Fellow, IEEE*

This work was supported by ODDR&E under MURI Grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from Grant No. 9900448 by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University, Brooklyn, NY 11201 USA.

V. Galdi is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, on leave from University of Sannio, Benevento, Italy (e-mail: vgaldi@bu.edu).

D.A. Castañon is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA (e-mail: dac@bu.edu).

L.B. Felsen is with the Department of Aerospace and Mechanical Engineering and the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, part-time. He is also University Professor Emeritus, Polytechnic University, Brooklyn, NY 11201 USA (e-mail: lfelsen@bu.edu).

Abstract

In this paper, we present a new technique for determining the surface profile of a moderately rough interface between air and a homogeneous dielectric half-space. Based on sparsely sampled step-frequency ground penetrating radar measurements, the proposed inversion scheme uses a quasi-ray Gaussian beam fast forward model, coupled with a low-order parameterization of the surface profile in terms of B-splines. The profile estimation problem is posed as a parameter optimization problem, which is solved using a multiresolution continuation method via *frequency hopping*. Numerical experiments establish that the algorithm is efficient, and yields accurate reconstructions throughout most of the illuminated region even in noisy environments, losing accuracy only in regions with very weak illumination.

Keywords

Ground penetrating radars, rough surfaces, inverse scattering, Gaussian beams.

I. INTRODUCTION

The problem of determining the properties of rough surfaces from electromagnetic (EM) reflected field data arises in many important applications, ranging from nondestructive testing to underground imaging. In this paper, we address underground imaging via ground-penetrating radar (GPR). In GPR systems, arrays of above-ground transmitters and receivers illuminate areas of interest, and receive backscattered signals from underground objects and from surface reflections. The shape of the air-ground interface is unknown, and constitutes a principal corruptor of the backscattered signal from subsurface targets of interest. In order to enhance subsequent detection, classification and inverse scattering processing, it is important to compensate for the distortion introduced by the air-ground interface.

One possible approach for this compensation is to characterize the statistics of the distortion caused by unknown rough surfaces, and then apply appropriate statistical signal processing techniques. Such an approach was used in [1]–[4] for detection of buried mines via both forward-looking and downward-looking GPR systems. However, this approach fails to exploit deterministic information present in the received signals due to scattering from the air-ground interface, and thus yields limited accuracy and robustness in classification and reconstruction (see, e.g., [5]). In this paper, we address the problem of estimating the profile of the air-ground interface from in-situ frequency-stepped GPR measurements, for use in subsequent imaging and classification processing.

Estimation of rough surfaces from inverse scattering has received considerable attention in the past decade. However, most of the available algorithms have focused on conducting surfaces. Wombell and DeSanto [6], [7] used Kirchhoff approximation and Fourier transform to estimate surface profiles based on measurements of the reflected field in all spectral directions. Noguchi and his colleagues [8], [9] used nonlinear optimization techniques for direct estimation of surfaces illuminated by monochromatic Gaussian beams, based on the far-field scattering amplitude for all spectral directions. In a different approach, Schatzberg and Devaney [10] used Rytov approximation and backpropagation to estimate surface profiles based on full measurements of the reflected wave.

In contrast with the above contributions, our work in this paper is focused on estimating surface profiles based on reflection from a moderately rough interface between air and a homogeneous dielectric half-space (soil), as illustrated in Fig. 1. Furthermore, we assume that the reflected field is measured only at receivers with discrete spatial locations, using a stepped-frequency GPR. We assume that the incident field arises from a discrete set of transmitters, and thus has a finite aperture.

First we employ a forward model relating the measured fields at the receivers to the surface profile. This model, detailed in [11], utilizes Gabor-based Gaussian beam algorithms in conjunction with the complex source point (CSP) method for generating beam-like wave objects [12]–[14]. With this model, together with a parametric representation of the surface in terms of B-splines [15], the surface estimation problem is posed as a nonlinear optimization problem, similar in spirit to the approach in [8]. We show that this optimization problem has local minima, and exploit a multiresolution continuation strategy based on *frequency hopping* [16]–[18], to approach convergence to globally optimal estimates of surface profiles.

The rest of this paper is organized as follows: Section II describes the problem of rough surface reconstruction from frequency-stepped GPR measurements, and poses the estimation problem as a nonlinear optimization problem. Section III describes the surface parameterization, the Gabor-based Gaussian beam algorithm, the multiresolution continuation algorithm, and the resulting optimization approach for determining the surface profile. Section IV details the results of numerical experiments to illustrate the accuracy

of the surface reconstructions under different conditions. Section V provides concluding remarks.

II. STATEMENT OF THE PROBLEM

The problem of interest is reconstructing the air-ground interface coarse-scale profile from a limited number of spatially sampled reflected field measurements, using a limited aperture illumination from a frequency-stepped GPR system. We consider a two-dimensional (2D) model, as depicted in Fig. 1, where a TM-polarized EM field with implicit time-harmonic dependence $\exp(-i\omega t)$ illuminates a dielectric half-space with known relative permittivity ϵ_r and a moderately rough interface, whose coarse scale profile is described by a continuous function $h(x)$. The field is assumed to be generated by an extended aperture distribution at $z = z_A$,

$$\mathbf{E}^{inc}(x, z_A) = f(x)\mathbf{u}_y, \quad |x| \leq d/2, \quad z = z_A, \quad (1)$$

where \mathbf{u}_y denotes a y -directed unit vector.

In this preliminary investigation, we neglect the presence of any buried object; surface estimation strategies in the presence of shallowly buried plastic mine-like targets, not very different in principle, are dealt with elsewhere [19]–[22]. Furthermore, we also neglect the *noisy* (incoherent) contribution of finer-scale roughness, and focus on estimation of the coarse shape, acknowledging the implicit limits of retrievable information through inverse scattering (see, e.g., [23], and the more relevant *near-proximity* extension [24]). The y -directed reflected electric field is sampled at a number N_r of fixed receiver locations $x_1^r, \dots, x_{N_r}^r$ at the observation plane $z = z^r$. As common in many GPR systems, we assume to work in a *step-frequency* regime with N_λ different operating wavelengths (frequencies), so that a set of $N_r \times N_\lambda$ complex (magnitude and phase) samples constitutes the *observed data* of the problem. It is well-known that this inverse scattering problem is *ill-posed*, and therefore a *blind* implementation of inverse scattering techniques would result in *ill-conditioned* numerical algorithms. We refer the interested reader to [23], [24] for analysis of the *retrievable information* from both theoretical and computational viewpoints.

Here, a *robust* inversion strategy via a well-posed inverse of the problem is achieved by

- i) Introducing a finite-dimensional compact geometrical parameterization of the unknown interface profile;
- ii) Estimating the unknown parameters by *fitting* the model-based forward scattering prediction to the available (measured/simulated) data, i.e., minimizing a suitable *cost functional*.

A key issue in this robust strategy is selection of an appropriate interface profile parameterization. This requires tradeoff between versatility, compactness and computational burden, bearing in mind that the number of unknown parameters to be estimated must be *smaller* than the collected reflected data size. Therefore, assuming that the collected data are *non-redundant* [23], [24], the maximum number of parameters that can be reliably estimated is $\sim N_r \times N_\lambda$. For the purposes of this paper, we assume that the surface shape can be approximated by a finite set of basis functions with unknown coefficients,

$$h(x) \approx \sum_{n=1}^N c_n s_n(x). \quad (2)$$

Let $\mathcal{E}_y^{refl}(x_p^r, \lambda_q)$ denote the complex reflected field *measured* at wavelength λ_q at receiver location x_p^r . Let \underline{c} denote the vector of coefficients $c_n, n = 1, \dots, N$. Given a vector of coefficients \underline{c} and the outgoing field from the aperture distribution in (1), we can use the Gabor-based Gaussian beam algorithm in [11] (see also Sec. III-B) for the surface profile in (2) to generate predictions of the reflected field at each receiver. Let $E_y^{refl}(x_p^r, \lambda_q; \underline{c})$ denote the complex reflected field *predicted* from \underline{c} at (free-space) wavelength λ_q at receiver location x_p^r . With this notation, we define the weighted approximation error $J(\underline{c})$ as follows:

$$J(\underline{c}) = \|\underline{E}^{refl}(\underline{c}) - \underline{\mathcal{E}}^{refl}\|^2 = \sum_{p=1}^{N_r} \sum_{q=1}^{N_\lambda} \alpha_{pq} \left| E_y^{refl}(x_p^r, \lambda_q; \underline{c}) - \mathcal{E}_y^{refl}(x_p^r, \lambda_q) \right|^2, \quad (3)$$

for receiver locations $x_1^r, \dots, x_{N_r}^r$ and operating wavelengths $\lambda_1, \dots, \lambda_{N_\lambda}$, $\alpha_{pq} > 0$ being (arbitrary) weight coefficients.

The regularized inverse scattering problem can now be formalized as finding the coefficient vector \underline{c} in (2) which minimizes the cost functional (3), i.e., finding $\hat{\underline{c}}$ such that

$$\hat{\underline{c}} = \arg \min_{\underline{c}} J(\underline{c}) \quad (4)$$

In general, the predictive model $E_y^{refl}(x_p^r, \lambda_q; \underline{c})$ is a highly nonlinear function of the coefficients \underline{c} . Thus, the resulting minimization problem may have multiple local minima. In the next section, we describe the choice of basis functions used in our representation, the forward scattering model, and the optimization approach used for determining global minima of the cost functional (3).

III. ALGORITHMIC ASPECTS

A. Interface Profile Parameterization

As our choice of basis functions in (2), we used shifted B-splines [15] selected with a fixed resolution matched to the coarse level of detail for the reconstruction. In particular, we chose a quartic-spline basis function $s^{(4)}(x)$, where

$$s^{(4)}(X) = \begin{cases} \frac{8}{115}X^4, & 0 \leq X < 1, \\ -\frac{8}{115}(5 - 20X + 30X^2 - 20X^3 + 4X^4), & 1 \leq X < 2, \\ \frac{8}{115}(155 - 300X + 210X^2 - 60X^3 + 6X^4), & 2 \leq X < 3, \\ -\frac{8}{115}(655 - 780X + 330X^2 - 60X^3 + 4X^4), & 3 \leq X < 4, \\ \frac{8}{115}(X - 5)^4, & 4 \leq X < 5, \end{cases} \quad (5)$$

and $X = x/\Delta_x$. This basis function has finite support and differentiable second derivatives (see Fig. 2). The surface profile is thus approximated as a linear combination of shifted B-spline basis functions,

$$h(x) \approx \sum_{n=-4}^{N_h-1} c_n s^{(4)}(x - x_n), \quad x_{min} \leq x \leq x_{max}, \quad (6)$$

$$x_n = x_{min} + n\Delta_x, \quad \Delta_x = (x_{max} - x_{min})/N_h. \quad (7)$$

The resulting linear combination (6) is a triply differentiable function with $N_h + 4$ degrees of freedom. We tried using cubic splines, but did not obtain satisfactory accuracy for the

forward (beam) solver. We can speculate that the differentiability of the *second* derivative (and hence of the radius of curvature) is required by the beam algorithm. In this investigation, we assume *a priori* knowledge of the scale parameter Δ_x in (6) (i.e., the number of B-spline basis functions), and focus on retrieving the unknown coefficients c_n only. A more general *adaptive* framework is presented in [25].

B. The Gabor-Based Gaussian Beam Algorithm

The forward scattering predictive model is detailed in [11]. It is based on a recently developed Gabor-based narrow-waisted Gaussian beam (NW-GB) algorithm for reflection from, and transmission through, moderately rough dielectric interfaces. The main steps of the algorithm can be summarized as follows:

1. The aperture field distribution in (1) is discretized self-consistently via Gabor expansion in terms of *narrow* Gaussian basis functions, which generate narrow-waisted *ray-like* Gaussian beams (GBs) launched from points on the aperture.
2. Each individual GB interaction with the rough interface is tracked via the complex-source-point (CSP) paraxial scheme (*quasi-real* ray tracing) developed in [14] for circular cylindrical dielectric layers, and generalized in [11] to rough surface geometries.
3. The various reflected/transmitted GB contributions are recombined at the observer.

In [11], the NW-GB algorithm has been validated and calibrated against an independently generated rigorous numerical solution [26], and has been shown to provide accurate and robust predictions over a range of calibrated combinations of the problem parameters, including moderate roughness with maximum slopes $\lesssim 40^\circ$, (average) curvature radii R_c larger than a wavelength, incidence directions far from grazing (incidence angles $\lesssim 30^\circ$ relative to z), and dielectric contrasts with $\text{Re}(\epsilon_r)$ ranging from 1.2 to 10 and $\text{Im}(\epsilon_r)$ up to 0.5. Though based on high-frequency asymptotics, the algorithm was found to provide satisfactory accuracy even for relatively *low-frequency* geometries ($R_c \sim 0.5\lambda$) and near-zone observation distances ($z^r \sim \lambda$). We refer the interested reader to [11] for theoretical and computational details, and to [27] for extension to pulsed excitation.

As noted earlier, the computational feasibility of the proposed nonlinear inverse scattering algorithm is strongly tied to the efficiency of the forward solver. In this connection, full-wave techniques are most likely not affordable in terms of computing time and

resources. Conversely, NW-GB algorithms, though not suffering from failures near caustics and other ray-field transition regions, preserve the attractive computational features of standard ray methods in the presence of large computational domains, with minimal memory requirements and typical computing times (for a field sample at a single position) of about 5 – 10ms on a 500 MHz PC, which are fairly shorter than those typical of full-wave techniques. Application of GB algorithms to inverse scattering scenarios was also suggested in [28], where they were found to provide a good tradeoff between accuracy and computational burden.

C. Optimization Strategy

As discussed in Section II, we want a vector of coefficients \underline{c} to minimize the weighted error functional (3). As stated previously, this minimization is nontrivial since the cost functional in (3) is likely non-convex with respect to \underline{c} ; therefore, unless an accurate initial guess is available, standard descent-based optimization algorithms [29] are likely to end in local minima, which correspond to *spurious* solutions. To illustrate this issue, we considered a simplified experiment where all unknown coefficients were set to their true value, with the exception of two that are left variable. The resulting cost function has two degrees of freedom, and can be easily displayed. The interface profile and experiment parameters are shown in Fig. 3. A nonphased cosine tapered aperture illumination

$$f(x) = \begin{cases} \cos(\pi x/d), & |x| \leq d/2, \\ 0, & |x| > d/2, \end{cases} \quad (8)$$

was assumed, with the aperture width d chosen so as to irradiate the region of interest while avoiding edge effects.

In this example, the surface profile was generated randomly by using 16 B-spline basis functions ($N_h = 12$). We chose c_2 and c_6 as unknown and set the remaining 14 coefficients to their true value. Figure 4 shows the resulting 2D cost function, as a function of c_2 and c_6 scaled with respect to their true values $c_2^{(true)}$, $c_6^{(true)}$, respectively. As the figure illustrates, the cost function has a deep *global* minimum at $(c_2, c_6) \approx (c_2^{(true)}, c_6^{(true)})$, but also has a number of *local* minima. This behavior was verified for multiple choices of weight coefficients α_{pq} , thus confirming that standard descent optimization techniques

(e.g., conjugate gradient [29]) may end up in local minima, and that some technique for global optimization is needed. However, popular global optimizers based on stochastic frameworks such as simulated annealing [30] and genetic algorithms [31] converge too slowly to be successfully exploited in realistic applications.

Our approach to finding global minima of (3) utilizes physics-based *multiresolution*, conceptually analogous to what in the optimization community is known as the *continuation method* [32]. The basic idea underlying continuation methods is illustrated in Fig. 5, with reference to a simple one-dimensional problem. Once a suitable smoothing parameter in the function to be minimized has been recognized, a smoothing procedure is applied in order to filter out the unwanted local minima. A standard descent minimization algorithm can be applied to the smoothed problem, yielding a rough estimation of the sought global minimum of the original problem. The smoothing is progressively relaxed, restoring the original problem, and the solution is progressively refined, using at each stage a standard descent optimizer and exploiting as initial guess the estimation obtained at the previous stage.

As typical of many inverse scattering problems, the smoothness of the cost functional (3) strongly depends on the choice of operating wavelength(s). In particular, using larger wavelengths (i.e., lower frequencies) will result in a smoother functional. Furthermore, using lower frequencies reduces the possibility of phase ambiguity, one of the major causes of local minima. As an example, in Fig. 6 is shown the reduced 2D cost function obtained using the same simulation parameters as in Fig. 4, but only the longer wavelength ($\lambda_1 = 0.2$) data. The resulting function is considerably smoother. The function still has a global minimum at $(c_2, c_6) \approx (c_2^{(true)}, c_6^{(true)})$ with a large basin of attraction, and *no* local minima. In this case, standard descent optimizers (e.g. conjugate gradient [29]) can be applied. The obtained estimate will need refining at other frequencies for two reasons: at lower frequencies, *i*) a *poorer* resolution can be expected, and *ii*) the accuracy of the beam forward solver is *poorer* [11]. Nonetheless, this first estimate provides a good initial guess that can be further refined through progressively introducing the higher-frequency information into the optimization, in the spirit of the continuation method [32]. The proposed multiresolution strategy can be thus viewed as a continuation method, where the

smoothing parameters are the weight coefficients α_{pq} associated with the different wavelengths. Among all possible ways of varying the weight coefficients α_{pq} , which corresponds to different ways of controlling the smoothing and the convergence behavior, we chose the simplest, i.e., an abrupt “on-off” variation. At each resolution stage, the frequency data to be included in the cost functional are selected by setting the corresponding weight coefficients α_{pq} to 0 or 1. This corresponds to what in the inverse scattering community is usually known as *frequency hopping* [16]–[18].

In our implementation, the partial optimization at each resolution level is performed using the Polak-Ribiere version of the conjugate gradient (CG) algorithm [29], particularly suited for non-quadratic functions. Specifically, the needed gradient of J is computed using a central difference formula, so that each gradient evaluation requires $2N_h + 8$ functional evaluations (i.e., $2N_h + 8$ solutions of a forward scattering problem), $N_h + 4$ being the number of unknown spline coefficients. The CG algorithm in [29] has been partially modified in order to enforce the consistency constraint

$$\max_x \{h(x)\} < z_A, z^r, \quad (9)$$

where z_A and z^r are the aperture and observation heights, respectively (see Fig. 1).

IV. NUMERICAL RESULTS

In order to test our surface profile estimation algorithms, we generated synthetic field measurement data using a reliable full-wave solution of the forward scattering problem by means of the multifilament current method in [26], in conjunction with rigorous Kirchhoff aperture integration [33], for a variety of surfaces. For all numerical experiments presented below, the accuracy of the NW-GB forward solver was preliminarily verified. The simulation parameters are summarized in Fig. 3, with the dielectric half-space constitutive parameters chosen so as to simulate a class of sandy soils in the GPR frequency range. In this example, there are 16 unknown spline coefficients to be estimated, based on 40 complex (magnitude and phase) field samples. Reconstruction results are shown in Fig. 7. Assuming as initial guess a flat interface ($c_n = 0$, $n = -4, \dots, N_h - 1$), the cost functional (3) was minimized using only the lowest frequency data ($\lambda_1 = 0.2$), obtaining the recon-

struction shown in Fig. 7a. Subsequent refinements of this reconstruction were obtained by including higher frequency data, resulting in the improved approximations shown in Figs. 7b-d. Specifically, each iterative improvement in Figs. 7b-d is obtained by using only a *single frequency* at a time, and exploiting the reconstruction at the previous (lower) frequency as initial guess. The more time-consuming alternative of using at each iteration the *current frequency plus* the lower ones was found not to yield significant improvement. The example illustrates that the reconstruction is accurate throughout most of the interval, except near the edges of the illuminated region; a similar phenomenon was observed in [8]. The likely explanation for this loss of accuracy is due to the weak illumination in these regions, corresponding to the aperture field tapering; the tapering is required to avoid numerical artifacts attributed to edge effects.

We performed a thorough calibration of the proposed inversion algorithm, by reconstructing a large number of randomly generated surface realizations with moderate roughness both in height and slope ($\lesssim 40^\circ$). Representative results are shown in Fig. 8. For all examples we used the four-stage frequency-hopping scheme as in Fig. 7. Again, except near the edge regions, the reconstructions are quite accurate. As a general comment, we found satisfactory reconstructions for problem parameters (roughness, permittivity, etc.) in the range of validity of the forward model summarized in Sec. III-B.

We also investigated the numerical stability of the algorithm with respect to errors in the reflected field simulated/measured data. In order to simulate the unavoidable measurement uncertainty, we added to the full-wave-computed reflected field data a uniformly distributed relative error. Furthermore, in order to roughly simulate the effect of possible clutter sources neglected in the model, such as the incoherent scattering contribution from fine-scale roughness, we added a background noise with uniformly distributed amplitude and phase. Reconstruction results are shown in Fig. 9, with the problem parameters as in Fig. 3. As one can see, the reconstruction obtained from corrupted data with a relative error of $\pm 5\%$ in amplitude and $\pm 10^\circ$ in phase, and a background noise of -20 dB, is not very different (apart from the edge regions) from that obtained using noise-free data, thus indicating the robustness of the proposed algorithm. Increasing the noise strength obviously results in a poorer reconstruction, especially in the edge regions whose *weaker*

scattering contribution is more noise-sensitive. It is interesting to notice that even with a considerably stronger background noise (-10 dB), the reconstruction of the central region of the interface profile is still relatively accurate.

Concerning the convergence rate, in the above examples an average number of $20 - 30$ conjugate gradient iterations per resolution stage was typically required, resulting in an overall computing time of ~ 1 min. on a 500 MHz PC; no particular effort was made to optimize the numerical code.

V. CONCLUSIONS

We presented a new inversion algorithm for the reconstruction of moderately rough dielectric interfaces using spatially sampled (multifrequency) reflected field data. The proposed approach is based on a compact parameterization of the unknown interface profile in terms of quartic splines, whose unknown parameters are estimated by minimizing the *difference* between model-based and measured reflected field data. The approach uses a fast forward model based on quasi-ray Gaussian beams [11]. In order to avoid local minima, a frequency hopping multiresolution approach is used, exploiting reconstructions based only on longer wavelengths to provide initial guesses for higher resolution reconstructions.

The proposed algorithm was evaluated on noisy data generated from simulated profiles, illustrating that accurate and robust reconstructions can be obtained for moderate roughness (maximum slopes $\lesssim 40^\circ$), with reasonable computing times. Thus, extensions of the algorithmic approach to three-dimensional (3D) geometries should be feasible for realistic situations where sparse (multifrequency) data and limited computing resources are available, and near real-time estimates are required.

The approach presented in this paper is similar in spirit to that of [8], although based on a different forward solver and optimization schemes. In particular, our approach generalizes naturally to sparsely sampled data.

The surface estimation algorithm has been extended to time dependent (short pulse) GPR excitation [25], and has been incorporated into adaptive techniques for subsurface GPR image reconstruction of shallowly buried plastic mine-like targets in the presence of unknown rough air-soil interfaces [19]–[22]; preliminary outcomes seem encouraging. Extensions presently under investigation include generalization to fully 3D geometries.

REFERENCES

- [1] T. Dogaru and L. Carin, "Time-domain sensing of targets buried under a rough air-ground interface," *IEEE Trans. Antennas Propagat.*, vol. 46, No. 3, pp. 360-372, Mar. 1998.
- [2] R.A. Weisenseel, W.C. Karl, D.A. Castañón, E.L. Miller, C.M. Rappaport, and C.A. DiMarzio, "Statistical fusion of GPR and EMI data," in *Detection and Remediation Technologies for Mines and Minelike Targets IV*, A.C. Dubey, J.F. Harvey, J. Broach, and R.E. Dugan, Eds., *Proc. SPIE*, vol. 3710, pp. 1179-1187, Aug. 1999.
- [3] T. Dogaru, L. Collins, and L. Carin, "Optimal time-domain detection of a deterministic target buried under a randomly rough interface," *IEEE Trans. Antennas Propagat.*, vol. 49, No. 3, pp. 313-326, Mar. 2001.
- [4] H. Zhan, C.M. Rappaport, M. El-Shenawee, and E.L. Miller, "Mine detection under rough ground surfaces using 2-D FDTD modeling and hypothesis testing," *Proc. 2001 IEEE Antennas Propagat. Int. Symposium*, Boston, MA, USA, July 8-13, 2001, vol. 3, p. 756.
- [5] H. Feng, D. A. Castañón, W. C. Karl, and E. L. Miller, "GPR imaging approaches for buried plastic landmine detection," in *Detection and Remediation Technologies for Mines and Minelike Targets V*, A.C. Dubey, J.F. Harvey, J.T. Broach, E.R. Dugan, Eds., *Proc. SPIE*, vol. 4038, pp. 1485-1496, Aug. 2000.
- [6] R.J. Wombell and J.A. DeSanto, "The reconstruction of shallow rough-surface profiles from scattered field data," *Inverse Probl.*, vol. 7, pp. L7-L12, 1991.
- [7] R.J. Wombell and J.A. DeSanto, "Reconstruction of rough-surface profiles with the Kirchhoff approximation," *J. Opt. Soc. Am. A*, vol. 8, No. 12, pp. 1892-1897, Dec. 1991.
- [8] C. Ying and A. Noguchi, "Rough surface inverse scattering problem with Gaussian beam illumination," *IEICE Trans. Electron.*, vol. E77-C, No. 11, pp. 1781-1785, Nov. 1994.
- [9] K. Harada and A. Noguchi, "Reconstruction of two dimensional rough surface with Gaussian beam illumination," *IEICE Trans. Electron.*, vol. E79-C, No. 10, pp. 1345-1349, Oct. 1996.
- [10] A. Schatzberg and A.J. Devaney, "Rough surface inverse scattering with the Rytov approximation," *J. Opt. Soc. Am. A*, vol. 10, No. 5, pp. 942-950, May 1993.
- [11] V. Galdi, L.B. Felsen, and D.A. Castañón, "Quasi-ray Gaussian beam algorithm for time-harmonic two-dimensional scattering by moderately rough interfaces," *IEEE Trans. Antennas Propagat.*, vol. 49, No. 9, pp. 1305-1314, Sept. 2001.
- [12] J.J. Maciel and L.B. Felsen, "Systematic study of fields due to extended apertures by Gaussian beam discretization," *IEEE Trans. Antennas Propagat.*, vol. 37, No. 7, pp. 884-892, July 1989.
- [13] J.J. Maciel and L.B. Felsen, "Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part I - plane layer," *IEEE Trans. Antennas Propagat.*, vol. 38, No. 10, pp. 1607-1617, Oct. 1990.
- [14] J.J. Maciel and L.B. Felsen, "Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part II - circular cylindrical layer," *IEEE Trans. Antennas Propagat.*, vol. 38, No. 10, pp. 1618-1624, Oct. 1990.
- [15] L.L. Shumaker, *Spline Functions: Basic Theory*. New York (NY): Wiley, 1981.
- [16] W.C. Chew and J.H. Lin, "A frequency-hopping approach for microwave imaging of large inhomogeneous bodies," *IEEE Microwave Guided Wave Lett.*, vol. 5, pp. 439-441, 1995.
- [17] A.G. Tijhuis and K. Belkebir, "Using multiple frequency information in the iterative solution of a two-dimensional nonlinear inverse problem," in *Proc. Progress Electromagnetic Research Symp. (PIERS '96)*, Innsbruck, Austria, 1996.

- [18] O.M. Bucci, L. Crocco, T. Isernia, and V. Pascazio, "Inverse scattering problems with multifrequency data: Reconstruction capabilities and solution strategies," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, No. 4 (Part 1), pp. 1749-1756, July 2000.
- [19] V. Galdi, W.C. Karl, D. A. Castañón, and L. B. Felsen, "Approaches to underground imaging for object localization," in *Detection and Remediation Technologies for Mines and Minelike Targets VI*, A.C. Dubey, J.F. Harvey, J.T. Broach, and V. George, Eds., *Proc. SPIE*, vol. 4394, Oct. 2001 (in print).
- [20] V. Galdi, L.B. Felsen, and D.A. Castañón, "Gaussian beam algorithm for rough surface underground imaging," *Proc. 7th Int. Conference on Electromagnetics in Advanced Applications (ICEAA '01)*, Torino, Italy, Sept. 10-14, 2001.
- [21] V. Galdi, H. Feng, D. A. Castañón, W.C. Karl, and L. B. Felsen, "Moderately rough surface underground imaging via short-pulse quasi-ray Gaussian beams," submitted to *IEEE Trans. Antennas Propagat.*, Oct. 2001.
- [22] V. Galdi, H. Feng, D.A. Castañón, W.C. Karl, and L.B. Felsen, "Multifrequency subsurface sensing in the presence of a moderately rough air-soil interface via quasi-ray Gaussian beams," invited paper in *Radio Science*, Special Issue on 2001 URSI EMT Symposium, submitted Nov. 2001.
- [23] O.M. Bucci and T. Isernia, "Electromagnetic inverse scattering: retrievable information and measurement strategies," *Radio Science*, vol. 32, No. 6, pp. 2123-2137, Nov.-Dec. 1997.
- [24] O.M. Bucci, L. Crocco, and T. Isernia, "Improving the reconstruction capabilities in inverse scattering problems by exploitation of close-proximity setups," *J. Opt. Soc. Am. A*, vol. 16, No. 7, pp. 1788-1798, July 1999.
- [25] V. Galdi, J. Pavlovich, W.C. Karl, D.A. Castañón, and L. B. Felsen, "Moderately rough dielectric interface reconstruction via short-pulse quasi-ray Gaussian beams," submitted to *IEEE Trans. Antennas Propagat.*, Oct. 2001.
- [26] Y. Leviatan and A. Boag, "Analysis of electromagnetic scattering from dielectric cylinders using a multifilament current model," *IEEE Trans. Antennas Propagat.*, vol. 35, No. 10, pp. 1119-1127, Oct. 1987.
- [27] V. Galdi, L. B. Felsen, and D. A. Castañón, "Quasi-ray Gaussian beam algorithm for short-pulse two-dimensional scattering by moderately rough dielectric interfaces," submitted to *IEEE Trans. Antennas Propagat.*, Apr. 2001.
- [28] B. Rao and L. Carin, "Beam-tracing-based inverse scattering for general aperture antennas," *J. Opt. Soc. Am. A*, vol. 16, No. 9, pp. 2219-2231, Sept. 1999.
- [29] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge (UK): Cambridge Univ. Press, 1992.
- [30] E. Aarts, and J. Korst, *Simulated Annealing and Boltzmann Machines: a Stochastic Approach to Combinatorial Optimization and Neural Computing*. New York (NY): Wiley, 1989.
- [31] K.F. Man, K.S. Tang, and S. Kwong, *Genetic Algorithms: Concepts and Designs*. New York (NY): Springer, 1999.
- [32] R. Horst and P.M. Pardalos (Eds.), *Handbook of Global Optimization*, Dordrecht: Kluwer, 1995.
- [33] B.Z. Steinberg, H. Heyman, and L.B. Felsen, "Phase-space methods for radiation from large apertures," *Radio Sci.*, vol. 26, No. 1, pp. 219-227, Jan.-Feb. 1991.

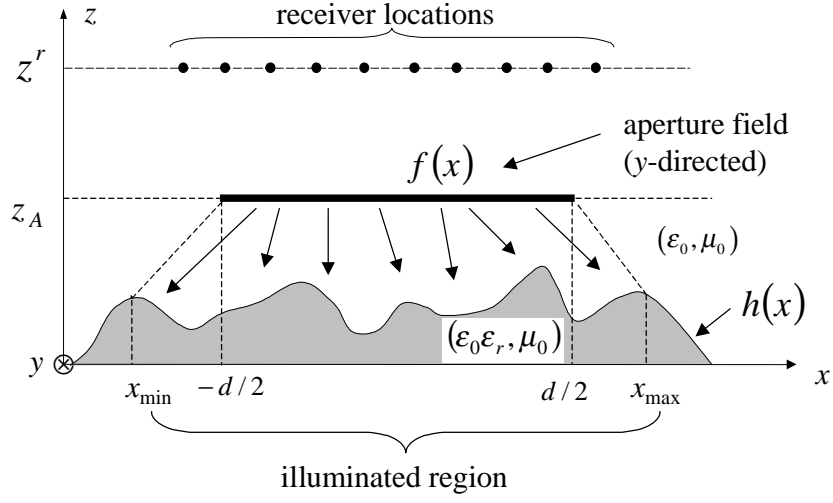


Fig. 1. Setup geometry for the inverse scattering problem. An aperture-excited, time-harmonic, TM-polarized EM field illuminates a homogeneous dielectric half-space with known relative permittivity ϵ_r and with a moderately rough interface whose coarse scale profile is described by the continuous function $h(x)$. The reflected field is sampled at N_r receiver locations at $z = z^r$.

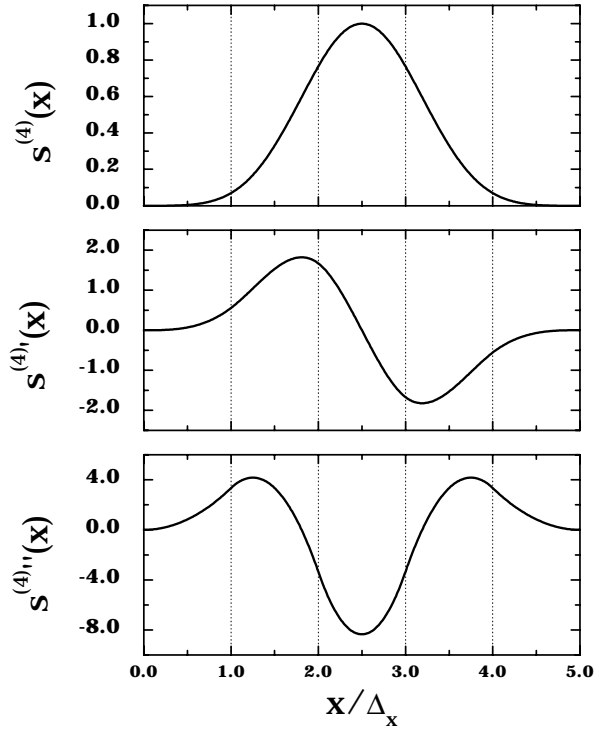


Fig. 2. Quartic B-spline basis function and its first two derivatives. ' denotes (d/dx) .

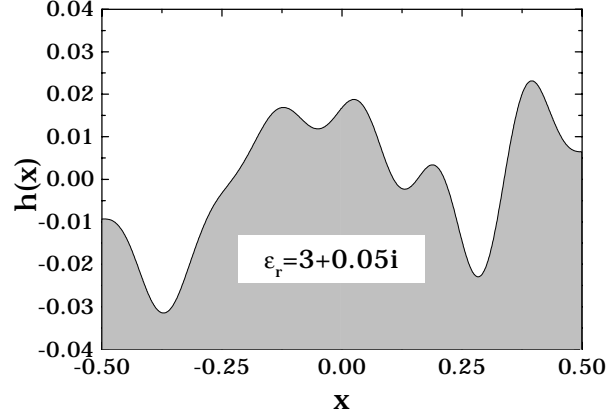


Fig. 3. Interface profile and simulation parameters (in arbitrary units) for the synthetic experiment discussed in Sec. III-C. Spline parameters: $N_h = 12$, $\Delta_x = 1/12$, $x_{min} = -0.5$, $x_{max} = 0.5$, maximum slope $\phi_{max} = 34^\circ$; aperture parameters (nonphased cosine distribution (8)): $d = 0.8$, $z_A = 0.6$; observation points: $N_r = 10$, $z^r = 0.6$, $x_p^r = -0.5 + (p - 1)/(N_r - 1)$, $p = 1, \dots, N_r$; operating wavelengths: $N_\lambda = 4$, $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, $\lambda_3 = 0.067$, $\lambda_4 = 0.05$; weight coefficients: $\alpha_{pq} = 1 \forall p, q$. Relative permittivity: $\epsilon_r = 3 + 0.05i$.

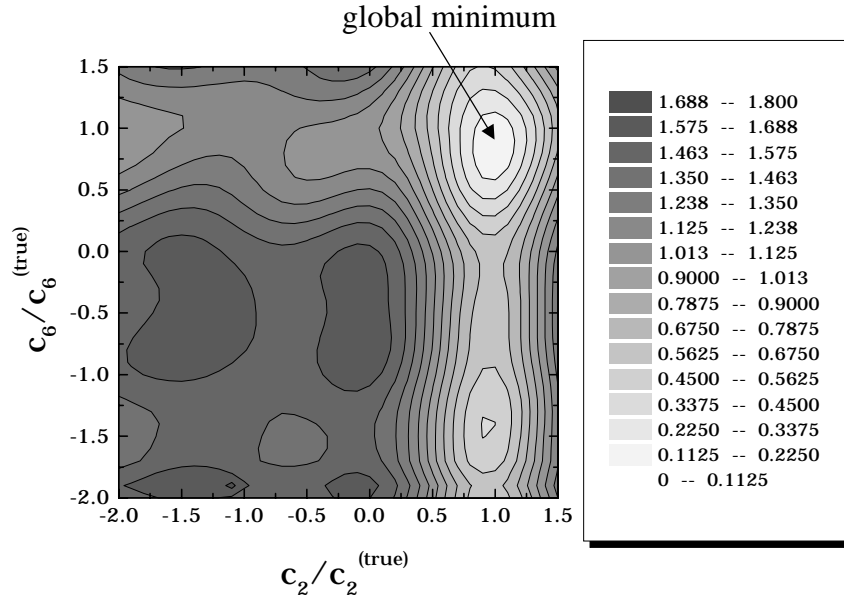


Fig. 4. Parameters as in Fig. 3. Reduced 2D cost functional (3) as a function of c_2 and c_6 scaled to their true values $c_2^{(true)}$, $c_6^{(true)}$. The remaining coefficients are set to their true value.

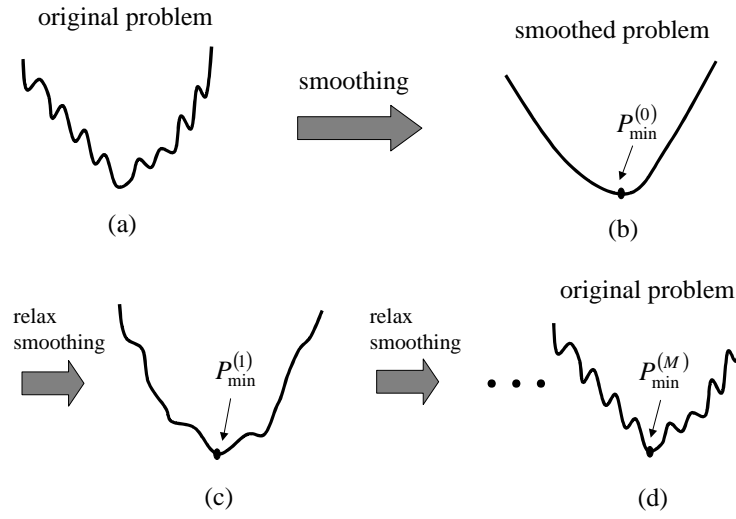


Fig. 5. Basic underlying scheme for continuation methods. (a): Original problem; (b): Local minima are filtered out through a smoothing procedure and a first rough estimate $P_{min}^{(0)}$ of the global minimum is obtained via a standard descent optimizer; (c): The smoothing is partially relaxed and a refined estimate $P_{min}^{(1)}$ is obtained using $P_{min}^{(0)}$ as initial guess; (d): The smoothing is gradually removed in M iterations, restoring the original problem and obtaining the final estimate $P_{min}^{(M)}$.

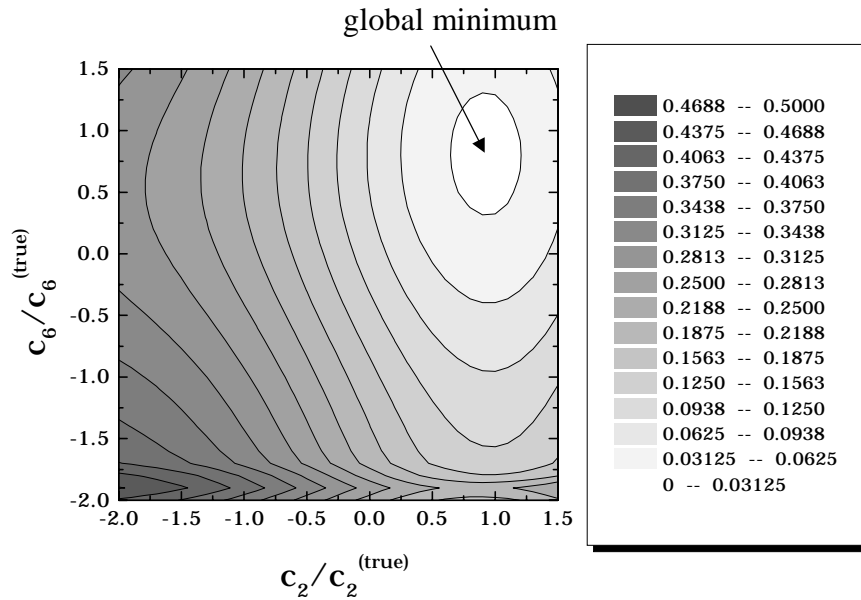


Fig. 6. Reduced 2D cost functional as in Fig. 4, but using the longer wavelength ($\lambda_1 = 0.2$) data only.

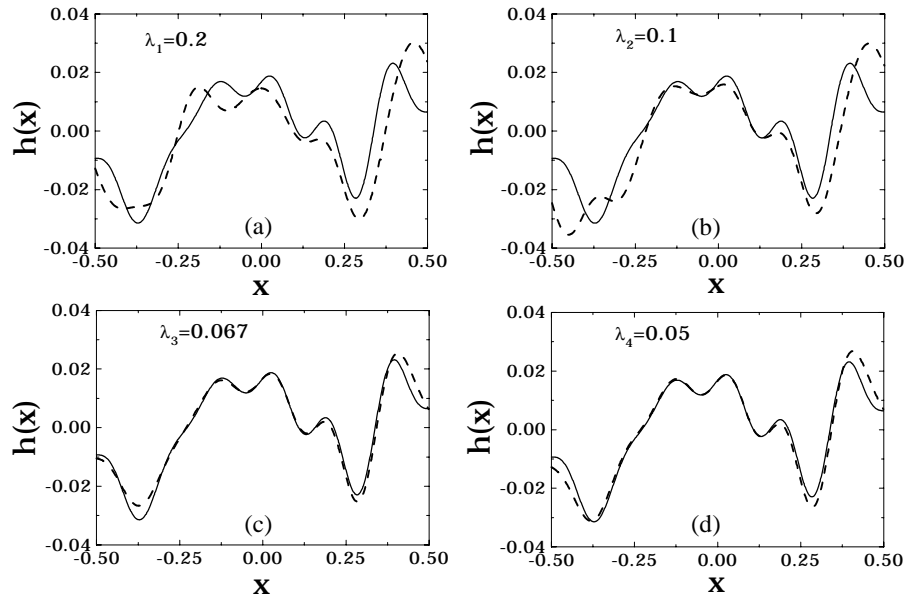


Fig. 7. Parameters as in Fig. 3. Convergence example for the proposed frequency-hopping strategy. (a): Reconstruction using only the $\lambda_1 = 0.2$ data and a flat interface ($c_n = 0, n = -4, \dots, N_h - 1$) initial guess; (b), (c), (d): Refinements obtained using single-frequency data for $\lambda_2 = 0.1$, $\lambda_3 = 0.067$, $\lambda_4 = 0.05$, respectively, and the previous stage reconstruction as initial guess.
— Actual profile; - - - Reconstruction.

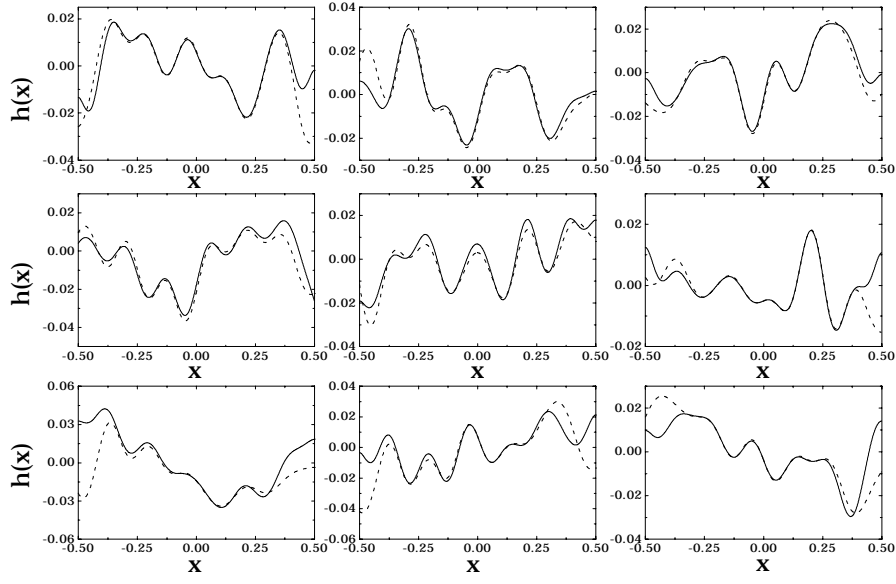


Fig. 8. Reconstruction of randomly chosen example profiles via the four-stage frequency-hopping strategy in Fig. 7. Spline parameters: $N_h = 12$, $\Delta_x = 1/12$, $x_{min} = -0.5$, $x_{max} = 0.5$; aperture parameters (nonphased cosine distribution (8)): $d = 0.8$, $z_A = 0.6$; observation points: $N_r = 10$, $z^r = 0.6$, $x_p^r = -0.5 + (p-1)/(N_r-1)$, $p = 1, \dots, N_r$; operating wavelengths: $N_\lambda = 4$, $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, $\lambda_3 = 0.067$, $\lambda_4 = 0.05$. — Actual profile; - - - Reconstruction.

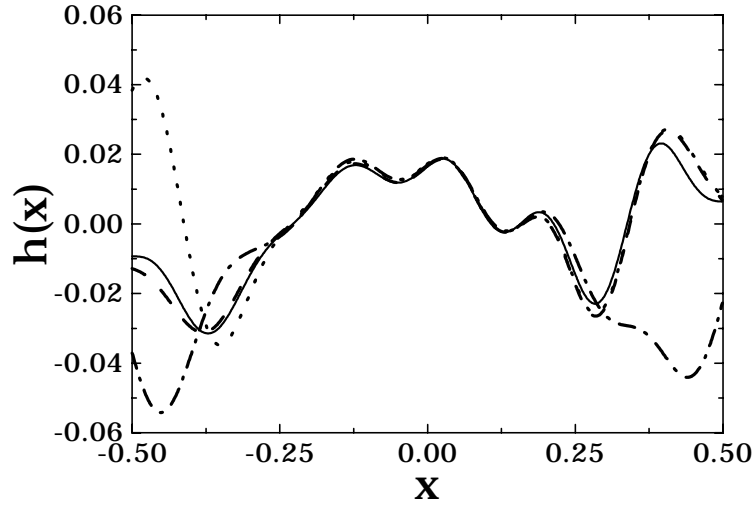


Fig. 9. Parameters as in Fig. 3. Reconstruction from noisy data via the four-stage frequency-hopping strategy in Fig. 7. — Actual profile; - - - Reconstruction (noise-free data); Reconstruction (relative error: $\pm 5\%$ in amplitude, $\pm 10^\circ$ in phase; background noise: -20 dB); - · - · - Reconstruction (relative error: $\pm 5\%$ in amplitude, $\pm 10^\circ$ in phase; background noise: -10 dB).

Time-Domain Radiation from Large Two-Dimensional Apertures via Narrow-Waisted Gaussian Beams

Vincenzo Galdi, *Member, IEEE*, Leopold B. Felsen, *Life Fellow, IEEE*,
and David A. Castañon, *Senior Member, IEEE*

This work was supported by ODDR&E under MURI Grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from Grant No. 9900448 by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University, Brooklyn, NY 11201 USA.

V. Galdi is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, on leave from University of Sannio, Benevento, Italy (e-mail: vgaldi@bu.edu).

L.B. Felsen is with the Department of Aerospace and Mechanical Engineering and the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, part-time. He is also University Professor Emeritus, Polytechnic University, Brooklyn, NY 11201 USA (e-mail: lfelsen@bu.edu).

D.A. Castañon is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA (e-mail: dac@bu.edu).

Abstract

This paper deals with the short-pulse radiation of three-dimensional (3D) vector electromagnetic fields from arbitrarily polarized large two-dimensional (2D) truncated aperture distributions, which are parameterized in terms of narrow-waisted ray-like pulsed Gaussian basis beams centered on a discretized Gabor lattice in a four-dimensional configuration-spectrum phase space. The study extends our previous Gabor-based investigation of time domain (TD) short-pulse radiation of 2D fields from 1D large truncated apertures with nonphased, linearly phased (delayed) and nonlinearly phased focusing aperture field profiles [1]. We begin with, and summarize, a Gabor-based frequency domain (FD) formulation of the 2D aperture problem which has been presented and tested elsewhere [2], [3], but we include additional numerical examples for validation and quality assessment. As in [1], we access the time domain by Fourier inversion from the FD, starting from the initial 3D space-time Kirchhoff formulation (whose numerical integration furnishes reference solutions), and then passing on to Gabor-parameterized field representations in terms of pulsed beam (PB) wavepackets which are launched by linearly and nonlinearly phase-delayed focusing aperture distributions. Example calculations and comparisons with numerically generated reference data serve to calibrate the Gabor-PB algorithms and assess their domains of validity.

Keywords

Gaussian beams, pulsed beam wavepackets, Gabor lattice representations.

I. INTRODUCTION

In an ongoing series of recent investigations, we have revisited a previously formulated discretized Gabor-based narrow-waisted Gaussian beam (GB) algorithm for two-dimensional (2D) time-harmonic high-frequency radiation from truncated 1D planar aperture distributions [4], which was then applied to the interaction of these GB-parameterized radiated fields with complex propagation environments [5], [6]. The algorithm had been found to be robust, accurate, physically appealing, and numerically efficient when compared with conventional Kirchhoff Physical Optics integration procedures. Our recent interest has been motivated in part by the need of robust forward solvers in forward and inverse scattering (identification and imaging) scenarios concerned with objects embedded within an irregularly bounded penetrable semi-infinite medium. In a stepwise approach toward constructing the necessary algorithms, we have proceeded along two parallel routes: 1. extension of the frequency domain (FD) algorithms for the 1D aperture/2D field configuration to the new rough interface propagation environment, and to the new general case of 2D aperture/3D vector fields; 2. extension of the FD results to the short-pulse-excited

time domain (TD). The FD interaction of the 1D aperture radiated field with a moderately rough interface has been addressed in [7] and [8] for the forward problem and in [9] and [10] for the inverse problem, whereas the 2D aperture/3D field radiation problem has been addressed in [2] and [3]. The TD extension of the 1D aperture problem has been carried out in [1], and the extension to 2D apertures is the subject of the present paper. The contents of this paper having been summarized in the abstract, we pass on directly to the problem formulation and its solution via the Gabor-based narrow-waisted beam algorithm.

II. FORMULATION OF THE PROBLEM

We consider the radiation of three-dimensional (3D) vector electromagnetic (EM) fields excited by pulsed arbitrarily polarized extended two-dimensional (2D) vector field distributions in a truncated planar aperture of characteristic dimension d_A on the $z = 0$ plane of Fig. 1. We shall only deal with the vector electric field $\mathbf{e}(\mathbf{r}, t)$, $\mathbf{r} = x\mathbf{u}_x + y\mathbf{u}_y + z\mathbf{u}_z$; the vector magnetic fields can be derived from Maxwell's equations. Here and henceforth, bold face symbols denote vector quantities, and \mathbf{u}_α denotes a unit vector in the α -direction. The 2D space-time aperture field distribution will be specified as

$$\mathbf{e}_0(x, y, z = 0, t) = \mathbf{f}(x, y, t), \quad (x, y) \in \Gamma_A, \quad \mathbf{u}_z \cdot \mathbf{f} = 0. \quad (1)$$

The corresponding field radiated into the halfspace $z > 0$ can be obtained by space-time superposition, using the 3D time-domain (TD) Green's function

$$g_{3D}(\mathbf{r}, t; \mathbf{r}'_0, t') = \frac{\delta(t - t' - R/c)}{4\pi R}, \quad R = |\mathbf{r} - \mathbf{r}'_0|, \quad (2)$$

where $\mathbf{r}'_0 = x'\mathbf{u}_x + y'\mathbf{u}_y$, $\delta(\cdot)$ represents the Dirac delta function, and c is the ambient propagation speed. The subscript “ $_0$ ” identifies quantities in the $z = 0$ plane. The resulting field radiated into the halfspace $z > 0$, synthesized by superposition of point source contributions (Kirchhoff integration), is given by [11, p. 175]

$$\mathbf{e}(\mathbf{r}, t) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\mathbf{r} - \mathbf{r}'_0) \times \left\{ \mathbf{u}_z \times \left[\frac{\mathbf{f}(x', y', t - R/c)}{R^3} + \frac{1}{cR^2} \frac{\partial}{\partial t} \mathbf{f}(x', y', t - R/c) \right] \right\} dx' dy'. \quad (3)$$

In Sec. IV-D, we shall be using a reference solution based on Gaussian quadrature numerical integration [12] of (3). In what follows, the time domain is accessed by Fourier inversion from the frequency domain (FD) via the Fourier transform pair

$$\mathbf{e}(\mathbf{r}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{E}(\mathbf{r}, \omega) \exp(-i\omega t) d\omega, \quad \mathbf{E}(\mathbf{r}, \omega) = \int_{-\infty}^{\infty} \mathbf{e}(\mathbf{r}, t) \exp(i\omega t) dt. \quad (4)$$

Capital letters identify FD field quantities.

III. RELEVANT FD RESULTS

This section contains a brief summary of relevant frequency domain (FD) results from [2], [3] and some new additions. The FD aperture field corresponding to (1) has an implicit $\exp(-i\omega t)$ time dependence and a spatial distribution $\mathbf{F}(x, y, \omega)$ (related to \mathbf{f} via the Fourier transform pair in (4)) over the aperture Γ_A in Fig. 1,

$$\mathbf{E}_0(x, y, z = 0, \omega) = \mathbf{F}(x, y, \omega), \quad (x, y) \in \Gamma_A, \quad \mathbf{u}_z \cdot \mathbf{F} = 0. \quad (5)$$

The resulting electric field in the halfspace $z > 0$, obtained by point-source (Kirchhoff) integration in configuration space, is given by [11, p. 107]

$$\mathbf{E}(\mathbf{r}, \omega) = 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \nabla G_{3D}(\mathbf{r}, \mathbf{r}'_0; \omega) \times [\mathbf{u}_z \times \mathbf{F}(x', y', \omega)] dx' dy', \quad (6)$$

where

$$G_{3D}(\mathbf{r}, \mathbf{r}'_0; \omega) = \exp(ikR)/(4\pi R), \quad R = |\mathbf{r} - \mathbf{r}'_0|, \quad (7)$$

is the 3D free-space FD Green's function, with $k = \omega/c = 2\pi/\lambda$ as the ambient wavenumber and λ as the wavelength. It can easily be verified (see [11, p. 175]) that $\mathbf{e}(\mathbf{r}, t)$ in (3) and $\mathbf{E}(\mathbf{r}, \omega)$ in (6) are related through the Fourier transform pair (4). In the remainder of this section, the ω dependence in \mathbf{E} and \mathbf{F} will be omitted for simplicity of notation, and the discussion relies heavily on references [1]–[3], with precise citations attached to each of the results extracted from these papers. Notationally, we use (2.5) to identify eqn. (5) in ref. [2], for example. We also call attention to the following differences between [2], [3] and the present paper: 1. [2] and [3] deal with the calculation of the vector potential \mathbf{P} whereas here we calculate the vector field \mathbf{E} . 2. The aperture potential field in [2] and [3]

is denoted by \mathbf{f} which should not be confused with the TD \mathbf{f} in (1) here. These differences do not affect the *basic format* of the Gabor expansions to which the citations refer.

A. Discretized Gabor-Based Field Representations in the Configuration-Spectrum Phase Space

A.1 Aperture Field

The vector aperture field $\mathbf{F}(x, y)$ can be expressed in the 2D Gabor Gaussian beam (GB) basis as follows (see (2.4))

$$\mathbf{F}(x, y) = \sum_{m,n,p,q=-\infty}^{\infty} \mathbf{A}_{mnpq} w(x - mL_x, y - pL_y) \exp[i(n\beta_x x + q\beta_y y)], \quad (8)$$

with $w(x, y)$ representing the normalized 2D Gaussian window in (2.10a),

$$w(x, y) = \left(\frac{2}{L_x L_y} \right)^{1/2} \exp \left\{ -\pi \left[(x/L_x)^2 + (y/L_y)^2 \right] \right\}, \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w^2(x, y) dx dy = 1. \quad (9)$$

The representation in (8) is tied to a four-dimensional lattice in the $(x, y; k_x, k_y)$ phase-space (Fig. 2), k_x, k_y being the x and y -domain wavenumbers, respectively, which can be expressed in terms of the spherical polar coordinate angles θ, ϕ (see Fig. 1) as

$$k_x = k \sin \theta \cos \phi, \quad k_y = k \sin \theta \sin \phi, \quad k_z = \sqrt{k^2 - k_x^2 - k_y^2} = k \cos \theta. \quad (10)$$

In this Gabor basis, each lattice point locates a Gaussian beam, whose initial profile matches the corresponding basis function. Indexes (m, p) tag x, y spatial shifts, respectively, whereas (n, q) tag k_x, k_y spectral shifts (i.e., beam tilts θ_{nq}, ϕ_{nq}), respectively. Spatial and spectral periods are related by the self-consistency relation (configuration-spectrum tradeoff) (2.4a),

$$L_x \beta_x = L_y \beta_y = 2\pi. \quad (11)$$

From (10), as the spectral shifts increase, the corresponding (n, q) beam polar angles $\theta_{nq} = \sin^{-1}[(k_x^2 + k_y^2)^{1/2}/k]_{nq}$, i.e.,

$$\theta_{nq} = \sin^{-1} \left[\frac{\sqrt{(n\beta_x)^2 + (q\beta_y)^2}}{k} \right] = \sin^{-1} \left[\sqrt{\left(\frac{n\lambda}{L_x} \right)^2 + \left(\frac{q\lambda}{L_y} \right)^2} \right], \quad (12)$$

become *complex* for $[(n\lambda/L_x)^2 + (q\lambda/L_y)^2] > 1$, and therefore the corresponding beams away from the aperture (see Sec. III-A.2) become *evanescent*. In the second equality of (12), use has been made of (11) and $k = 2\pi/\lambda$.

The expansion coefficients in (8) can be computed by introducing an auxiliary *bi-orthogonal* function $\gamma(x, y)$ (cf. (2.8)),

$$\mathbf{A}_{mnpq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{F}(x, y) \gamma^*(x - mL_x, y - pL_y) \exp[-i(n\beta_x x + q\beta_y y)] dx dy, \quad (13)$$

where * denotes the complex conjugate. For the Gaussian window in (9), a closed-form expression for $\gamma(x, y)$ is derived in [2, App. C]. Alternatively, the Gabor coefficients can be computed effectively by schemes employed in image processing [13]. However, we shall avoid this calculation later on by specializing to narrow-waisted Gaussian basis functions (see Sec. III-B).

A.2 Radiated Field

The Gabor lattice representation for the field radiated into the halfspace $z > 0$ follows by combining (6) and (8),

$$\mathbf{E}(\mathbf{r}) = \sum_{m,n,p,q=-\infty}^{\infty} (\mathbf{A}_{mnpq} \times \mathbf{u}_z) \times \mathbf{B}_{mnpq}(\mathbf{r}), \quad (14)$$

with the (vector) beam propagator \mathbf{B}_{mnpq} expressed by Gabor-weighted point-source superposition as

$$\mathbf{B}_{mnpq}(\mathbf{r}) = 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \nabla G_{3D}(\mathbf{r}, \mathbf{r}_0) w(x' - mL_x, y' - pL_y) \exp[i(n\beta_x x' + q\beta_y y')] dx' dy'. \quad (15)$$

B. Narrow-Waisted Beams

When the observer is located in the *paraxial far-zone* of a beam, the integral in (15) can be evaluated by saddle-point asymptotics. It is shown in [3, Sec. III], that the result can be

expressed in an explicit closed form in terms of *complex source point* (CSP) propagators, provided the beams are nontilted ($n = q = 0$), narrow-waisted ($L_{x,y} \lesssim \lambda \ll d_A$), and distributed over a symmetric lattice ($L_x = L_y = L$). Narrow-waisted (ray-like) GBs have been shown in a series of previous studies of 2D fields excited by 1D apertures [5]–[8] to yield highly efficient and accurate algorithms for synthesis of interactions with complex propagation and scattering environments. The narrow-waisted nontilted GB can be related to the 3D-CSP Green’s function \tilde{G}_{3D} as follows (cf. (3.14))

$$\tilde{\mathbf{B}}_{m0p0}(\mathbf{r}) \sim 2\sqrt{2}L \exp(-kb) \nabla \tilde{G}_{3D}(\mathbf{r}, \tilde{\mathbf{r}}'_{mp}) \quad (16)$$

$$\sim ikL \frac{\exp[ik(\tilde{R}_{mp} + ib)]}{\sqrt{2\pi}\tilde{R}_{mp}^2} (\mathbf{r} - \tilde{\mathbf{r}}'_{mp}), \quad (17)$$

where

$$\tilde{R}_{mp} = |\mathbf{r} - \tilde{\mathbf{r}}'_{mp}| = \sqrt{(x - x'_m)^2 + (y - y'_p)^2 + (z - \tilde{z}')^2}, \quad \text{Re}(\tilde{R}_{mp}) \geq 0, \quad (18)$$

is the complex distance between the real observation point \mathbf{r} and the complex source point

$$\tilde{\mathbf{r}}'_{mp} = x'_m \mathbf{u}_x + y'_p \mathbf{u}_y + \tilde{z}' \mathbf{u}_z = mL \mathbf{u}_x + pL \mathbf{u}_y + ib \mathbf{u}_z. \quad (19)$$

Here and henceforth the tilde \sim identifies dependence on analytically continued spatial source coordinates as well as the field produced thereby. The complex displacement parameter is given by (3.11c),

$$\tilde{z}' = ib, \quad b = L^2/\lambda; \quad (20)$$

one notes that b is the beam Fresnel length corresponding to an “effective” beam waist L . Equations (16) and (17) are valid in the paraxial far-zone of each beam, $|\tilde{R}_{mn}| \gg b$, and are restricted to symmetric and nontilted beams only. However, recalling (12), one observes that these nontilted sufficiently *narrow* beams with $n = q = 0$, $L \lesssim \lambda \ll d_A$, yield the dominant contribution to the radiated field because all other beams are evanescent.

An important attractive feature of narrow-waisted GBs is that their Gabor coefficients can be estimated effectively by *aperture sampling* (see (3.19)), avoiding the time-consuming operations required in (13) and (2.11a,b) or [13],

$$\mathbf{A}_{mpq} \approx \begin{cases} (L/\sqrt{2})\mathbf{F}(mL, pL), & n, q = 0, \\ 0, & n, q \neq 0, \end{cases} \quad (21)$$

so that

$$\tilde{\mathbf{E}}(\mathbf{r}) \sim \sum_{m,p} (\mathbf{A}_{m0p0} \times \mathbf{u}_z) \times \tilde{\mathbf{B}}_{m0p0}(\mathbf{r}). \quad (22)$$

Due to the finite extent of the aperture, the (m, p) summations in (22) include only a finite number of beams, given roughly by $(d_A/L)^2$. The approximation in (21) and (22) neglects the tilted $(n, q \neq 0)$ beams in the Gabor expansion which, however, are *evanescent* for $[(n\lambda/L_x)^2 + (q\lambda/L_y)^2] > 1$, as noted earlier. Subject to the constraints $n = q = 0$, $L \lesssim \lambda \ll d_A$, the CSP paraxial *far-zone* approximation (17) can be invoked at *moderate* distance, within and beyond the radiating *near-zone* of the *aperture*. The resulting beam synthesis in (22), with (21) and (17), yields an efficient algorithm which has been tested and validated in Sec. III-D.

C. Linearly Phased Aperture

For linearly-phased distributions, the aperture field becomes

$$\mathbf{F}(x, y) = \mathbf{g}(x, y) \exp[ik \sin \theta_A (x \cos \phi_A + y \sin \phi_A)], \quad (23)$$

where $\mathbf{g}(x, y)$ is a real function and θ_A, ϕ_A specify the direction of the main radiation lobe. Now, a more effective implementation of the narrow-waisted beam discretization can be achieved by exploiting *propagation-matched tilted* beams, which has not been investigated in [3]. As for the 2D field/1D aperture problem in [1], this is achieved by Gabor-expanding the real function $\mathbf{g}(x, y)$ only, and including the linear phasing in the beam integral (15) for the \mathbf{B}_{m0p0} beam propagator. Accordingly, generalizing the 1D aperture expressions in (1.18)–(1.20) to the 2D aperture case, the narrow-waisted beam expansion (22) can be recast as (for simplicity, the subscript “ $m0p0$ ” is henceforth replaced by “ mp ”)

$$\tilde{\mathbf{E}}(\mathbf{r}) \sim \sum_{m,p} (\mathbf{C}_{mp} \times \mathbf{u}_z) \times \tilde{\mathbf{B}}_{mp}(\mathbf{r}), \quad (24)$$

where

$$\mathbf{C}_{mp} \approx (L/\sqrt{2})\mathbf{g}(mL, pL), \quad (25)$$

$$\tilde{\mathcal{B}}_{mp}(\mathbf{r}) = ikL \frac{\exp[ik(\tilde{R}_{mp} + ib + \mathcal{L}_{mp})]}{\sqrt{2\pi}\tilde{R}_{mp}^2} (\mathbf{r} - \tilde{\mathbf{r}}'_{mp}), \quad (26)$$

$$\mathcal{L}_{mp} = L \sin \theta_A (m \cos \phi_A + p \sin \phi_A), \quad (27)$$

$$\tilde{R}_{mp} = |\mathbf{r} - \tilde{\mathbf{r}}'_{mp}|, \quad (28)$$

$$\begin{aligned} \tilde{\mathbf{r}}'_{mp} &= \tilde{x}'_m \mathbf{u}_x + \tilde{y}'_p \mathbf{u}_y + \tilde{z}' \mathbf{u}_z = \\ &= (mL + ib \sin \theta_A \cos \phi_A) \mathbf{u}_x + (pL + ib \sin \theta_A \sin \phi_A) \mathbf{u}_y + ib \cos \theta_A \mathbf{u}_z, \end{aligned} \quad (29)$$

$$b = (L \cos \theta_A)^2 / \lambda. \quad (30)$$

Instead of \mathbf{A}_{mp} , \mathbf{C}_{mp} in (24) and (25) identifies Gabor coefficients associated with the tilt-matched linearly phased aperture formulation. The beam propagator (26) differs from $\tilde{\mathbf{B}}_{m0p0}$ in (17) by the phase shift ($ik\mathcal{L}_{mp}$) and by the different definitions of the CSP $\tilde{\mathbf{r}}'_{mp}$ and b in (29) and (30), respectively (see (1.21) and (1.22)), which produce the propagation-matched tilt (θ_A, ϕ_A) in the beam direction. Equation (26) can be obtained by the same procedure that leads to (17), or even by inspection, recalling the 1D aperture results [1]. As shown in Sec. III-D below, for linearly phased apertures, the tilted beam expansion (24) allows significantly coarser lattice sampling than (22), for the same degree of accuracy. For *moderately* nonlinear phasing it may still be convenient to split the phase function into a linear term (included in the beam propagator as above) and a nonlinear remainder (included in the Gabor-expanded function \mathbf{g}), but no particular benefit is expected for *strongly* nonlinear phasing (e.g., strongly focused apertures).

D. Numerical Results

In order to compare the nontilted and tilted formulations in (22) and (24), respectively, we consider the linearly phased distribution in (23), with an x -polarized tapered square cosine amplitude distribution

$$\mathbf{g}(x, y) = \begin{cases} \cos(\pi x/d_A) \cos(\pi y/d_A) \mathbf{u}_x, & |x|, |y| \leq d_A/2, \\ 0, & |x|, |y| > d_A/2. \end{cases} \quad (31)$$

It follows from (6) that the resulting vector electric field radiated into the $z > 0$ half-space has x, z components only. All other parameters used in our simulations are listed in the figure captions. In our simulations, we assume $\theta_A = 30^\circ$, $\phi_A = 45^\circ$, and a fixed (symmetric) beam lattice period ($L_x = L_y = L = 0.8\lambda$). In Figs. 3a and 3b, the x -components of the radiated fields synthesized under the same conditions with the same number ($(d_A/L)^2 \sim 150$) of narrow-waisted nontilted (22) and tilted (24) beams are compared with the reference solution (Gaussian quadrature numerical integration of (6)), showing representative x - and y -cuts, respectively, at a fixed value of z well within the aperture near-zone. The tilted beam synthesis is barely distinguishable from the reference solution, whereas the nontilted synthesis is noticeably more *inaccurate*. The accuracy is more clearly quantified in Figs. 3c and 3d, where the relative error magnitudes

$$\Delta E_{x,z} \equiv \frac{|E_{x,z}^{(ref)} - E_{x,z}^{(beam)}|}{\max |E_{x,z}^{(ref)}|}, \quad (32)$$

are displayed as gray-scale plots. It is observed that the tilted beam synthesis attains a maximum error $< -30\text{dB}$, whereas the nontilted beam synthesis error is evidently greater. In order to obtain comparable accuracy with nontilted beams, a considerably smaller beam lattice period ($L \sim 0.1\lambda$) should be used, resulting in an overall computation time increase of almost *two orders of magnitude*. The corresponding results for the z component are shown in Fig. 4, and the same considerations apply. The nonphased case ($\theta_A = 0$) yields the same results as in [3].

As a further more challenging test, we consider a *quadratically-phased*, cosine-tapered aperture field distribution

$$\mathbf{F}(x, y) = \mathbf{g}(x, y) \exp \left[-ik \frac{(x^2 + y^2)}{2L_f} \right], \quad (33)$$

with $\mathbf{g}(x, y)$ defined in (31), which produces focusing ($L_f > 0$) or defocusing ($L_f < 0$), with L_f representing the conventional focal length. Figure 5a shows the magnitude profile plot of the x component of radiated electric field at the focal plane $z = L_f > 0$ (chosen well within the aperture near-zone), computed via numerical Kirchhoff integration of (6). The wavefield is peaked around the point $(x, y) = (0, 0)$, clearly highlighting the focusing effects. Figures 5b and 5c show representative x - and y -cuts, respectively, where the reference solutions are compared with the narrow-waisted beam synthesis (22). Note that, as compared with the linearly phased example, a smaller beam lattice period with a correspondingly larger number of beams ($(d_A/L)^2 = 2500$) is required for the strongly nonlinear phasing in order to attain comparable accuracy. As shown in Fig. 5d, a value $L = 0.2\lambda$ suffices to keep the maximum error below -30dB . The corresponding results for the z component are shown in Fig. 6, and the same conclusions hold. It should be emphasized that the near-zone focal plane field poses a relatively strong challenge; for moderate observation distances away from the focal plane, a coarser discretization is adequate.

IV. TD FORMULATION

A. TD Narrow-Waisted Beam Discretization: Linear-Delay Aperture Fields

Rigorous extension, into the time domain (TD), of the Gaussian beam discretization for the time-harmonic aperture distribution in Sec. III would require a six-index Gabor series tied to a discretized lattice in the 6D (space-time, wavenumber-frequency) phase space. However, for the narrow-waisted beams in Sec. III-B, which we shall use throughout, implementation of the Gabor algorithm simplifies substantially. In particular, we extend the TD formulation for 2D field radiation by 1D pulsed aperture distributions in [1] to pulsed radiation from 2D aperture distributions. As in [1, Sec. III-A], generalizing (1.28), we first consider space-time separable aperture fields with linear time delay and θ_A, ϕ_A real,

$$\mathbf{f}(x, y, t) = \mathbf{g}(x, y)p[t - c^{-1} \sin \theta_A(x \cos \phi_A + y \sin \phi_A)], \quad (34)$$

and its FD counterpart via (4),

$$\mathbf{F}(x, y, \omega) = P(\omega)\mathbf{g}(x, y) \exp[ik \sin \theta_A(x \cos \phi_A + y \sin \phi_A)]. \quad (35)$$

Here, $p(t)$ is a pulse of short length with respect to the characteristic aperture width d_A , i.e., $T \ll d_A/c$, and $P(\omega)$ is its Fourier transform,

$$P(\omega) = \int_{-\infty}^{\infty} p(t) \exp(i\omega t) dt. \quad (36)$$

The results in this section are related to those in Sec. III through the Fourier transform in (4) filtered by the pulse spectrum $P(\omega)$. When the aperture field is expressed in the Gabor representation based on the CSP beam propagator in (26), the presence of its evanescent spectra motivates use of the one-sided analytic transform [11, p. 222] as in [1],

$$f^+(t) = \frac{1}{\pi} \int_0^{\infty} F(\omega) \exp(-i\omega t) d\omega, \quad \text{Im}(t) \leq 0, \quad (37)$$

where $F(\omega)$ is the conventional Fourier spectrum of the real signal $f(t)$ (see (4)), and the real signal for real t is recovered via

$$f(t) = \text{Re} \left[f^+(t) \right]. \quad (38)$$

Then by FD inversion of (24),

$$\mathbf{e}^+(\mathbf{r}, t) \sim \sum_{m,p} (\mathbf{c}_{mp} \times \mathbf{u}_z) \times \mathbf{b}_{mp}^+(\mathbf{r}, t), \quad (39)$$

with $\mathbf{c}_{mp} = \mathbf{C}_{mp}$ in (25), and the analytic pulsed beam (PB) propagator \mathbf{b}_{mp}^+ expressed via (37) in terms of the paraxial, far-zone FD propagator $\tilde{\mathcal{B}}_{mp}$ in (26) as

$$\mathbf{b}_{mp}^+(\mathbf{r}, t) = \frac{1}{\pi} \int_0^{\infty} \tilde{\mathcal{B}}_{mp}(\mathbf{r}, \omega) P(\omega) \exp(-i\omega t) d\omega, \quad \text{Im}(t) \leq 0. \quad (40)$$

To simplify the evaluation of the integral in (40), it has been customary to choose the complex displacement ib in (26) to be frequency independent so as to yield “isodiffracting”

basis beam wavepackets which remain collimated over a range of frequencies [14]. Via (30), the isodiffracting excitation requires a frequency-dependent beam lattice. However, as noted before (see also [1]), this leads to frequency-dependent Gabor coefficients which cannot be computed by aperture profile sampling when the narrow-waisted constraint is imposed. We therefore choose L to be frequency independent, to obtain the approximate Gabor coefficients

$$\mathbf{c}_{mp} = \mathbf{C}_{mp} \approx (L/\sqrt{2})\mathbf{g}(mL, pL), \quad (41)$$

but at the expense of a frequency-dependent b , which is dealt with below. From here on, we continue by generalizing the 1D aperture algorithm in [1], and consider a class of differentiated Gaussian pulses (cf. (1.37) and (1.38)),

$$p(t) = P_0 p_g^{(j)}(t - T/2, T), \quad (42)$$

$$P(\omega) = P_0 (-i\omega)^j \exp(i\omega T/2) P_g(\omega, T), \quad (43)$$

where P_0 is a normalization constant, the superscript $^{(j)}$ indicates j -th order differentiation with respect to t , and $p_g(t)$, $P_g(\omega)$ denote the standard Gaussian pulse and its spectrum, respectively,

$$p_g(t, T) = \exp \left[- \left(\frac{t}{\sigma T} \right)^2 \right], \quad P_g(\omega, T) = \sqrt{\pi} \sigma T \exp \left(- \frac{\sigma^2 T^2 \omega^2}{4} \right); \quad (44)$$

the variance σ is chosen so that the pulse width of $p(t)$ is $\sim T$.

To de-emphasize the effect of the frequency-dependent parameter b on the complex distance \tilde{R}_{mp} in (28), we assume L/λ (and therefore b) sufficiently small, so that the resulting amplitude factor in the FD beam propagator (26) can be approximated by (see (29) or (1.39))

$$\frac{\mathbf{r} - \tilde{\mathbf{r}}'_{mp}}{\tilde{R}_{mp}} \approx \frac{\mathbf{r} - \mathbf{r}'_{mp}}{R_{mp}}, \quad \mathbf{r}'_{mp} = mL\mathbf{u}_x + pL\mathbf{u}_y, \quad R_{mp} = |\mathbf{r} - \mathbf{r}'_{mp}|, \quad b/L = L(\cos \theta_A)^2/\lambda \ll 1, \quad (45)$$

where the source point \mathbf{r}'_{mp} and the distance R_{mp} are now *real* and *frequency independent*. In the phase, under the same assumptions, we retain the first order paraxial correction (extending (1.40)),

$$\tilde{R}_{mp} \approx z_{bmp} - ib + \frac{\rho_{bmp}^2}{2(z_{bmp} - ib)} \approx z_{bmp} - ib + \frac{\rho_{bmp}^2(z_{bmp} + ib)}{2z_{bmp}^2}, \quad (46)$$

subject to (see (1.42) and (1.43))

$$|z_{bmp} - ib| \gg \rho_{bmp}, \quad b = \frac{(L \cos \theta_A)^2 \omega}{2\pi c} \ll z_{bmp}, \quad \omega \leq \Omega. \quad (47)$$

Here, $\rho_{bmp} = (x_{bmp}^2 + y_{bmp}^2)^{1/2}$, Ω denotes the bandwidth of the pulse spectrum $P(\omega)$, and $(x_{bmp}, y_{bmp}, z_{bmp})$ are the 3D beam coordinates (cf. (1.41))

$$\begin{bmatrix} x_{bmp} \\ y_{bmp} \\ z_{bmp} \end{bmatrix} = \begin{bmatrix} \cos \theta_A \cos \phi_A & \cos \theta_A \sin \phi_A & -\sin \theta_A \\ -\sin \phi_A & \cos \phi_A & 0 \\ \sin \theta_A \cos \phi_A & \sin \theta_A \sin \phi_A & \cos \theta_A \end{bmatrix} \begin{bmatrix} x - mL \\ y - pL \\ z \end{bmatrix}. \quad (48)$$

Applying (45)–(48) to (26), the tilted beam propagator $\tilde{\mathcal{B}}_{mp}$ in (40) can be written concisely as

$$\tilde{\mathcal{B}}_{mp}(\mathbf{r}, \omega) \approx -\Lambda_{mp} i\omega \exp\left(-\frac{\omega^2 T_{mp}^2}{4} + i\omega \tau_{mp}\right) (\mathbf{r} - \mathbf{r}'_{mp}), \quad (49)$$

where

$$\Lambda_{mp}(\mathbf{r}) = -\frac{L}{\sqrt{2\pi c R_{mp}^2}}, \quad (50)$$

$$\tau_{mp}(\mathbf{r}) = c^{-1} \left[z_{bmp} + \frac{\rho_{bmp}^2}{2z_{bmp}} + L \sin \theta_A (m \cos \phi_A + p \sin \phi_A) \right], \quad (51)$$

$$T_{mp}(\mathbf{r}) = \frac{\rho_{bmp} L \cos \theta_A}{\sqrt{\pi c z_{bmp}}}. \quad (52)$$

In contrast with the 1D aperture problem [1], the beam propagator in (49) *does not* contain fractional powers of the frequency; therefore, for the class of pulses in (42)–(44), the PB propagator can be evaluated more easily. In particular, by substituting (43), (44)

and (49) into (40), and recalling basic properties (differentiation, time-shift) of analytic signals, one obtains

$$\mathbf{b}_{mp}(\mathbf{r}, t) = \text{Re} \left[\mathbf{b}_{mp}^+(\mathbf{r}, t) \right] = \frac{P_0 \Lambda_{mp} \sigma T}{\mathcal{T}_{mp}} p_g^{(j+1)} [(t - T/2 - \tau_{mp}) \sigma, \mathcal{T}_{mp}] (\mathbf{r} - \mathbf{r}'_{mp}), \quad (53)$$

$$\mathcal{T}_{mp} = \sqrt{T_{mp}^2 + \sigma^2 T^2}. \quad (54)$$

Thus, the analytic TD narrow-waisted beam expansion in (39) reduces to an efficiently computable closed form in terms of simple analytic functions, with the real field obtained from

$$\mathbf{e}(\mathbf{r}, t) = \text{Re} \left[\mathbf{e}^+(\mathbf{r}, t) \right] \sim \sum_{m,p} (\mathbf{c}_{mp} \times \mathbf{u}_z) \times \mathbf{b}_{mp}(\mathbf{r}, t), \quad (55)$$

Modulated Gaussian pulses can be treated similarly, but require a slightly more complex algebra involving error functions [15].

B. TD Narrow-Waisted Beam Discretization: Nonlinear-Delay Aperture Fields

We now extend the results in Sec. IV-A to the more general aperture field distribution (c.f. (1.52))

$$\mathbf{f}(x, y, t) = \mathbf{s}(x, y) p \left[t - c^{-1} \Phi(x, y) \right], \quad (56)$$

obtained by Fourier inversion of

$$\mathbf{F}(x, y, \omega) = P(\omega) \mathbf{s}(x, y) \exp \left[i \frac{\omega}{c} \Phi(x, y) \right], \quad (57)$$

where $\mathbf{s}(x, y)$ and $\Phi(x, y)$ are real functions. For moderately nonlinear phasing $\Phi(x, y)$, as discussed in Sec. III-C for the FD synthesis, it may be convenient to split the phase (delay) function into a linear part plus a nonlinear remainder

$$\Phi(x, y) = \sin \theta_A (x \cos \phi_A + y \sin \phi_A) + \Phi_{NL}(x, y), \quad (58)$$

where $\Phi_{NL}(x, y)$ does not contain linear terms. Referring to (23), $\mathbf{g}(x, y)$ is now given by

$$\mathbf{g}(x, y) = \mathbf{s}(x, y) \exp[ik\Phi_{NL}(x, y)], \quad (59)$$

with the *complex* Gabor coefficients (see (25)),

$$\mathbf{C}_{mp} = (L/\sqrt{2})\mathbf{s}(mL, pL) \exp[ik\Phi_{NL}(mL, pL)]. \quad (60)$$

The corresponding TD beam expansion differs slightly from (39) (see (1.57)),

$$\mathbf{e}(\mathbf{r}, t) \sim \sum_{m,p} (\mathbf{s}_{mp} \times \mathbf{u}_z) \times \mathbf{b}_{mp}(\mathbf{r}, t - t_{mp}), \quad (61)$$

with the PB propagator \mathbf{b}_{mp} taken from (53), and

$$\mathbf{s}_{mp} = (L/\sqrt{2})\mathbf{s}(mL, pL), \quad t_{mp} = c^{-1}\Phi_{NL}(mL, pL). \quad (62)$$

For strongly nonlinear phasing, the splitting in (58) offers no computational advantages over the *nontilted* version (i.e., $\theta_A = \phi_A = 0$ in (53)) applied to the entire phase function $\Phi(x, y)$.

C. Assessment of Accuracy

The explicit closed form expressions developed in Sec. IV-A for the beam propagator integral in (40) are based on a sequence of approximations which have been expressed in terms of interdependent inequalities that involve all relevant problem parameters and are stated in (45) and (47). To systematize estimates of the range of validity in this multiscale problem, it is useful to define a nondimensional (ND) estimator which embodies all of the problem parameters and is structured around specifically scaled ND parameter combinations containing quantities of special interest.

The most stringent overall constraint in (45) and (47) is the *smallness* of the beam parameter b over the whole bandwidth of interest (the second inequality in (47)), which actually determines the maximum allowable lattice period (i.e., the minimum number of beams) for specified Ω , θ_A , and observation point. For specified observation plane at $z = z_{obs}$, recalling that $z_{bmp} = z_{obs}/\cos\theta_A$, this constraint can be written as

$$\frac{(L \cos \theta_A)^2 \Omega}{2\pi c} \ll \frac{z_{obs}}{\cos \theta_A}. \quad (63)$$

Multiplying and dividing by $d_A^2 T$ on the right hand side, and introducing the auxiliary ND parameters

$$N_b = \left(\frac{d_A}{L} \right)^2 \quad (\text{number of beams in (39)}), \quad (64)$$

$$\kappa = \frac{\Omega T}{2\pi} \quad (\text{normalized bandwidth of the pulse}), \quad (65)$$

$$F_d = \frac{d_A^2}{cT} \quad (\text{Fresnel distance of the aperture}), \quad (66)$$

$$\chi = \frac{z_{obs}}{F_d}, \quad (\text{normalized distance from the aperture}), \quad (67)$$

the constraint in (63) can be written compactly in terms of the ND estimator Q ,

$$Q \equiv \frac{\kappa(\cos \theta_A)^3}{\chi N_b} \ll 1, \quad (68)$$

which highlights the important role of N_b in the algorithm architecture. Actually, except for square apertures, N_b does not represent the precise number of effective beams (i.e., those with nonzero Gabor coefficients) in the expansion (39), and it should be regarded as a conservative estimate. For robustness, it is required that the number of beams, N_b , is large enough to render the narrow-waisted beam synthesis insensitive to “scramblings” of the beam-lattice parameters, i.e., to different combinations of these parameters within the narrow-waisted constraint. With this in mind, one observes from (68), for example, that decreasing the number of beams N_b can be compensated for by a corresponding increase in χ , i.e., in the observation distance z_{obs} when F_d is specified. In principle, the ND estimator Q in (68) can be used for linear as well as non-linear delay aperture field distributions (see Sec. IV-D). However, the actual value of Q ensuring a prescribed accuracy may vary for different distributions. In particular, a finer discretization (i.e., larger N_b) is usually required in the presence of nonlinear phasing/delay, for the same degree of accuracy.

D. Numerical Results

Extensive numerical simulations have been performed in order to validate and calibrate the PB syntheses introduced in Secs. IV-A and IV-B. Here we have selected some rep-

representative results, comparing the beam syntheses with a reference solution obtained via Gaussian quadrature numerical integration [12] of (3). In all simulations, the aperture field distributions are assumed to be x -polarized and therefore (see (3)) the radiated fields contain only x, z components. The assumed temporal excitation is the wideband Rayleigh pulse (four-times differentiated Gaussian) shown in Fig. 7, obtained from (42)-(44) with

$$j = 4, \quad P_0 = T^4/30000, \quad \sigma = 1/\sqrt{50}. \quad (69)$$

The examples presented below represent the TD counterparts of the FD cases discussed in Sec. III-D. In all simulations, we used a pulse-bandwidth value $\Omega T = 40$ (see Fig. 7), which sets the reference level for the ND estimator Q .

We first consider the linear-delay aperture field distribution in (34), with $\mathbf{g}(x, y)$ taken as the x -polarized, cosine-tapered distribution in (31). Figure 8 shows typical results for the x -component of the radiated field for $\theta_A = 30^\circ, \phi_A = 45^\circ$. The problem parameters are listed in the caption. In Fig. 8a, a gray-scale snapshot of the reference field solution is observed in the near-zone of the aperture at a fixed value of z and ct . To highlight details of the PB synthesis in (55), various 1D cuts of the space-time wavefield are displayed in Figs. 8b–8d for a specified beam lattice period, and are compared with the reference solution. Excellent agreement has been obtained with a relatively small number (20×20) of beams (the two solutions are almost indistinguishable on the scale of the plots); the corresponding value of the nondimensional estimator in (68) is $Q \approx 0.1$. Similar results are observed for the z -component in Fig. 9. The convergence of the PB synthesis and the role of the nondimensional estimator Q are illustrated in Fig. 10, which displays as a function of Q the r.m.s. (energy) error $\Delta e_{x,z}$ at fixed observation points \mathbf{r} in the near, intermediate, and far zone of the aperture,

$$\Delta e_{x,z} \equiv \frac{\int_{-\infty}^{\infty} \left| e_{x,z}^{(ref)}(\mathbf{r}, t) - e_{x,z}^{(beam)}(\mathbf{r}, t) \right|^2 dt}{\left[\int_{-\infty}^{\infty} \left| e_{x,z}^{(ref)}(\mathbf{r}, t) \right|^2 dt \int_{-\infty}^{\infty} \left| e_{x,z}^{(beam)}(\mathbf{r}, t) \right|^2 dt \right]^{1/2}}. \quad (70)$$

Each curve in Fig. 10 refers to a fixed observation point, and is parameterized in terms of Q by varying the remaining free parameter in (68), i.e., the number of beams N_b . One observes that the error decreases almost monotonically with decreasing Q (i.e., increasing

N_b). Moreover, the error vs. Q is *weakly* dependent on the observation distance, confirming our previous experience that, for the same degree of accuracy, a coarser discretization can be used as the observation distance increases, subject to (68). Quantitatively, values of $Q \approx 0.1$ (as in Figs. 8 and 9) yield an error < -20 dB, which is usually acceptable.

As a final example, we consider an x -polarized, cosine tapered, aperture distribution with quadratic delay

$$\mathbf{f}(x, y, t) = \mathbf{g}(x, y)p \left[t + (x^2 + y^2)/(2cL_f) \right], \quad (71)$$

where $\mathbf{g}(x, y)$ is defined in (31). The above distribution represents the TD counterpart of the FD focusing/defocusing distribution in (33), and constitutes a particularly stringent test for the nonlinear-delay aperture PB synthesis in (61). Figure 11 contains representative results for the x -component of the radiated field, observed in the critical focal plane $z = L_f$ which has been chosen at moderate distance from the aperture, $L_f = 0.2F_d$. In particular, Fig. 11a shows a snapshot of the reference solution at the time instant $ct = 20.44$, where the focusing effects are clearly evident. The PB synthesis in (61), for a fixed value of the beam lattice period, is compared with the reference solution in a number of representative 1D cuts shown in Figs. 11b–11d, with excellent agreement. Similar observations apply to the z -component of the radiated field in Fig. 12. Note that standard (non-uniform) ray asymptotics would fail due to the presence of caustic transition regions. As in the FD case (cf. Sec. III-D), the narrow-waisted pulsed beams still yield an accurate wavefield synthesis, with an increase (in this example of a factor 16) in the computational effort, as compared with the linear-delay distributions. In this example, we used 80×80 beams, yielding $Q \approx 0.005$. The r.m.s. errors in (70) vs. Q are shown in Fig. 13 for various observation points. It is observed, for both components, that away from the focal plane $z = L_f$, the error and its numerical values behave similarly to those for the linear-delay case in Fig. 10. In particular, the error behavior is weakly dependent on the observation point. In the more critical focal plane, the error behavior is quite different: it is uniformly larger and reaches a plateau (< -20 dB) for $Q \lesssim 0.005$, indicating an intrinsic lower bound independent of the discretization. Such behavior was not observed in the FD, where the *full* CSP GB propagators are used, and therefore is probably related to the

paraxial approximation in (46), whose implications are not accounted for in the ND estimator Q (derived from (47)). For the moderate focusing in this example, this lower bound still yields reasonably accurate syntheses in the focal plane (see Figs. 11 and 12), but for considerably stronger focusing, decreased accuracy can be expected and has actually been encountered in numerical experiments. For these near-zone focal-plane field observations, our numerical simulations usually produced satisfactory accuracy ($\Delta e_{x,z} < -15\text{dB}$) for focused distributions with $L_f > 0.1F_d$. As seen from Fig. 13, reasonable accuracy can be achieved away from the focal plane within the $Q \lesssim 0.1$ constraint observed in the linear-delay-case.

The above beam syntheses were obtained with computing times ranging from 2ms (linearly-phased aperture with 400 beams in Figs. 8 and 9) to 32ms (focused aperture with 6400 beams in Figs. 11 and 12) per space-time sample on a 700 MHz PC. The reference solution was obtained via numerical quadrature of the Kirchhoff integration in (3), using a 7-point Gaussian formula [12] with discretization steps $\Delta_x = \Delta_y = d_A/350 = cT/35$, resulting in computing times of about 600 ms per space-time field sample.

V. CONCLUSIONS

Referring for background to the abstract and introduction (Sec. I), we have developed here a discretized Gabor-based narrow-waisted pulsed beam (PB) algorithm for synthesis of 3D vector fields radiated by truncated planar 2D aperture distributions with amplitude-tapered coordinate-separable linearly and nonlinearly phased profile functions. Building upon the frequency domain 2D aperture radiation analysis in [2], [3] and the time domain 1D aperture radiation study in [1], we have obtained the simple, readily computable, robust, explicit asymptotic PB expansions in (55) and (61). Calibration against reference solutions for a variety of numerical test configurations have established the accuracy and range of validity of the PB algorithms. The validity constraints have been formalized through the nondimensional estimator in (68), which combines all relevant problem parameters in terms of physically meaningful spatially and temporally scaled groupings. Computationally, the proposed PB syntheses appear attractive when compared with numerical Kirchhoff integrations, with typical computing times from about 20 to 300 times smaller.

Our next goal is the interaction of the radiated PBs with complex propagation environments. In this connection, forward and inverse scattering applications to 1D aperture/2D field configurations have been addressed in [16]–[18]. Extensions to 2D aperture/3D field configurations, by combining the PB discretization algorithm in the present paper with Kirchhoff Physical Optics techniques, are currently under consideration.

REFERENCES

- [1] V. Galdi, L.B. Felsen, and D.A. Castañón, “Narrow-waisted Gaussian beam discretization for two-dimensional time-dependent radiation from large apertures,” *IEEE Trans. Antennas Propagat.*, vol. 49, No. 9, pp. 1322-1332, Sept. 2001.
- [2] J.J. Maciel and L.B. Felsen, “Discretized Gabor-based beam algorithm for time-harmonic radiation from two-dimensional truncated planar aperture distributions - I: Formulation and solution,” *IEEE Trans. Antennas Propagat.*, 2002 (in print).
- [3] J.J. Maciel and L.B. Felsen, “Discretized Gabor-based beam algorithm for time-harmonic radiation from two-dimensional truncated planar aperture distributions - II: Asymptotics and numerical tests,” *IEEE Trans. Antennas Propagat.*, 2002 (in print).
- [4] J.J. Maciel and L.B. Felsen, “Systematic study of fields due to extended apertures by Gaussian beam discretization,” *IEEE Trans. Antennas Propagat.*, vol. 37, No. 7, pp. 884-892, July 1989.
- [5] J.J. Maciel and L.B. Felsen, “Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part I - plane layer,” *IEEE Trans. Antennas Propagat.*, vol. 38, No. 10, pp. 1607-1617, Oct. 1990.
- [6] J.J. Maciel and L.B. Felsen, “Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part II - circular cylindrical layer,” *IEEE Trans. Antennas Propagat.*, vol. 38, No. 10, pp. 1618-1624, Oct. 1990.
- [7] L.B. Felsen and V. Galdi, “Complex-source-point narrow-waisted ray-like Gaussian beams for frequency and time domain radiation and scattering,” in *Ultra-Wideband, Short Pulse Electromagnetics 5*. New York, NY: Kluwer/Plenum Publishers (S.R. Cloude and P.D. Smith, Eds.), 2001 (in print).
- [8] V. Galdi, L.B. Felsen, and D.A. Castañón, “Quasi-ray Gaussian beam algorithm for time-harmonic scattering by moderately rough interfaces,” *IEEE Trans. Antennas Propagat.*, vol. 49, No. 9, pp. 1305-1314, Sept. 2001.
- [9] V. Galdi, D.A. Castañón, and L.B. Felsen, “Multifrequency reconstruction of moderately rough interfaces via quasi-ray Gaussian beams,” *IEEE Trans. Geosci. Remote Sensing*, 2002 (in print).
- [10] V. Galdi, H. Feng, D.A. Castañón, W.C. Karl, and L.B. Felsen, “Multifrequency subsurface sensing in the presence of a moderately rough air-soil interface via quasi-ray Gaussian beams,” invited paper in *Radio Science*, Special Issue on 2001 URSI EMT Symposium, submitted Nov. 2001.
- [11] T.B. Hansen and A.D. Yaghjian, *Plane-Wave Theory of Time-Domain Fields: Near-Field Scanning Applications*, Piscataway (NJ): IEEE Press, 1999.
- [12] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge (UK): Cambridge Univ. Press, 1992.
- [13] L. Wang, C.T. Chen, and W.C. Lin, “An efficient algorithm to compute the complete set of discrete Gabor coefficients,” *IEEE Trans. Image Proc.*, vol. 3, No. 1, pp. 87-92, Jan. 1994.

- [14] E. Heyman and L.B. Felsen, "Gaussian beam and pulsed beam dynamics: complex source and spectrum formulation within and beyond paraxial asymptotics," *J. Opt. Soc. Am. A*, Feature Issue on Free and Guided Optical Beams, vol. 18, No. 7, pp. 1588-1611, July 2001.
- [15] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*. New York (NY): Dover, 1964.
- [16] V. Galdi, L.B. Felsen, and D.A. Castañon, "Quasi-ray Gaussian beam algorithm for short-pulse two-dimensional scattering by moderately rough dielectric interfaces," submitted to *IEEE Trans. Antennas Propagat.*, Apr. 2001.
- [17] V. Galdi, J. Pavlovich, W.C. Karl, D.A. Castañon, and L. B. Felsen, "Moderately rough dielectric interface reconstruction via short-pulse quasi-ray Gaussian beams," submitted to *IEEE Trans. Antennas Propagat.*, Oct. 2001.
- [18] V. Galdi, H. Feng, D. A. Castañon, W.C. Karl, and L. B. Felsen, "Moderately rough surface underground imaging via short-pulse quasi-ray Gaussian beams," submitted to *IEEE Trans. Antennas Propagat.*, Oct. 2001.

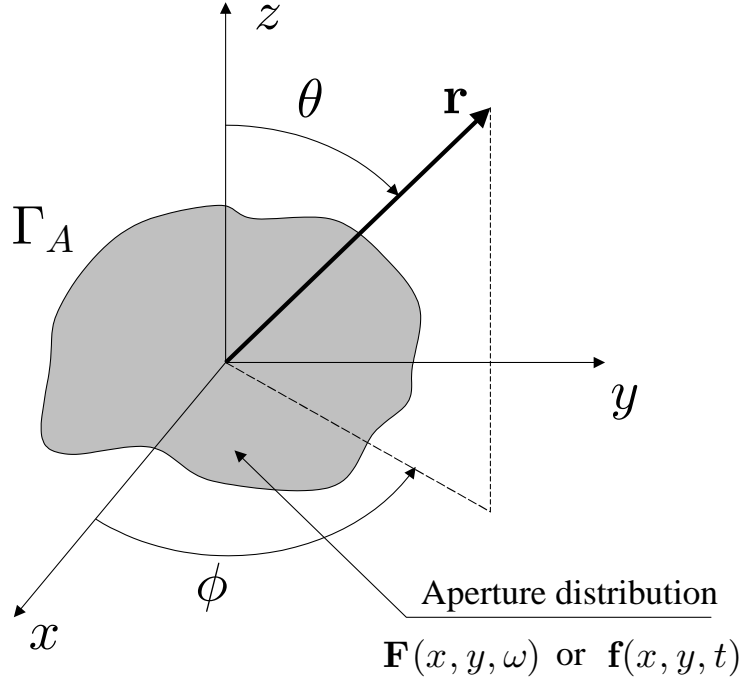


Fig. 1. Two-dimensional aperture field distribution and coordinate systems.

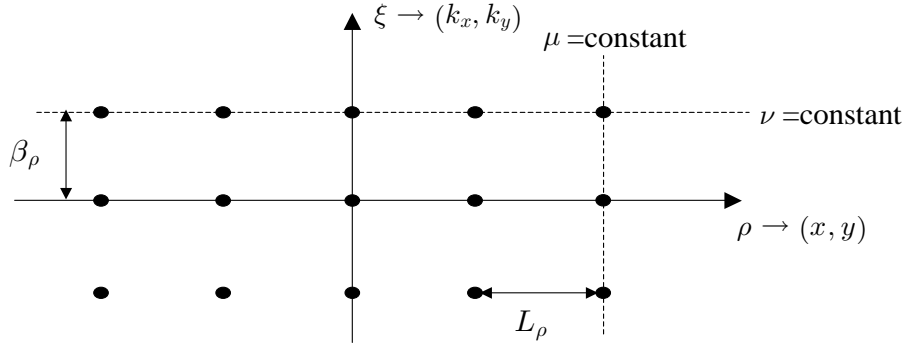


Fig. 2. 2D projection of the 4D $(x, y; k_x, k_y)$ discretized Gabor lattice onto 2D (ρ, ξ) subspaces, where ρ stands for either x or y and ξ stands for either k_x or k_y , with μ and ν denoting the corresponding spatial and spectral integer indexes (m, p) or (n, q) , respectively. $\nu = \text{const.}$: spatially displaced beams centered at $\rho_\mu = \mu L_\rho$, with fixed phase gradient (beam tilt) $\nu \beta_\rho$; $\mu = \text{const.}$: spectrally displaced beam tilts $(\nu \beta_\rho)$ at fixed beam location ρ_μ .

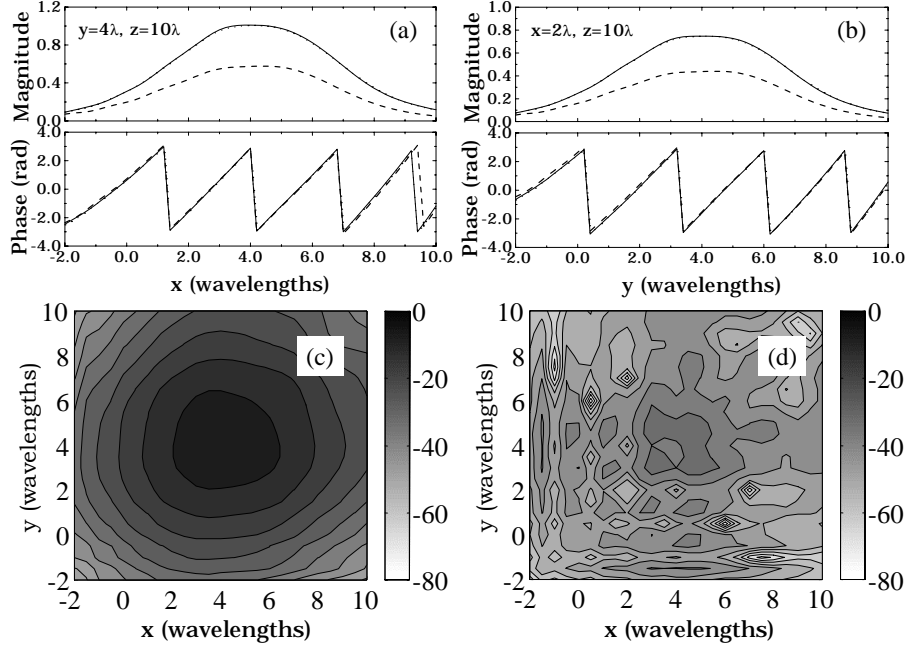
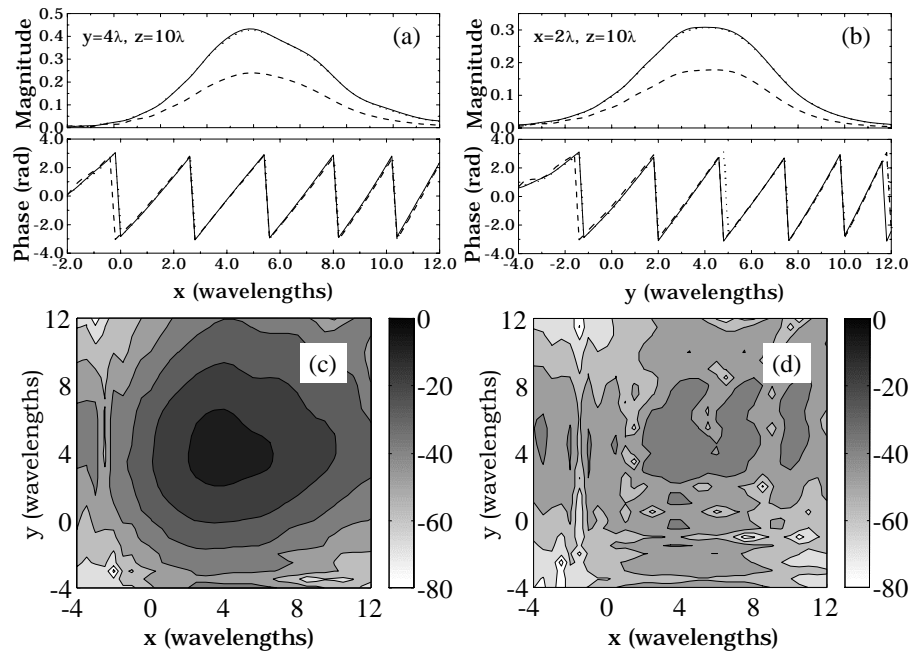


Fig. 3. x -polarized, linearly-phased, square cosine aperture distribution in (23) with (31) ($d_A = 10\lambda$, $\theta_A = 30^\circ$, $\phi_A = 45^\circ$): **x -component** of near-zone ($z = 10\lambda$) radiated field synthesis using narrow-waisted nontilted and tilted beams with a fixed beam lattice period ($L = 0.8\lambda$) is compared with the reference solution (numerical integration of (6)). The number of beams used is $(d_A/L)^2 \sim 150$. (a): Cut at $y = 4\lambda$; (b): Cut at $x = 2\lambda$; — Reference solution; - - - Nontilted beam synthesis; Tilted beam synthesis. The reference solution and tilted beam synthesis coincide on the scale of plots. (c), (d): Gray-scale plots of the relative error (32) (in dB) for nontilted and tilted beam synthesis, respectively.

Fig. 4. As in Fig. 3, but **z-component**.

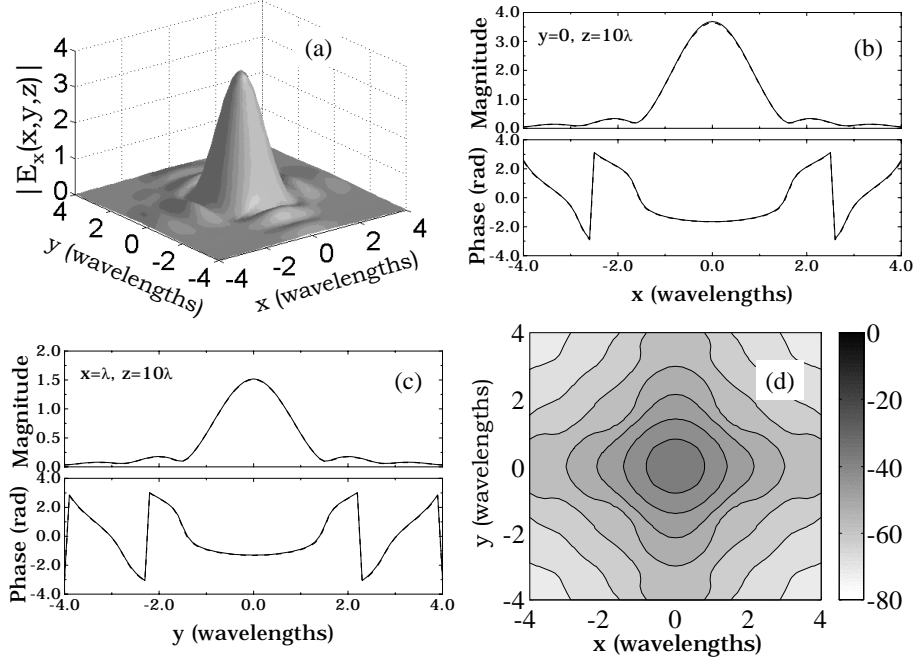
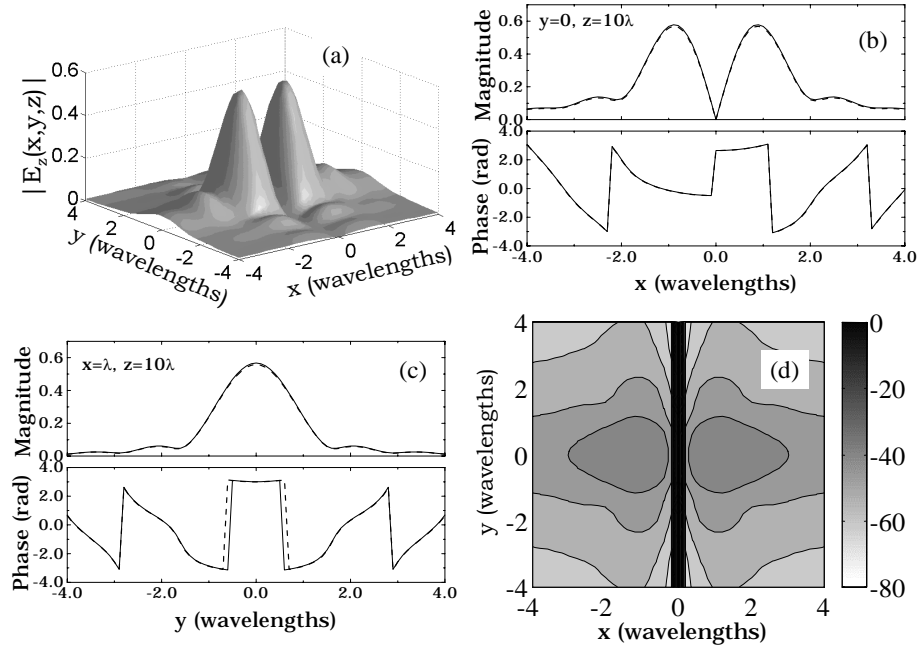
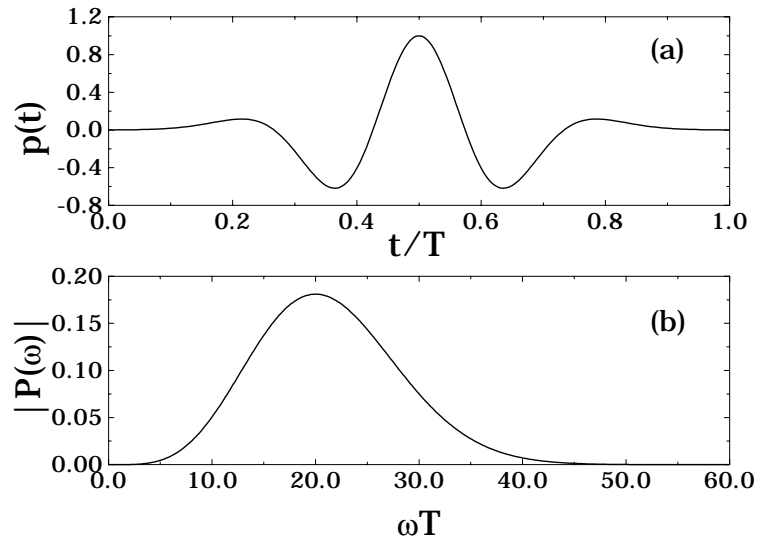


Fig. 5. x -polarized, focused, square cosine aperture distribution in (33) ($d_A = 10\lambda, L_f = 10\lambda$): **x-component** of radiated field at the focal plane $z = L_f$. (a): Magnitude profile plot of the reference solution (numerical integration of (6)); (b),(c): Cuts at $y = 0$ and $x = \lambda$, respectively, comparing the reference solution with the narrow-waisted beam synthesis ((22) with $L = 0.2\lambda$). (d): Gray-scale plot of the relative error (32) (in dB) for the beam synthesis ($L = 0.2\lambda$). The number of beams used is $(d_A/L)^2 \sim 2500$. — Reference solution; - - - Beam synthesis.

Fig. 6. As in Fig. 5, but **z-component**.Fig. 7. Rayleigh pulse. (a): Temporal profile in (42). (b): Spectrum (magnitude) in (43) ($j = 4$, $P_0 = T^4/30000$, $\sigma = 1/\sqrt{50}$).

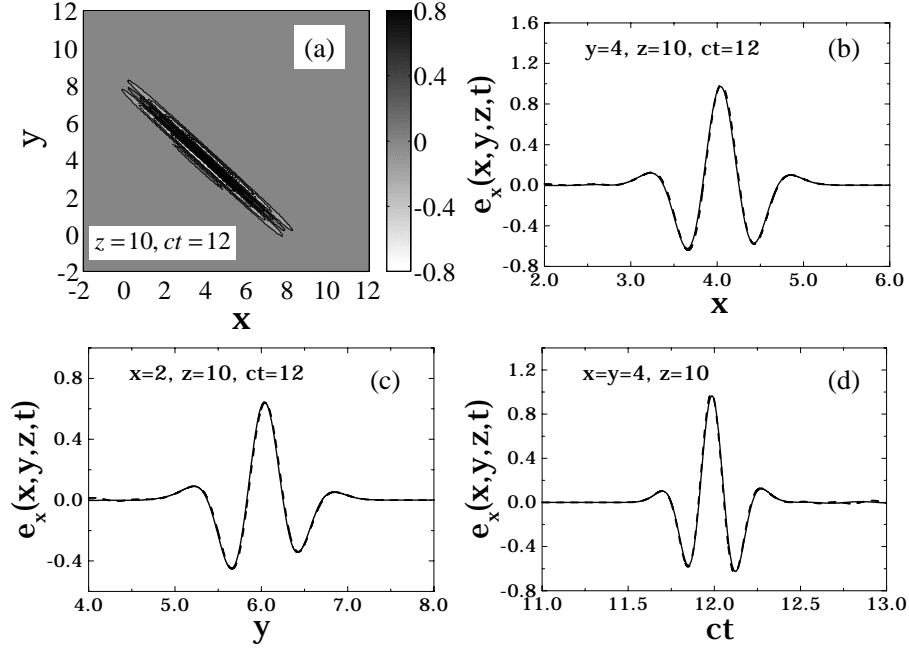


Fig. 8. x -polarized, pulsed, linear-delay, square cosine aperture distribution in (34) with (31), $cT = 1$, $d_A = 10 = 10cT$ (arbitrary units), $\theta_A = 30^\circ$, $\phi_A = 45^\circ$: **x-component** of radiated field at $z = 10 = 0.1F_d$. (a): Gray-scale snapshot of the reference solution (numerical integration of (3)) at $ct = 12$; (b),(c),(d): Cuts at ($y = 4, ct = 12$), ($x = 2, ct = 12$), and ($x = y = 4$), respectively, comparing the reference solution with the PB beam synthesis (55) ($L = d_A/20$, i.e., 400 beams). For the above parameters, $Q \approx 0.1$. — Reference solution; - - - Beam synthesis.

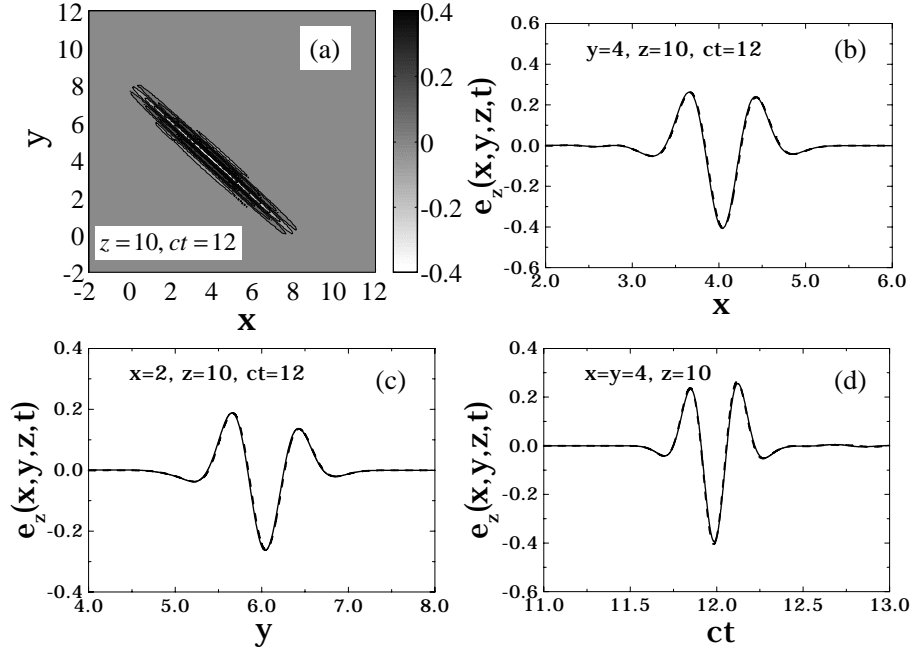
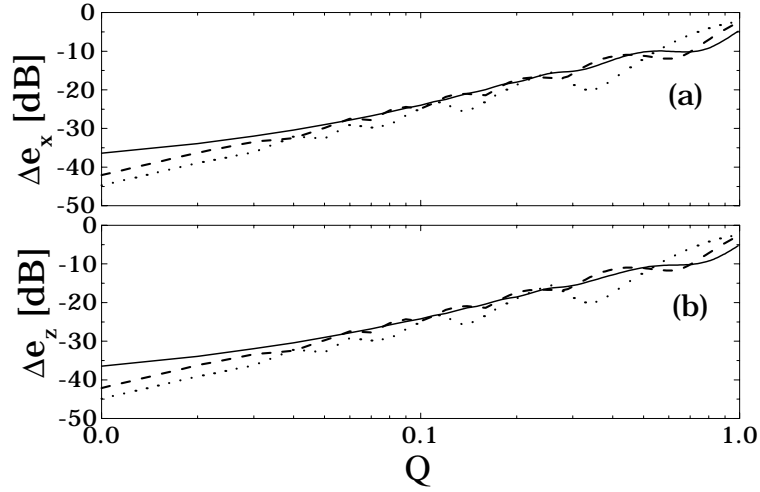
Fig. 9. As in Fig. 8, but **z-component**.

Fig. 10. r.m.s error in (70) for x - and z -components, respectively, as a function of the nondimensional estimator Q in (68), adjusted by varying N_b . Parameters as in Figs. 8 and 9. Observation points (on the beam axis (θ_A, ϕ_A)): — $x = y = 4, z = 10 = 0.1F_d$; - - - $x = y = 16, z = 40 = 0.4F_d$; ····· $x = y = 40, z = 100 = F_d$.

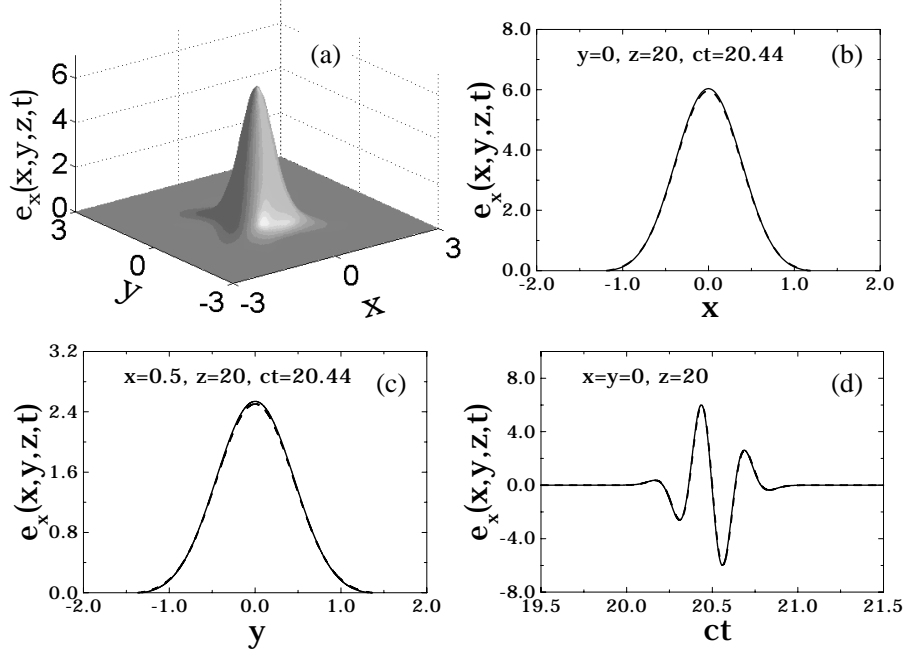


Fig. 11. x -polarized, pulsed, focused, square cosine aperture distribution in (71), with $cT = 1, d_A = 10 = 10cT, L_f = 20 = 0.2F_d$ (arbitrary units). **x-component** of radiated field at the focal plane $z = L_f$. (a): Snapshot of the reference solution (numerical integration of (3)) at $ct = 20.44$; (b),(c),(d): Cuts at $(y = 0, ct = 20.44)$, $(x = 0.5, ct = 20.44)$, and $(x = y = 0)$, respectively, comparing the reference solution with the PB beam synthesis (61) ($L = d_A/80$, i.e., 6400 beams). For the above parameters, $Q \approx 0.05$. — Reference solution; - - - Beam synthesis.

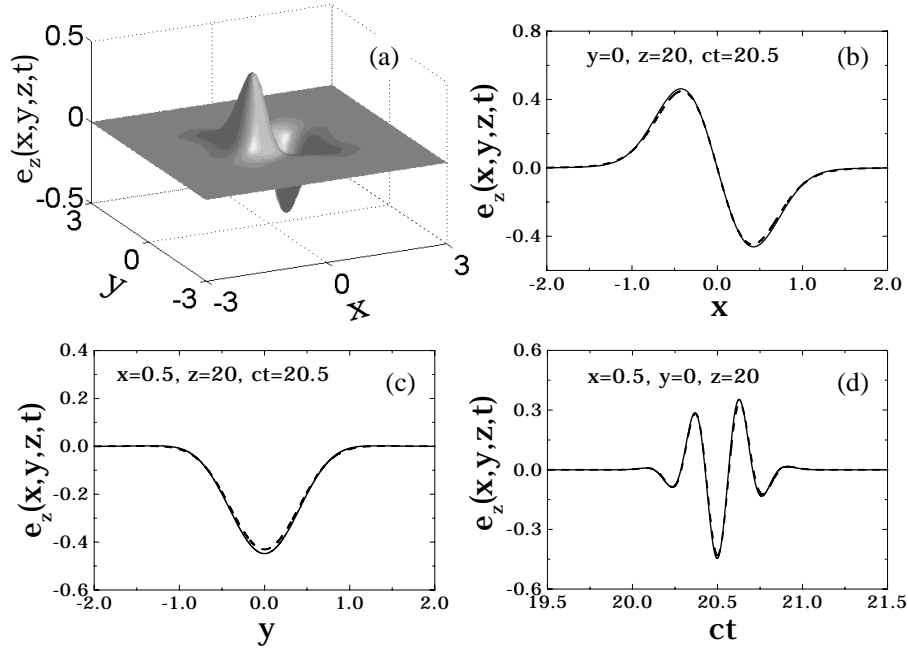


Fig. 12. Parameters as in Fig. 11: **z-component**. (a): Snapshot of the reference solution at $ct = 20.5$; (b),(c),(d): Cuts at $(y = 0, ct = 20.5)$, $(x = 0.5, ct = 20.5)$, and $(x = 0.5, y = 0)$, respectively, comparing the reference solution with the PB beam synthesis (61) ($L = d_A/80$, i.e., 6400 beams, $Q \approx 0.05$). — Reference solution; - - - Beam synthesis.

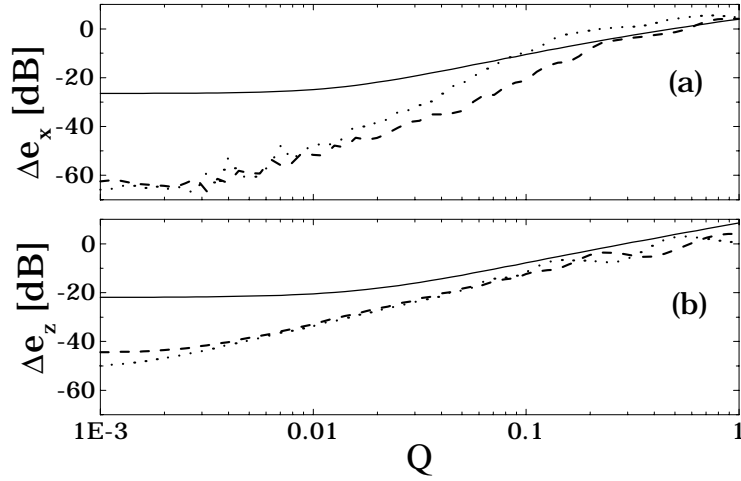


Fig. 13. r.m.s error in (70) for x - and z -components, respectively, as a function of the nondimensional estimator Q in (68), adjusted by varying N_b . Parameters as in Figs. 11 and 12: (a): x -component. Observation points: — $x = y = 0, z = 20 = L_f = 0.2F_d$; - - - $x = y = 0, z = 70 = 0.7F_d$; ····· $x = y = 0, z = 100 = F_d$. (b): z -component. Observation points: — $x = 0.5, y = 0, z = 20 = L_f$; - - - $x = 3.7, y = 0, z = 70$; ····· $x = 4.7, y = 0, z = 100$.

Quasi-Ray Gaussian Beam Algorithm for Short-Pulse Two-Dimensional Scattering by Moderately Rough Dielectric Interfaces

Vincenzo Galdi, *Member, IEEE*, Leopold B. Felsen, *Life Fellow, IEEE*,
and David A. Castañon, *Senior Member, IEEE*

This work was supported by ODDR&E under MURI Grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from Grant No. 9900448 by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University, Brooklyn, NY USA.

V. Galdi is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, on leave from University of Sannio, Benevento, Italy (e-mail: vgaldi@bu.edu).

L.B. Felsen is with the Department of Aerospace and Mechanical Engineering and the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, part-time. He is also University Professor Emeritus, Polytechnic University, Brooklyn, NY 11201 USA (e-mail: lfelsen@bu.edu).

D.A. Castañon is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA (e-mail: dac@bu.edu).

Abstract

We consider short-pulse (SP) time domain (TD) two-dimensional (2D) scattering by moderately rough dielectric interfaces excited by one-dimensional SP-TD aperture field distributions. This study extends to the SP-TD our previous investigation of time-harmonic *high frequency* 2D-scattering of Gabor-based quasi-ray Gaussian beam fields excited by 1D aperture field distributions in the presence of moderately rough dielectric interfaces [1]. The proposed approach is based on the Kirchhoff Physical Optics (PO) approximation in conjunction with the Gabor-based quasi-ray narrow-waisted Gaussian pulsed-beam (PB) discretization [2], which is applied to the equivalent induced SP “surface currents” on the interface that establish the TD reflected/transmitted fields. We show that, for well-collimated truncated SP incident fields, the PO-PB synthesis of the reflected/transmitted fields yields an approximate *explicit* physically appealing, numerically efficient asymptotic algorithm, with well-defined domains of validity based on the problem parameters. An extensive series of numerical experiments verifies the accuracy of our method by comparison with a rigorously based numerical reference solution. The algorithm is intended for use as a rapid forward solver in SP-TD inverse scattering and imaging scenarios in the presence of moderately rough dielectric interfaces.

Keywords

Rough surface scattering, Gaussian beams, short pulses.

I. INTRODUCTION

This paper is concerned with the short-pulse (SP) time domain (TD) scattering of two-dimensional (2D) fields by 1D moderately rough dielectric interfaces which separate two semi-infinite air and dielectric regions, respectively, and are excited by incident fields from SP planar 1D truncated aperture distributions. The aperture emits a well-collimated SP which illuminates a portion of the interface and induces there local “equivalent SP surface currents” which produce the TD reflected and transmitted fields. Restricting the analysis to the high frequency range, the equivalent surface currents are approximated by their local Physical Optics (PO) values, which renders the reflection and transmission problem equivalent to radiation from a locally varying time-delayed SP aperture field distribution along the interface profile. This SP-PO current distribution is parameterized in terms of Gabor-based narrow-waisted ray-like discretized Gaussian initial fields on the interface which generate reflected/transmitted pulsed beam (PB) propagators. The field at the observer in either half space is established by summation over the SP-PO basis beams.

The problem strategy outlined above implements a further step in our planned sequential approach toward a robust physically incisive, accurate and numerically efficient high frequency asymptotic algorithm that quantifies SP scattering of 3D vector fields by 2D moderately rough interface profiles between different dielectric media (which may be lossy). The intended application is as a rapid forward solver in SP-TD inverse scattering and imaging scenarios where a dielectric rough interface plays a critical role (our approach differs from several numerical/statistical approaches that have been explored in the past (see, e.g., [3]–[5] and the references therein)). The general problem of SP-TD fields excited by 1D planar aperture distributions parameterized in terms of SP Gaussian beams (GB) has already been addressed in one of our previous investigations [2]. The new step in the present problem is the inversion to the SP-TD of the scattered fields in our previous study that deals with the same complete 2D problem geometry but excited by time-harmonic high frequency illumination; here the discretized Gabor basis for the 1D planar truncated aperture distribution involves ray-like frequency domain GB propagators [1]. The coordination of [2] with the SP-TD inversion of the reflected/transmitted fields in [1] occupies the remainder of this paper. After the problem statement in Sec. II, we re-examine in Sec. III the high frequency solution in [1], and modify it by introducing the PO approximation for the induced surface fields on the interface in order to facilitate *explicit* analytic inversion to the SP-TD later on; the corresponding PO-modified asymptotic Gabor beam propagators excited by illumination from a quasi-linearly-phased truncated aperture distribution are developed here. The inversion to the SP-TD is treated in Sec. IV, with preliminary frequency domain approximations for slightly lossy dielectrics as well as certain basis beam parameters that adapt the inversion to the SP-TD aperture field parameterization in [2]. The numerical experiments in Sec. V, performed for a variety of different problem parameters, assess and quantify the accuracy and range of validity of the explicit, compact, physics-based computationally efficient algorithm in Sec. IV through comparison with a rigorously-based numerical reference solution. Conclusions are briefly stated in Sec. VI.

II. STATEMENT OF THE PROBLEM

The problem geometry is sketched in Fig. 1. All fields and geometries are two-dimensional (2D) in the y -independent (x, z) space. A transverse magnetic (TM) po-

larized pulsed electromagnetic (EM) field is assumed to impinge from free space onto a homogeneous lossy dielectric half-space bounded by an irregular interface described by a continuous function $h(x)$ (Fig. 1a).

A. Incident Field

We restrict our attention to the y -directed electric field $e(x, z, t)$ from which all other field quantities of interest can be computed via Maxwell's equations. The y -directed incident field e^i is generated by a large truncated 1D aperture of width d at $z = z_A$, with an assigned space-time electric field distribution $f(x, t)$,

$$e^i(x, z = z_A, t) = \begin{cases} f(x, t), & |x - x_A| \leq d/2, \\ 0, & |x - x_A| > d/2, \end{cases} \quad (1)$$

where x_A is a spatial shift which can be used to adjust the illumination. In what follows, we shall be concerned with pulsed well-collimated Gaussian beams (GBs) generated by the separable linear-delay space-time field distributions

$$f(x, t) = g[(x - x_A) \cos \theta_A] p[t - c^{-1}(x - x_A) \sin \theta_A]. \quad (2)$$

In (2), $g(x)$ is a Gaussian taper function and $c = (\epsilon_0 \mu_0)^{-1/2}$ is the free-space wavespeed, with ϵ_0, μ_0 denoting the free-space permittivity and permeability, respectively. Moreover, $p(t)$ is a short pulse of length $T \ll d/c$, and θ_A denotes the tilt angle of the radiated beam relative to the z -axis (Fig. 1b). The variance of $g(x)$ and the spatial shift x_A are chosen so that $f(x, t)$ tapers to zero for $|x| > d/2$, as shown in Fig. 1b. The linear-delay aperture field distribution in (2) generates a *wide-waisted* tilted GB which, in the collimation zone of the aperture, can be approximated by a pulsed tapered plane wave,

$$e^i(x, z, t) \sim g(x_B) p(t - c^{-1} z_B), \quad (3)$$

where (x_B, z_B) are beam centered coordinates (Fig. 1b)

$$\begin{bmatrix} x_B \\ z_B \end{bmatrix} = \begin{bmatrix} \cos \theta_A & \sin \theta_A \\ \sin \theta_A & -\cos \theta_A \end{bmatrix} \begin{bmatrix} x - x_A \\ z - z_A \end{bmatrix}. \quad (4)$$

The reference solutions employed later on are based on the incident field generated by numerical evaluation of the rigorous Kirchhoff aperture (line source Green's function) integration using (1) and (2) (with (10)), whereas all subsequent beam-derived results rely on the plane wave approximation in (3).

B. Reflected and Transmitted Fields

The irradiated half-space is assumed to be nonmagnetic, i.e., with relative permeability $\mu_r = 1$, and characterized by the constant relative permittivity ϵ_r and electric conductivity σ , both being frequency independent. The interface profile is assumed to be *moderately* rough (both in height and slope) with respect to the pulse length T , i.e., the undulations in $h(x)$ are assumed to be on the order of cT . The reflected and transmitted fields are modeled via the Gabor-based narrow-waisted discretized pulsed beam (PB) algorithm in [2], to which we shall refer frequently throughout the paper. Notationally, we shall use $(\alpha.\beta)$ to denote eq. (β) in ref. $[\alpha]$; for instance, (1.6) means eq. (6) in ref. [1].

III. FREQUENCY DOMAIN FORMULATION

A Gabor-based, narrow-waisted beam approach for time-harmonic scattering by, and transmission through, moderately rough interfaces has been introduced in [1], [6], extending the results in [7], [8], and yielding a robust and efficient numerical algorithm. Although extension of this approach to *pulsed* excitation, via analytic Fourier inversion, is possible in principle, its implementation is indirect and cumbersome. In particular, multiple interactions, which can be treated conveniently in the frequency domain (FD) (see Appendix A in [1]), would involve substantial preliminary algebra in the time domain (TD). Accordingly, we pursue here a simplified approach based on the FD Kirchhoff Physical Optics (PO) approximation which, via the algorithm in [2], can be more easily inverted to the TD. In this section, we briefly review the FD PO formulation and its Gabor-based narrow-waisted beam discretization. Capital letters identify FD field quantities.

A. Physical Optics Approximation

The Kirchhoff PO approximation has been used extensively for scattering from smoothly curved structures which are large on the wavelength scale [9]. Although most

applications have been carried out for impenetrable structures, penetrable objects can be handled as well (see, e.g., [10]). Applications to *conducting* rough surface scattering have also received much attention (see, e.g., [11]–[14]). For *penetrable* rough interfaces, we return to the geometry in Fig. 1a, subject to the pulsed tapered plane-wave illumination in (3), with FD spectrum

$$E^i(x, z, \omega) \sim g(x_B)P(\omega) \exp(ik_0 z_B). \quad (5)$$

Here, (x_B, z_B) are the beam coordinates defined in (4), $k_0 = \omega/c = 2\pi/\lambda_0$ is the free-space wavenumber, λ_0 is the free-space wavelength, and $P(\omega)$ is the spectrum of the pulse $p(t)$,

$$P(\omega) = \int_{-\infty}^{\infty} p(t) \exp(i\omega t) dt. \quad (6)$$

The y -directed FD PO reflected field in the half-space $z > h(x)$ can be expressed as (see, e.g., [15])

$$E^r(x, z, \omega) \sim - \int_{C_{PO}} J_{PO}^r(x', \omega) \frac{\partial}{\partial \zeta} G_{2D}(x, z; x', h(x'); k_0) d\ell', \quad (7)$$

where C_{PO} extends over the illuminated portion of the 1D surface $z = h(x)$, $d\ell'$ is the incremental arc-length measured along the surface tangent,

$$d\ell' = \sqrt{1 + \left(\frac{dh}{dx}(x')\right)^2} dx', \quad \frac{dh}{dx}(x') \equiv \frac{dh}{dx}(x) \Big|_{x'}, \quad (8)$$

and $\partial/\partial\zeta$ denotes the normal derivative (Fig. 2)

$$\frac{\partial}{\partial \zeta} \equiv \left[1 + \left(\frac{dh}{dx}(x')\right)^2 \right]^{-1/2} \left[\frac{\partial}{\partial z} - \frac{dh}{dx}(x') \frac{\partial}{\partial x} \right]. \quad (9)$$

As stated previously, the incident field tapering is chosen so that the illuminated portion of the interface, C_{PO} , is essentially confined to the interval $|x| \leq d/2$ (see Fig. 1b). In (7), G_{2D} is the FD line-source Green's function,

$$G_{2D}(x, z; x', z'; k) = \frac{i}{4} H_0^{(1)}(kR), \quad R = [(x - x')^2 + (z - z')^2]^{1/2}, \quad (10)$$

with $H_0^{(1)}(\cdot)$ denoting the zeroth order Hankel function of the first kind. The induced PO “surface current” density J_{PO}^r is given by twice the tangential reflected field at the interface, obtained from the canonical solution of infinite plane-wave scattering by a plane dielectric boundary locally tangent to the rough surface profile (Fig. 2)

$$J_{PO}^r(x, \omega) = 2\mathcal{R}(x, \omega)E^i(x, h(x), \omega). \quad (11)$$

In (11), \mathcal{R} denotes the local TM plane-wave Fresnel reflection coefficient,

$$\mathcal{R}(x, \omega) = \frac{\cos \theta^i - (\epsilon_r^e - \sin^2 \theta^i)^{1/2}}{\cos \theta^i + (\epsilon_r^e - \sin^2 \theta^i)^{1/2}}, \quad (12)$$

where

$$\epsilon_r^e(\omega) = \epsilon_r + i \frac{\sigma}{\omega \epsilon_0} \quad (13)$$

is the *effective* complex relative permittivity, and θ^i is the local incidence angle relative to the normal, which can be expressed in terms of the incident beam tilt angle θ_A and the local slope α as (see Fig. 2)

$$\theta^i(x) = \theta_A - \alpha(x), \quad \alpha(x) = \tan^{-1} \left[\frac{dh}{dx}(x) \right]. \quad (14)$$

The same considerations applied to the field transmitted into the half-space $z < h(x)$ yield

$$E^t(x, z, \omega) \sim \int_{C_{PO}} J_{PO}^t(x', \omega) \frac{\partial}{\partial \zeta} G_{2D}(x, z; x', h(x'); k) d\ell', \quad (15)$$

where $k = \sqrt{\epsilon_r^e} k_0$ and the PO surface current density J_{PO}^t is given by

$$J_{PO}^t(x, \omega) = 2[1 + \mathcal{R}(x, \omega)]E^i(x, h(x), \omega). \quad (16)$$

The limitations of the Kirchhoff PO approximation have been thoroughly investigated in the past and are well documented in the technical literature (for the case of conducting rough surface scattering see, e.g., [11]). In general, this approximation works well for large, smooth scatterers, and for observation directions not far from backscatter. The

formulations in (7) and (15) neglect multiple interactions, which can be incorporated in principle through higher-order models (see, e.g., [14]). Here, we restrict the range of validity to moderate roughness (both in height and slope) and incidence directions far from grazing, thereby avoiding multiple scattering.

B. Gabor-Based Narrow-Waisted Gaussian Beam Discretization

The FD PO integrals in (7), (15) are formally analogous to the FD Kirchhoff aperture radiation integrals in [2, Sec. II] (see (2.2)). The only difference is that the line integration in (7), (15) is performed along the 1D rough surface profile $z = h(x)$ instead of a 1D planar aperture as in (2.2). In [2], the assigned FD planar aperture field distribution is parameterized in terms of x -domain discretized m -indexed Gabor basis functions with narrow width L , centered on the Gabor lattice points $x_m = mL$ on the aperture; these initial conditions generate *narrow-waisted*, quasi-ray Gaussian beams (GB) which can be approximated efficiently in terms of *complex source point* (CSP) propagators (see (2.14)). *Nontilted* beams, launched from the Gabor lattice points and propagating along the direction *normal* to the aperture plane, are superposed to synthesize the radiated field (see (2.13)), with the Gabor coefficients approximated efficiently by *sampling* the aperture field distribution (see (2.12)). All other *tilted* beams in the full Gabor GB expansion (2.8), which are evanescent for $L \lesssim \lambda_0 \ll d$, are neglected. For linearly phased apertures (cf. (2.17)), a more efficient parameterization is obtained by exploiting propagation-matched tilted beams (cf. (2.18)-(2.22)). For *plane* dielectric interfaces, the plane-wave-excited PO “surface currents” would be *exactly* linearly phased; therefore the tilted beam discretization in (2.18)-(2.22) could be applied straightforwardly and, as shown in [2, Sec. II-C], would be considerably more efficient than the nontilted algorithm in (2.13)-(2.16). For *moderately rough* interfaces, the phasing in the PO surface currents is no longer *globally* linear. However, for narrow-waisted beams, one can still exploit the *locally* linear behavior. To this end, it is expedient to rewrite the PO currents by separating out the locally linear phase term that the incident plane wave would induce on the locally tangent plane, i.e., (generalizing (2.17))

$$J_{PO}^\nu(x) = \mathcal{J}_{PO}^\nu(x) \exp[ik_0 x \sin \theta^i(x) \cos \alpha(x)], \quad \nu = r \text{ or } t, \quad (17)$$

with $\theta^i(x)$ defined in (14). Here and henceforth the ω dependence is omitted for simplicity of notation. The weakly phased *reduced* PO currents

$$\mathcal{J}_{PO}^\nu(x) = J_{PO}^\nu(x) \exp[-ik_0 x \sin \theta^i(x) \cos \alpha(x)], \quad \nu = r \text{ or } t \quad (18)$$

can thus be parameterized approximately as (see (2.18)-(2.19))

$$\mathcal{J}_{PO}^\nu(x) \approx \sum_{|m| \leq (d/2L)} C_m^\nu w(x - x_m), \quad \nu = r \text{ or } t, \quad (19)$$

where the subscript m tags the m -th GB in the discretization and $w(\cdot)$ represents the normalized Gaussian window in (2.6),

$$w(x) = \left(\frac{\sqrt{2}}{L} \right)^{1/2} \exp \left[-\pi (x/L)^2 \right], \quad \int_{-\infty}^{\infty} w^2(x) dx = 1, \quad (20)$$

with the Gabor coefficients C_m^ν given approximately by

$$C_m^\nu = \left(\frac{L}{\sqrt{2}} \right)^{1/2} \mathcal{J}_{PO}^\nu(x_m), \quad \nu = r \text{ or } t. \quad (21)$$

The sum in (19) extends up to $|m| \leq d/2L$ because the PO surface currents (subject to verification) are assumed to be negligible for $|x| \geq d/2$, outside the illumination window (see Fig. 1b). By combining (7) and (15) with (17) and (19), the reflected and transmitted fields can be discretized as (see also (2.18))

$$E^\nu(x, z) \sim \frac{1}{2} \sum_{|m| \leq (d/2L)} C_m^\nu \mathcal{B}_m^\nu(x, z), \quad \nu = r \text{ or } t, \quad (22)$$

where the beam propagators $\mathcal{B}_m^r, \mathcal{B}_m^t$ are given by

$$\mathcal{B}_m^r(x, z) = -2 \int_{C_{PO}} w(x' - x_m) \frac{\partial}{\partial \zeta} G_{2D}(x, z; x', h(x'); k_0) \exp[ik_0 x' \sin \theta^i(x') \cos \alpha(x')] d\ell', \quad (23)$$

$$\mathcal{B}_m^t(x, z) = 2 \int_{C_{PO}} w(x' - x_m) \frac{\partial}{\partial \zeta} G_{2D}(x, z; x', h(x'); k) \exp[ik_0 x' \sin \theta^i(x') \cos \alpha(x')] d\ell'. \quad (24)$$

For narrow Gaussian windows, i.e., $L \lesssim \lambda_0 \ll d$, and moderate surface roughness, the paraxial far zone approximation in [7] can be applied to (23) and (24). For the reflected field, one obtains the following *complex source point* (CSP) approximation which applies (2.20) locally (see Appendix A for details)

$$\begin{aligned} \mathcal{B}_m^r(x, z) \sim & -ik_0 2^{5/4} \left(\frac{L}{8\pi k_0} \right)^{1/2} \frac{(\zeta_m - ib_m^r \cos \theta_m^i)}{(\tilde{R}_m^r)^{3/2}} \\ & \cdot \exp \left\{ i \left[k_0 \left(\tilde{R}_m^r + x_m \sin \theta_m^i \cos \alpha_m + ib_m^r \right) + \pi/4 \right] \right\}, \end{aligned} \quad (25)$$

where (see Fig. 3) $\theta_m^i \equiv \theta^i(x_m)$, $\alpha_m \equiv \alpha(x_m)$,

$$\zeta_m = -(x - x_m) \sin \alpha_m + [z - h(x_m)] \cos \alpha_m. \quad (26)$$

Furthermore,

$$\tilde{R}_m^r = \sqrt{(x_{bm}^r)^2 + (z_{bm}^r - ib_m^r)^2}, \quad \text{Re}[\tilde{R}_m^r] \geq 0, \quad (27)$$

is the complex distance, in the beam coordinates of Fig. 3, between the observation point

$$\begin{bmatrix} x_{bm}^r \\ z_{bm}^r \end{bmatrix} = \begin{bmatrix} \cos \gamma_m^r & -\sin \gamma_m^r \\ \sin \gamma_m^r & \cos \gamma_m^r \end{bmatrix} \begin{bmatrix} x - x_m \\ z - h(x_m) \end{bmatrix}, \quad \gamma_m^r = \theta_m^i - \alpha_m = \theta_A - 2\alpha_m, \quad (28)$$

and the CSP $(0, ib_m^r)$, with the complex displacement parameter b_m^r given by (generalizing (2.22)),

$$b_m^r = (L \cos \alpha_m \cos \theta_m^i)^2 / \lambda_0. \quad (29)$$

Here and henceforth, the tilde \sim identifies dependence on analytically continued spatial source coordinates. The beam discretization in (22) with the CSP propagators in (25) is physically appealing because it represents a superposition of GBs launched from points $(x_m, h(x_m))$ on the illuminated portion of the interface along the reflection directions γ_m^r (see Fig. 3), which are *locally matched* to the surface currents. Similarly, for the transmitted field one obtains

$$\mathcal{B}_m^t(x, z) \sim ik 2^{5/4} \left(\frac{L}{8\pi k} \right)^{1/2} \frac{(\zeta_m - ib_m^t \cos \theta_m^R)}{(\tilde{R}_m^t)^{3/2}}.$$

$$\cdot \exp \left\{ i \left[k \left(\tilde{R}_m^t + x_m \sin \theta_m^R \cos \alpha_m + i b_m^t \right) + \pi/4 \right] \right\}, \quad (30)$$

where θ_m^R is the local refraction angle (see Fig. 4), related to θ_m^i and ϵ_r^e via Snell's law

$$\sin \theta_m^i = \sqrt{\epsilon_r^e} \sin \theta_m^R. \quad (31)$$

Moreover, in the beam coordinates of Fig. 4, \tilde{R}_m^t is the complex distance

$$\tilde{R}_m^t = \sqrt{(x_{bm}^t)^2 + (z_{bm}^t - i b_m^t)^2}, \quad \text{Re}[\tilde{R}_m^t] \geq 0, \quad (32)$$

with

$$\begin{bmatrix} x_{bm}^t \\ z_{bm}^t \end{bmatrix} = \begin{bmatrix} \cos \gamma_m^t & \sin \gamma_m^t \\ \sin \gamma_m^t & -\cos \gamma_m^t \end{bmatrix} \begin{bmatrix} x - x_m \\ z - h(x_m) \end{bmatrix}, \quad \gamma_m^t = \theta_m^R + \alpha_m, \quad (33)$$

and the complex displacement parameter given by

$$b_m^t = \sqrt{\epsilon_r^e} (L \cos \alpha_m \cos \theta_m^R)^2 / \lambda_0. \quad (34)$$

Like the reflected field, the transmitted field is synthesized via (22) and (30) with *surface current-matched* CSP GBs propagating along the *local refraction direction* γ_m^t (see Fig. 4). Since the dielectric half-space is assumed to be lossy, θ_m^R , γ_m^t and b_m^t are complex.

IV. TIME DOMAIN FORMULATION

A. Preliminary Considerations

The FD beam propagators in (25) and (30) are nearly identical with that in (2.20). Therefore, extension to the time domain (TD) for pulsed excitation can be pursued by following the procedure in [2, Sec. III-A]. However, the TD inversion is now complicated by the dispersive (Ohmic) properties of the dielectric half-space (cf. (13)). Although analytic approaches to PB propagation in Ohmic-dispersive media are available [16], here we use a simpler approach, restricted to slightly lossy materials, i.e.,

$$\frac{\sigma}{\omega \epsilon_0 \epsilon_r} \ll 1, \quad \omega < \Omega, \quad (35)$$

with Ω denoting the effective bandwidth of the pulse $p(t)$. This condition is fulfilled for a class of actual ground penetrating radar (GPR) applications which is of interest to us

(see, e.g., the discussion in [17]), and which allows the complex wavenumber to be written as [17]

$$k = \frac{\omega}{c} \sqrt{\epsilon_r^e} \approx \sqrt{\epsilon_r} \frac{\omega}{c} \left(1 + i \frac{\sigma}{2\omega\epsilon_0\epsilon_r} \right) = \sqrt{\epsilon_r} \frac{\omega}{c} + i \frac{\sigma}{2c\epsilon_0\sqrt{\epsilon_r}}, \quad (36)$$

i.e., with a *frequency-independent* imaginary part. This *nondispersive* approximation considerably simplifies the TD inversion. In what follows, we shall be using (36) for calculating the transmitted field. Moreover, we approximate the frequency-dependent reflection coefficient in (12) by its value at the center angular frequency Ω_0 of the pulse,

$$\mathcal{R}(x, \omega) \approx \mathcal{R}_0(x) \equiv \mathcal{R}(x, \Omega_0). \quad (37)$$

As in [2], we assume a frequency-independent Gabor lattice parameter L in order to conveniently estimate the TD Gabor coefficients by aperture sampling (see the discussion in [2, Sec. III-A]). Anticipating Fourier inversion, recalling (5), (11), (16), (18) and (37), the FD Gabor coefficients in (21) can be written as

$$C_m^\nu \approx 2c_m^\nu P(\omega) \exp(i\omega t_m), \quad \nu = r \text{ or } t, \quad (38)$$

where

$$c_m^r = (L/\sqrt{2})^{1/2} g(x_{Bm}) \mathcal{R}_0(x_m), \quad c_m^t = (L/\sqrt{2})^{1/2} g(x_{Bm}) [1 + \mathcal{R}_0(x_m)] \quad (39)$$

are frequency independent, and

$$t_m = c^{-1} \left(z_{Bm} - x_m \sin \theta_m^i \cos \alpha_m \right), \quad (40)$$

with (x_{Bm}, z_{Bm}) denoting the lattice points $(x_m, h(x_m))$ in the incident-beam coordinates (4),

$$\begin{bmatrix} x_{Bm} \\ z_{Bm} \end{bmatrix} = \begin{bmatrix} \cos \theta_A & \sin \theta_A \\ \sin \theta_A & -\cos \theta_A \end{bmatrix} \begin{bmatrix} x_m - x_A \\ h(x_m) - z_A \end{bmatrix}. \quad (41)$$

The FD Gabor expansions in (22) can thus be rewritten as

$$E^\nu(x, z, \omega) \sim \sum_{|m| \leq (d/2L)} c_m^\nu P(\omega) \mathcal{B}_m^\nu(x, z, \omega) \exp(i\omega t_m), \quad \nu = r \text{ or } t. \quad (42)$$

As in [2], in order to deal with the evanescent spectra in the beam propagators \mathcal{B}_m^r and \mathcal{B}_m^t , the TD is accessed by Fourier inversion of (42) via the analytic signal formulation (cf. (2.31), (2.32)) [18]

$$f^+(t) = \frac{1}{\pi} \int_0^\infty F(\omega) \exp(-i\omega t) d\omega, \quad \text{Im}(t) \leq 0, \quad (43)$$

where $F(\omega)$ is the conventional Fourier spectrum of the real signal $f(t)$, and the real signal for real t is recovered via

$$f(t) = \text{Re} \left[f^+(t) \right]. \quad (44)$$

Accordingly (cf. (2.33)),

$$e^\nu(x, z, t) \sim \sum_{|m| \leq (d/2L)} c_m^\nu \mathbf{b}_m^\nu(x, z, t - t_m), \quad \nu = r \text{ or } t, \quad (45)$$

with the analytic PB propagators \mathbf{b}_m^ν given by

$$\mathbf{b}_m^\nu(x, z, t) = \frac{1}{\pi} \int_0^\infty \mathcal{B}_m^\nu(x, z, \omega) P(\omega) \exp(-i\omega t) d\omega, \quad \text{Im}(t) \leq 0, \quad \nu = r \text{ or } t, \quad (46)$$

and the real TD fields obtained from

$$e^\nu(x, z, t) = \text{Re} \left[\mathbf{e}^\nu(x, z, t) \right], \quad \nu = r \text{ or } t. \quad (47)$$

The analytic PB propagators \mathbf{b}_m^r and \mathbf{b}_m^t in (46) should not be confused with the complex displacement parameters b_m^r and b_m^t in (29) and (34), respectively.

As in [2], we consider a class of Rayleigh (differentiated Gaussian) pulses (cf. (2.37), (2.38)),

$$p(t) = \Xi \frac{d^j}{dt^j} \exp \left[- \left(\frac{t - T/2}{\varsigma T} \right)^2 \right], \quad (48)$$

$$P(\omega) = \sqrt{\pi} \Xi \varsigma T (-i\omega)^j \exp \left(-\frac{\varsigma^2 \omega^2 T^2}{4} + i \frac{\omega T}{2} \right), \quad (49)$$

where Ξ is a normalization constant, and the variance ς is chosen so that the pulse width of $p(t)$ is $\sim T$. We now generalize the TD results in [2, Sec. III-A], starting with the reflected field.

B. Reflected Field

As in [2], in order to evaluate the integral in (46) with (25) and (49), we first approximate the complex distance \tilde{R}_m^r in (27), which is frequency dependent via (29). Assuming L/λ_0 (and therefore b_m^r) sufficiently small in the amplitude factor of (25) we let (cf. (2.39))

$$\frac{(\zeta_m - ib_m^r \cos \theta_m^i)}{(\tilde{R}_m^r)^{3/2}} \approx \frac{\zeta_m}{(R_m^r)^{3/2}}, \quad R_m^r = \sqrt{(x_{bm}^r)^2 + (z_{bm}^r)^2}, \quad b_m^r \ll \zeta_m, z_{bm}^r, \quad (50)$$

where R_m^r is real and frequency-independent, with ζ_m and (x_{bm}^r, z_{bm}^r) defined in (26) and (28), respectively. In the phase factor, instead of the paraxial approximation in (2.40), we use a perturbation (first-order McLaurin) approximation in terms of b_m^r (see (29)),

$$\tilde{R}_m^r(\omega) \approx R_m^r - ib_m^r \frac{z_{bm}^r}{R_m^r} = R_m^r - i\omega \frac{z_{bm}^r (L \cos \alpha_m \cos \theta_m^i)^2}{2\pi c R_m^r}, \quad b_m^r \ll z_{bm}^r, \quad (51)$$

which was found to give better results. Using (50) and (51), the beam propagator in (25) can be rewritten as

$$\mathcal{B}_m^r(x, z, \omega) \sim \Lambda_m^r \omega^{1/2} \exp[-(\omega T_m^r)^2/4 + i\omega \tau_m^r], \quad (52)$$

where

$$\Lambda_m^r = -i2^{5/4} \exp(i\pi/4) c^{-1} \sqrt{\frac{L}{8\pi}} \frac{\zeta_m}{(R_m^r)^{3/2}}, \quad (53)$$

$$\tau_m^r = c^{-1} (R_m^r + x_m \sin \theta_m^i \cos \alpha_m + T/2), \quad T_m^r = \frac{L \cos \alpha_m \cos \theta_m^i}{2c\sqrt{2\pi}} \sqrt{1 - \frac{z_{bm}^r}{R_m^r}}. \quad (54)$$

Substituting (49) and (52) into (46), one obtains a canonical integral [19],

$$\begin{aligned}
I(p, q, j) &= \int_0^\infty \omega^{j+1/2} \exp(ip\omega - q^2\omega^2/4) d\omega = \\
&= 2^{j+1/2} q^{-j-5/2} \left[q \Gamma\left(\frac{3+2j}{4}\right) M_1^{(j)}\left(\frac{p}{q}\right) + 2ip \Gamma\left(\frac{5+2j}{4}\right) M_2^{(j)}\left(\frac{p}{q}\right) \right], \quad q \in R^+, \quad (55)
\end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function [20], and

$$M_1^{(j)}(t) = {}_1F_1\left(\frac{3+2j}{4}, \frac{1}{2}, -t^2\right), \quad M_2^{(j)}(t) = {}_1F_1\left(\frac{5+2j}{4}, \frac{3}{2}, -t^2\right), \quad (56)$$

with ${}_1F_1(u, v, t)$ denoting the Kummer confluent hypergeometric function [20]. The reflected analytic PB propagator can be thus written as

$$b_m^+(x, z, t) = (-1)^j \beta_m^r \left[\mathcal{T}_m^r \Gamma\left(\frac{3+2j}{4}\right) M_1^{(j)}\left(\frac{t - \tau_m^r}{\mathcal{T}_m^r}\right) - 2i(t - \tau_m^r) \Gamma\left(\frac{5+2j}{4}\right) M_2^{(j)}\left(\frac{t - \tau_m^r}{\mathcal{T}_m^r}\right) \right], \quad (57)$$

where

$$\beta_m^r = 2^{j+1/2} \pi^{-1/2} (\mathcal{T}_m^r)^{-j-5/2} \Lambda_m^r \Xi \zeta T, \quad \mathcal{T}_m^r = \sqrt{(T_m^r)^2 + \zeta^2 T^2}. \quad (58)$$

Equation (57) generalizes (2.44) to arbitrary order j of the derivative in (48). The functions $M_{1,2}^{(j)}$ can be efficiently computed using the rapidly converging expansions in Appendix B.

C. Transmitted Field

The transmitted PB propagator can be obtained similarly. The only slight difference is due to the complex wavenumber k and the complex parameters θ_m^R and b_m^t in (30) (see (31) and (34)). For slightly lossy materials as in (35), the complex wavenumber is approximated via (36), whereas θ_m^R and b_m^t are approximated by the real values

$$\theta_m^R \approx \theta_{m0}^R \equiv \sin^{-1} \left(\sin \theta_m^i / \sqrt{\epsilon_r} \right), \quad b_m^t \approx b_{m0}^t \equiv \sqrt{\epsilon_r} (L \cos \alpha_m \cos \theta_{m0}^R)^2 / \lambda_0. \quad (59)$$

Accordingly,

$$b_m^+(x, z, t) = (-1)^j \beta_m^t \left[\mathcal{T}_m^t \Gamma\left(\frac{3+2j}{4}\right) M_1^{(j)}\left(\frac{t - \tau_m^t}{\mathcal{T}_m^t}\right) - 2i(t - \tau_m^t) \Gamma\left(\frac{5+2j}{4}\right) M_2^{(j)}\left(\frac{t - \tau_m^t}{\mathcal{T}_m^t}\right) \right], \quad (60)$$

where

$$\beta_m^t = 2^{j+1/2} \pi^{-1/2} (\mathcal{T}_m^t)^{-j-5/2} \Lambda_m^t \Xi \zeta T, \quad \mathcal{T}_m^t = \sqrt{(T_m^t)^2 + \zeta^2 T^2}, \quad (61)$$

$$\Lambda_m^t = i(\epsilon_r)^{1/4} 2^{5/4} \exp(-\kappa R_{m0}^t + i\pi/4) c^{-1} \sqrt{\frac{L}{8\pi}} \frac{\zeta_m}{(R_{m0}^t)^{3/2}}, \quad \kappa = \frac{\sigma}{2c\epsilon_0 \sqrt{\epsilon_r}}, \quad (62)$$

$$\tau_m^t = c^{-1} \left[\sqrt{\epsilon_r} (R_{m0}^t + x_m \sin \theta_{m0}^R \cos \alpha_m) + T/2 \right], \quad T_m^t = \sqrt{\epsilon_r} \frac{L \cos \alpha_m \cos \theta_{m0}^R}{2c\sqrt{2\pi}} \sqrt{1 - \frac{z_{bm0}^t}{R_{m0}^t}}, \quad (63)$$

$$R_{m0}^t = \sqrt{(x_{m0}^t)^2 + (z_{m0}^t)^2}, \quad \begin{bmatrix} x_{bm0}^t \\ z_{bm0}^t \end{bmatrix} = \begin{bmatrix} \cos \gamma_{m0}^t & \sin \gamma_{m0}^t \\ \sin \gamma_{m0}^t & -\cos \gamma_{m0}^t \end{bmatrix} \begin{bmatrix} x - x_m \\ z - h(x_m) \end{bmatrix}, \quad \gamma_{m0}^t = \theta_{m0}^R + \alpha_m, \quad (64)$$

D. Limitations

The limitations of the proposed approach can be divided into two categories. The first category includes the *model constraints*, i.e., the underlying PO approximation and all other simplifying assumptions. These constraints are known *a priori*, and can be summarized as follows

- i) Moderately rough interfaces (both in height and slope) with local curvature radii large compared to the pulse length cT ;
- ii) Plane-wave excitation (cf. (3)) with incidence direction far from grazing;
- iii) Slightly lossy dielectrics (cf. (35)).

Strong roughness and/or near-grazing incidence would require more sophisticated models than the simple PO approximation in Sec. III-A. More general (e.g., focused) excitations would require a two-step procedure: first discretizing the aperture field distribution in terms of narrow-waisted GBs as in [2] and subsequently applying the PO algorithm to each incident beam. More sophisticated loss/dispersion models could also be included in principle [16].

The second type of constraint is related to the adequacy of the narrow-waisted PB discretization in Secs. IV-B and IV-C, in terms of the number of beams required to guar-

antee stability of the outcome. We have referred to this as the *scrambling* criterion, i.e., the insensitivity of the result with respect to different combinations of the beam/lattice configuration. Overall reliability requires that both the model constraints and the scrambling criterion are satisfied.

V. NUMERICAL RESULTS

A. Reference Solution

The PB syntheses presented in Sec. IV have been validated and calibrated against an independent reference solution based on the time-harmonic multifilament current method in [21], adapted to moderately rough interfaces (cf. App. B in [1]). The frequency spectra of the reflected and transmitted fields were obtained by solving the FD problem at 100 different frequencies within the pulse bandwidth. The incident FD field was computed via numerical integration of the rigorous Kirchhoff aperture distribution (the spectrum of (2)), without resorting to the plane wave approximation, and with use of the full dispersive permittivity model in (13) for the dielectric half-space. The resulting frequency samples were smoothed through local Padé-approximation [22] and filtered by the pulse spectrum $P(\omega)$ in (49). The TD solution was then obtained via standard inverse FFT routines [22].

B. Simulation Parameters

The numerical simulations that follow are based on the pulsed aperture field distribution in (2) with the Gaussian taper

$$g(x) = \exp[-18x^2/(d \cos \theta_A)^2], \quad (65)$$

and the wideband fourth-order Rayleigh pulse $p(t)$, obtained from (48) with $j = 4$, $\Xi = T^4/30000$, $\varsigma = 1/\sqrt{50}$, and length $cT = 0.08d$ (see Fig. 5a), which has a center angular frequency $\Omega_0 = 20/T$ (see Fig. 5b). For the reflected/transmitted fields, the special functions $M_{1,2}^{(4)}$ in (57) and (60) were computed via the rapidly converging expansions in Appendix B with $N = 5$ (cf. (2.50) and (2.51)). The aperture height z_A was chosen so as to place the rough dielectric interface within the collimation zone of the aperture, thereby justifying the plane-wave incidence approximation in (3); depending on the tilt

angle θ_A , the spatial shift x_A was adjusted so that the illuminated region was confined in the interval $[-d/2, d/2]$ (see Fig. 1b). In each of the examples below, it was verified that the illumination at the edges $x = \pm d/2$ was at least 30 dB below the maximum strength so that numerical artifacts due to edge effects were negligible. The rough surface profile was generated via the quartic spline model in [1]. For the GPR applications of interest in our studies, the dielectric half-space parameters were chosen so as to simulate a class of realistic soils (Puerto Rican clay loam [23]), with the frequency range chosen to satisfy (35).

C. Results

We begin with the problem geometry in Fig. 6, with the relevant parameters specified in the figure caption. For this configuration, the roughness is moderate both in height ($h_{max} = 0.36cT$) and slope ($\phi_{max} = 31^\circ$), and the average radius of curvature \bar{r}_c is large with respect to the pulse length ($\bar{r}_c = 5.9cT$). The incident field direction is vertical ($\theta_A = 0$), and the constitutive parameters ϵ_r, σ satisfy (35). The previously stated conditions for validity of the proposed PB synthesis in (47) should be thus satisfied, and good accuracy should be expected with an “adequate” number of beams. The reflected and transmitted fields have been computed via (47) with (57) and (60), respectively, at nine observation points (three different horizontal positions, spanning the illuminated region, on three observation planes). The reflected/transmitted temporal waveforms obtained via the PB synthesis with 150 beams are shown in Figs. 7 and 8, respectively, and are compared with the reference solution. Very good agreement is observed. The reflected waveforms reveal details which can be tied to the space-time dependent scattering processes. At certain observation points, especially at smaller observation distance (cf. Figs. 7a, 7c, 7f), the dominant contributions turn out to be well separated and replicate the incident waveform (inverted due to reflection). This is not the case in the central region ($x = 0$), where the almost simultaneous arrivals from the surface do not allow their separate resolution for the specified input pulse width, thereby resulting in the chirped oscillations in Figs. 7b, 7e, 7h. Separate (non-inverted) arrivals are also visible in the transmitted field (cf. Fig. 8i); at smaller observation distance (e.g., Figs. 8a, 8c), however, the transmitted field waveforms coalesce into a close replication of the incident pulse. In all examples, the PB

syntheses capture even the finer details.

The number of beams in these examples was arrived at using the pragmatic stability criterion discussed in Sec. IV-D. In order to better quantify the accuracy, and address convergence issues, we have computed the r.m.s. (energy) errors

$$\Delta e^\nu = \frac{\int_{-\infty}^{\infty} |e_r^\nu(x, z, t) - e_b^\nu(x, z, t)|^2 dt}{\left[\int_{-\infty}^{\infty} |e_r^\nu(x, z, t)(x, z, t)|^2 dt \int_{-\infty}^{\infty} |e_b^\nu(x, z, t)(x, z, t)|^2 dt \right]^{1/2}}, \quad \nu = r \text{ or } t, \quad (66)$$

where the subscript $_r$ denotes the reference solution and $_b$ denotes the PB synthesis. The r.m.s. errors for Figs. 7 and 8 are < -25 dB. The convergence of the algorithm in Figs. 7 and 8 is illustrated in Fig. 9, where the r.m.s. errors for both reflected and transmitted fields are plotted vs. the number of beams $N_b = d/L$. The various curves pertain to different observation points. It is noted that beyond a critical threshold, the error becomes *practically insensitive* to a further increase in the number of beams, indicating that convergence has been achieved. “Convergence” here implies that the *PO integral* is adequately beam-discretized but, as noted in Sec. IV-D, this *does not* necessarily imply that the *overall* solution is accurate; good *overall* accuracy requires that the model constraints are likewise satisfied. The test configuration was chosen deliberately so that this is the case. In this example, the convergence turns out to be weakly dependent on the observation point, and a robust threshold for *uniformly* good accuracy ($\Delta e^{r,t} < -25$ dB) can be set around $N_b = 150$. If the *model constraints* in Sec. IV-D were violated, the PB synthesis, although stabilized, could be *inaccurate*. For illustration, we strained the algorithm by selecting simulation parameters near the limit of their range of validity. The results in Fig. 10 pertain to the profile in Fig. 6, conformally scaled so as to increase the roughness up to a maximum height $h_{max} = 0.6cT$ and a maximum slope $\phi_{max} = 46^\circ$, with an average curvature radius $\bar{r}_c = 4.4cT$. The number of beams was chosen to satisfy the stability criterion. Remarkably, the PB synthesis still performs well, but now one observes discrepancies with respect to the reference solution (particularly in one late-time peak (Fig. 10a)) which *cannot* be repaired by increasing the number of beams. Note that the increased roughness admits multiple reflections which are accounted for in the reference solution but

ignored in the PO algorithm. This may contribute to the diminished accuracy, which now yields r.m.s. errors $\Delta e^r = -10$ dB, $\Delta e^t = -20$ dB. The deterioration is aggravated further for oblique incidence, where multiple interactions are more pronounced, as seen in Fig. 11, where the profile of Fig. 10 is illuminated by a tilted beam with $\theta_A = 40^\circ$. For this example, $\Delta e^r = -8$ dB and $\Delta e^t = -9$ dB. We also investigated the accuracy degradation in the “low frequency” (i.e., long pulse) limit. As an illustration, the reflected and transmitted waveforms at various observation points in Fig. 12 pertain to the geometry in Fig. 6, but using a ten-times longer excitation pulse ($cT = 0.8$). For this configuration, $cT = 0.8d$ and $\bar{r}_c = 0.44cT$; the reduced radius of curvature stretches the validity of the asymptotics in the PO model and in the CSP approximations (25) and (30). First, one notes from Fig. 12 that the reflected/transmitted waveforms contain less structure than those in Figs. 7, 8, 10, 11, since the surface is now *flatter* on the pulse length scale ($h_{max} = 0.06cT$). Second, due to the lower-frequency content of the pulse, recalling (29) and (51), a larger Gabor lattice period L (i.e., fewer beams) should be adequate to stabilize the beam synthesis. This was confirmed in numerical simulations, where the beam syntheses were found to stabilize around $N_b = 30$. As seen from Fig. 12, the beam syntheses, though reasonably good, are no longer *highly* accurate, but have r.m.s. errors ranging from -25 dB to -5 dB. As in Figs. 10 and 11, the accuracy *cannot* be improved by increasing the number of beams, indicating that the limit in range of validity of the PO/CSP asymptotics has been reached.

From a practical viewpoint, we have found fairly accurate predictions ($\Delta e^{r,t} \lesssim -20$ dB) for roughness with maximum height $h_{max} \lesssim 0.5cT$, (average) curvature radii $\bar{r}_c \gtrsim 2cT$, maximum slopes $\phi_{max} \lesssim 40^\circ$, for nearly-vertical incidence ($\theta_A \lesssim 30^\circ$), and for dielectrics with $\sigma/(\Omega_0\epsilon_0\epsilon_r) \lesssim 0.05$. For the examples in Figs. 7, 8, 10 and 11, 150–200 beams were usually found to be sufficient to reach convergence, resulting in an average computing time $T_c = 8$ ms per field time-sample (at a fixed observation point) on a 500 MHz PC. Note that the computing time scales linearly with the number of beams; accordingly, for the examples in Fig. 12 (i.e., 30 beams) we found $T_c = 2$ ms.

VI. CONCLUSIONS

A Gabor-based, quasi-ray, pulsed beam (PB) algorithm has been presented for short-pulse 2D reflection by, and transmission through, a 1D moderately rough interface separating free space from a slightly lossy dielectric half-space. The approach is based on the Kirchhoff PO approximation and the PB 1D aperture field discretization in [2], and has been validated and calibrated against an independently generated reference solution. Numerical simulations show that the proposed algorithm yields fast, robust and accurate predictions in a calibrated range of parameters, and therefore is promising as a useful forward model for inverse scattering scenarios that involve buried objects. Extension to more general dispersion/loss models and to non-collimated aperture excitations remain to be investigated. Extension to 2D surfaces, and 3D field scattering, based on 2D aperture PB discretization [24], is straightforward in principle, but its computational features remain to be explored.

APPENDIX A

DETAILS PERTAINING TO EQ. (25)

For narrow Gaussian windows, with $L \lesssim \lambda_0 \ll d$, the integrand in (23) is strongly localized around $x = x_m$. Thus, for moderate roughness, the integration path near x_m can be approximated by the local tangent plane, and the phase function can be assumed to be linear. Accordingly, in the (ξ_m, ζ_m) coordinate system of Fig. 3,

$$\begin{bmatrix} \xi_m \\ \zeta_m \end{bmatrix} = \begin{bmatrix} \cos \alpha_m & \sin \alpha_m \\ -\sin \alpha_m & \cos \alpha_m \end{bmatrix} \begin{bmatrix} x - x_m \\ z - h(x_m) \end{bmatrix}, \quad \alpha_m \equiv \alpha(x_m), \quad (67)$$

the reflected-beam integral in (23) can be approximated as

$$\mathcal{B}_m^r(x, z) \approx -2 \int_{-\infty}^{\infty} w_{\alpha m}(\xi'_m) \frac{\partial}{\partial \zeta_m} G_{2D}(\xi_m, \zeta_m; \xi'_m, 0; k_0) \exp[ik_0(\xi'_m + x_m \cos \alpha_m) \sin \theta_m^i] d\xi'_m, \quad (68)$$

where $w_{\alpha m}(\cdot)$ is the Gaussian window in (20) projected onto the local tangent plane (ξ_m axis in Fig. 3),

$$w_{\alpha m}(x) = w(x/\cos \alpha_m) = \left(\frac{\sqrt{2}}{L}\right)^{1/2} \exp \left[-\pi \left(\frac{x}{L \cos \alpha_m} \right)^2 \right]. \quad (69)$$

The integral in (68) has the same form as (2.9). Therefore, straightforward application of the paraxial far zone approximation detailed in the Appendix of [7] yields the result in (25). Similar considerations applied to the transmitted beams in (24) yield (30) (see Fig. 4).

APPENDIX B

RAPIDLY CONVERGING EXPANSIONS FOR $M_{1,2}^{(j)}$

As in [2, App. C], using the Kummer transformation and the truncated Taylor series for the confluent hypergeometric function ${}_1F_1$, one obtains [20]

$$M_1^{(j)}(t) = \exp(-t^2) {}_1F_1 \left(-\frac{5+j}{4}, \frac{1}{2}, t^2 \right) \approx \exp(-t^2) \sum_{n=0}^N \frac{\left(-\frac{5+j}{4} \right)_n}{\left(\frac{1}{2} \right)_n n!} t^{2n}, \quad (70)$$

$$M_2^{(j)}(t) = \exp(-t^2) {}_1F_1 \left(-\frac{3+j}{4}, \frac{3}{2}, t^2 \right) \approx \exp(-t^2) \sum_{n=0}^N \frac{\left(-\frac{3+j}{4} \right)_n}{\left(\frac{3}{2} \right)_n n!} t^{2n}, \quad (71)$$

with $(u)_n$ denoting the Pochhammer symbol [20],

$$(u)_n = u(u+1)(u+2) \cdots (u+n-1), \quad (u)_0 = 1. \quad (72)$$

As shown in [2, Fig. 5], the expansions in (70) and (71) guarantee satisfactory accuracy with few terms ($N \approx 5$).

REFERENCES

- [1] V. Galdi, L.B. Felsen, and D.A. Castañón, "Quasi-ray Gaussian beam algorithm for time-harmonic two-dimensional scattering by moderately rough interfaces," *IEEE Trans. Antennas Propagat.*, vol. 49, No. 9, Sept. 2001 (in print).
- [2] V. Galdi, L.B. Felsen, and D.A. Castañón, "Narrow-waisted Gaussian beam discretization for short-pulse radiation from one-dimensional large apertures," *IEEE Trans. Antennas Propagat.*, vol. 49, No. 10, Oct. 2001 (in print).
- [3] T. Dogaru and L. Carin, "Time-domain sensing of targets buried under a rough air-ground interface," *IEEE Trans. Antennas Propagat.*, vol. 46, No. 3, pp. 360-372, Mar. 1998.
- [4] G.A. Tsihrintzis, C.M. Rappaport, S.C. Winton, and P.M. Johansen, "Statistical modeling of rough surface scattering for ground-penetrating radar applications," in *Detection and Remediation Technologies for Mines and Mine-Like Targets III, Proc. SPIE*, vol. 3392, pp. 735-744, 1998.

- [5] T. Dogaru and L. Carin, "Multiresolution time domain analysis of scattering from a rough dielectric surface," *Radio Sci.*, vol. 35, No. 6, pp. 1279-1292, Nov.-Dec. 2000.
- [6] L.B. Felsen and V. Galdi, "Complex-source-point narrow-waisted ray-like Gaussian beams for frequency and time domain radiation and scattering," in *Ultra-Wideband, Short Pulse Electromagnetics 5*. New York (NY): Kluwer/Plenum Publishers (S.R. Cloude and P.D. Smith, Eds.), 2001 (in print).
- [7] J.J. Maciel and L.B. Felsen, "Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part I - plane layer," *IEEE Trans. Antennas Propagat.*, vol. 38, No. 10, pp. 1607-1617, Oct. 1990.
- [8] J.J. Maciel and L.B. Felsen, "Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part II - circular cylindrical layer," *IEEE Trans. Antennas Propagat.*, vol. 38, No. 10, pp. 1618-1624, Oct. 1990.
- [9] R.F. Harrington, *Time-Harmonic Electromagnetic Fields*. New York (NY): McGraw-Hill, 1961, p. 127.
- [10] M. V. Vesnik and P.Ya. Ufimtsev, "An asymptotic feature of corner waves scattered by polygonal plates," *Electromagn.*, vol. 12, pp. 265-272, 1992.
- [11] E.I. Thorsos, "The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum," *J. Acoust. Soc. Am.*, vol. 83, pp. 78-82, Aug. 1988.
- [12] A. Collaro, G. Franceschetti, M. Migliaccio, and D. Riccio, "Gaussian rough surfaces and Kirchhoff approximation," *IEEE Trans. Antennas Propagat.*, vol. 47, No. 2, pp. 392-398, Feb. 1999.
- [13] G. Franceschetti, A. Iodice, M. Migliaccio, and D. Riccio, "Scattering from natural rough surfaces modeled by fractional Brownian motion two-dimensional processes," *IEEE Trans. Antennas Propagat.*, vol. 47, No. 9, pp. 1405-1415, Sep. 1999.
- [14] J.A. DeSanto and G.S. Brown, "Analytical techniques for multiple scattering from rough surfaces," in *Progress in Optics XXIII*, E. Wolf, Ed. Amsterdam: North-Holland, 1986.
- [15] L.C. Kempel, J.L. Volakis, T.B.A. Senior, S.S. Locus, and K.M. Mitzner, "Scattering by S-shaped surfaces," *IEEE Trans. Antennas Propagat.*, vol. 41, No. 5, pp. 701-708, June 1993.
- [16] E. Heyman, A.G. Tijhuis, and J. Boersma, "Spherical and collimated pulsed fields in conducting media," *Proc. URSI Trium Int. Symp. on Electromagnetic Theory*, St. Petersburg, Russia, 1995, pp. 643-645.
- [17] T.B. Hansen and P.M. Johansen, "Inversion scheme for ground penetrating radar that takes into account the planar air-soil interface," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, No. 1, pp. 496-506, Jan. 2000.
- [18] T.B. Hansen and A.D. Yaghjian, *Plane-Wave Theory of Time-Domain Fields: Near-Field Scanning Applications*, Piscataway (NJ): IEEE Press, 1999.
- [19] S. Wolfram, *The Mathematica Book*, 3rd ed. New York (NY): Wolfram Media/Cambridge University Press, 1996.
- [20] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*. New York (NY): Dover, 1964.
- [21] Y. Leviatan and A. Boag, "Analysis of electromagnetic scattering from dielectric cylinders using a multifilament current model," *IEEE Trans. Antennas Propagat.*, vol. 35, No. 10, pp. 1119-1127, Oct. 1987.
- [22] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge (UK): Cambridge Univ. Press, 1992.
- [23] J.E. Hipp, "Soil electromagnetic parameters as functions of frequency, soil density, and soil moisture," *Proc. IEEE*, vol. 62, pp. 98-103, Jan. 1974.
- [24] V. Galdi, L.B. Felsen, and D.A. Castañón, "Time-domain radiation from large two-dimensional apertures via narrow-waisted Gaussian beams," submitted to *IEEE Trans. Antennas Propagat.*, Dec. 2000.

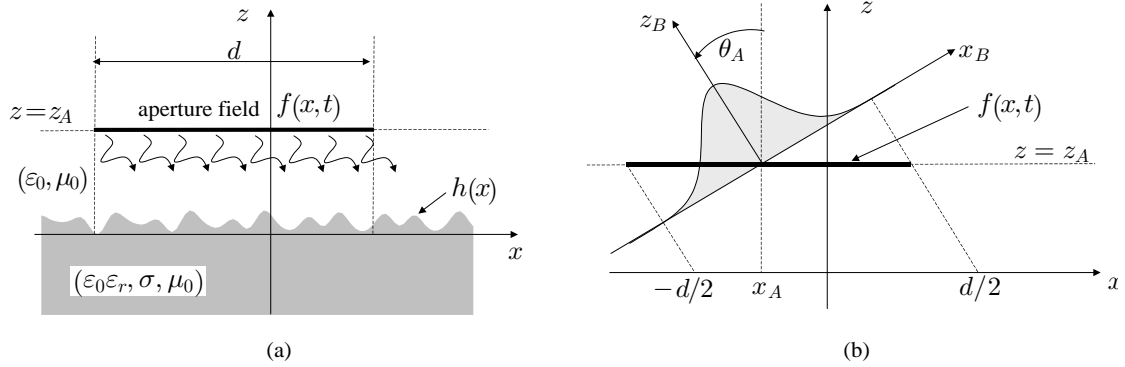


Fig. 1. (a): Problem geometry. An aperture-generated TM-polarized pulsed field impinges from free-space onto a dielectric half-space with relative permittivity ϵ_r and conductivity σ , bounded by a moderately rough interface $z = h(x)$. (b): Wide-waisted Gaussian beam excitation in (2), to be approximated as in (3).

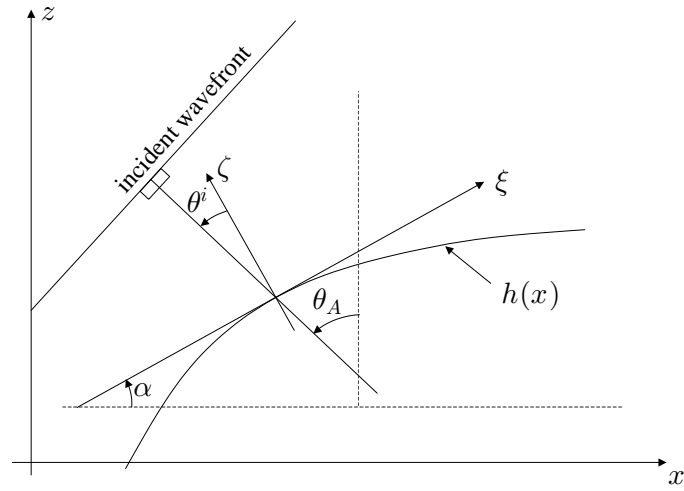


Fig. 2. Global (x, z) and local surface-based tangent-normal (ξ, ζ) coordinates. $\alpha(x) = \tan^{-1}[dh(x)/dx]$ is the local slope of the surface profile; $\theta^i(x) = \theta_A - \alpha(x)$ is the local incidence angle measured from the surface normal.

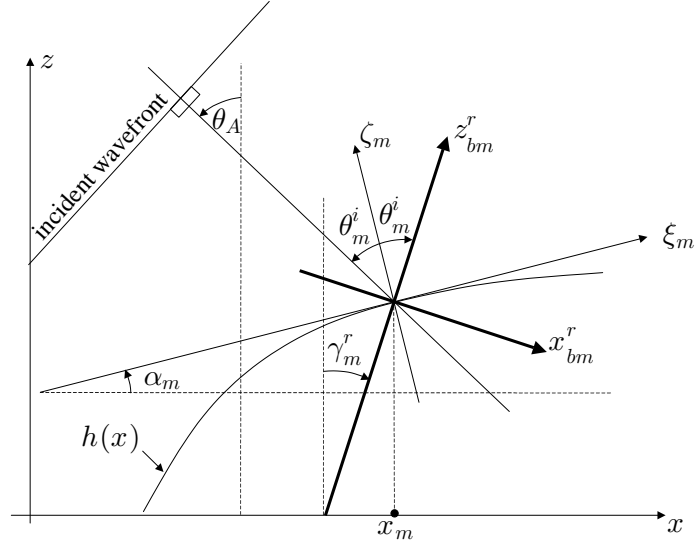


Fig. 3. Global and local coordinates for reflected beams. (x_{bm}^r, z_{bm}^r) is the m -th reflected-beam coordinate system; (ξ_m, ζ_m) is the local tangent-normal coordinate system of Fig. 2, centered at lattice points $(x_m, h(x_m))$; $\alpha_m \equiv \alpha(x_m) = \tan^{-1}[dh(x_m)/dx]$ is the local slope; $\theta_m^i = \theta_A - \alpha(x_m)$ is the local incidence angle; $\gamma_m^r = \theta_m^i - \alpha_m$ is the local reflection angle measured from the z -axis.

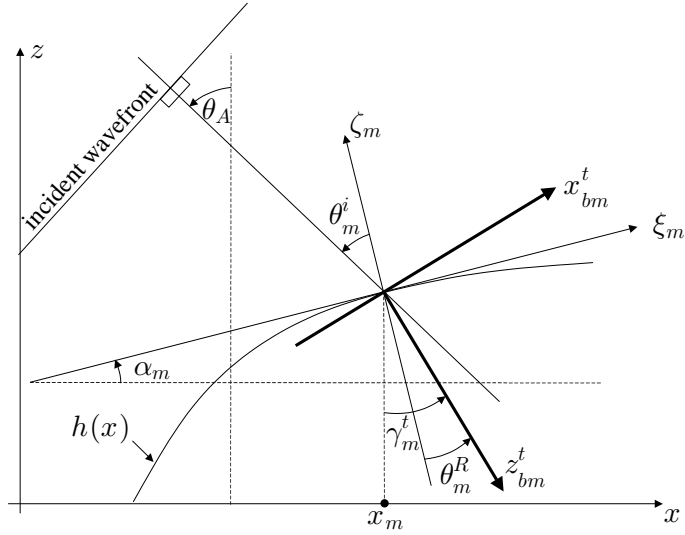


Fig. 4. As in Fig. 3, but for transmitted beams. (x_{bm}^t, z_{bm}^t) is the m -th transmitted-beam coordinate system; θ_m^R is the local refraction angle measured from the surface normal; $\gamma_m^t = \theta_m^R + \alpha_m$ is the local refraction angle measured from the z -axis. Note that for lossy materials, θ_m^R and γ_m^t are complex; for slight losses, the real (lossless) refraction angles in the figure are good approximations.

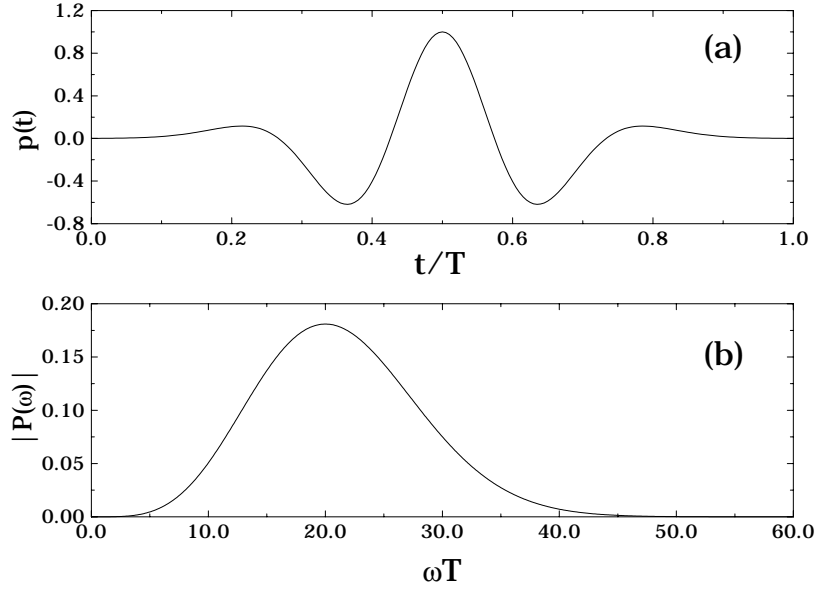


Fig. 5. Fourth-order Rayleigh pulse. (a): Temporal profile in (48); Spectrum (magnitude) in (49) ($j = 4$, $\Xi = T^4/30000$, $\varsigma = 1/\sqrt{50}$).

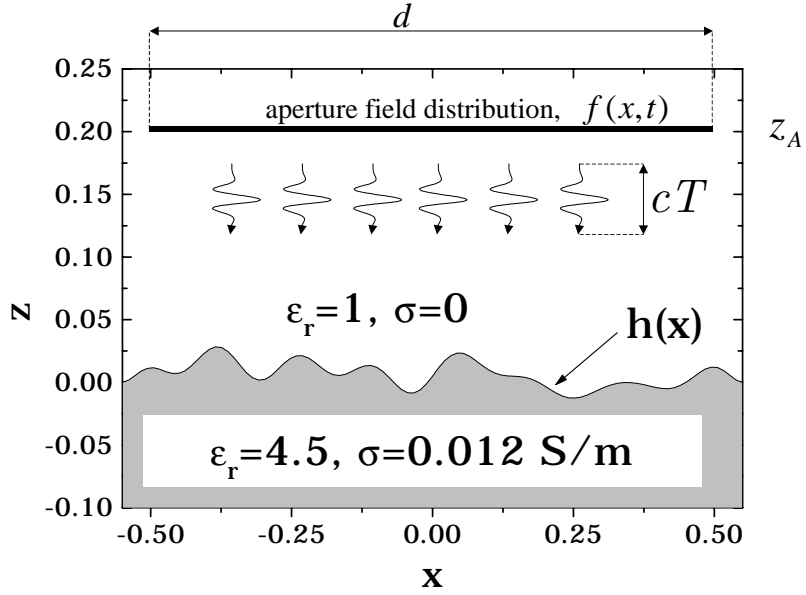


Fig. 6. Simulation geometry (cf. Fig. 1). Aperture field distribution in (2), with (65): $d = 1$, $\theta_A = 0$, $x_A = 0$, $z_A = 0.2$, and $p(t)$ as in Fig. 5 with $cT = 0.08$. Rough surface profile: maximum height $h_{max} = 0.36cT$; maximum slope $\phi_{max} = 31^\circ$; average curvature radius $\bar{r}_c = 5.9cT$. Dielectric parameters: $\epsilon_r = 4.5$, $\sigma = 0.012 \text{ S/m}$.

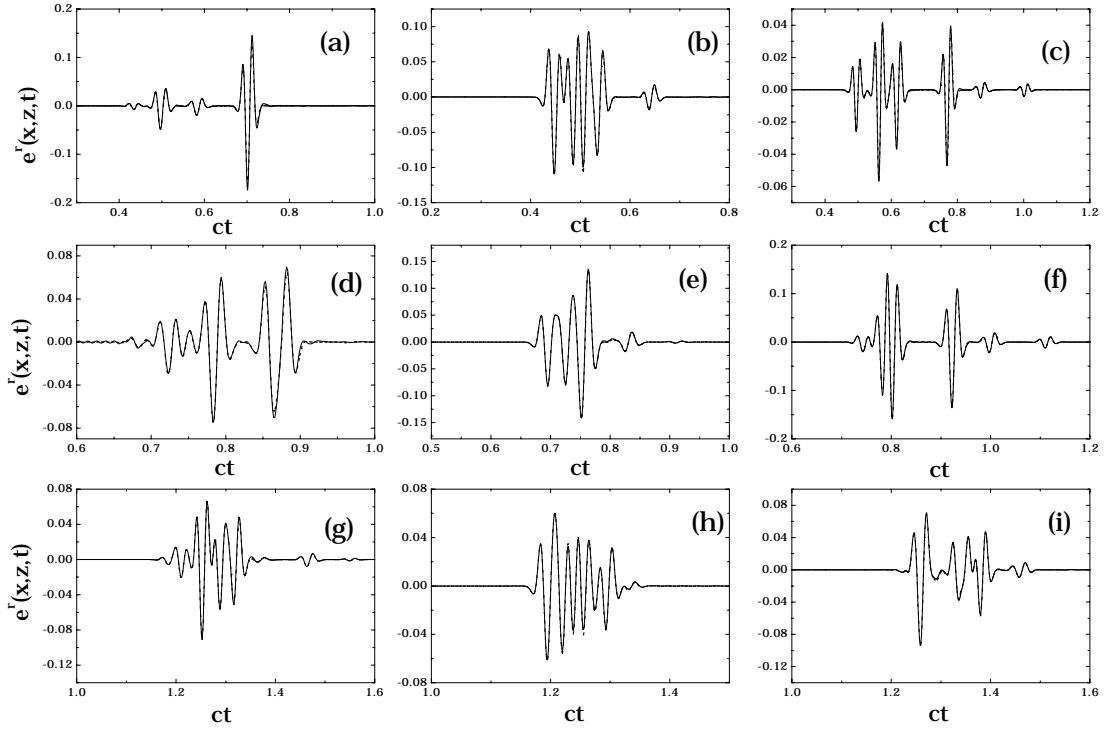


Fig. 7. Reflected field at various observation points. Parameters as in Fig. 6. (a), (b), (c): $z = 0.25$, $x = -0.4, 0, 0.4$, respectively; (d), (e), (f): $z = 0.5$, $x = -0.4, 0, 0.4$, respectively; (g), (h), (i): $z = 1$, $x = -0.4, 0, 0.4$, respectively. — Reference solution; - - - Beam synthesis ($d/L = 150$, i.e., 150 beams).

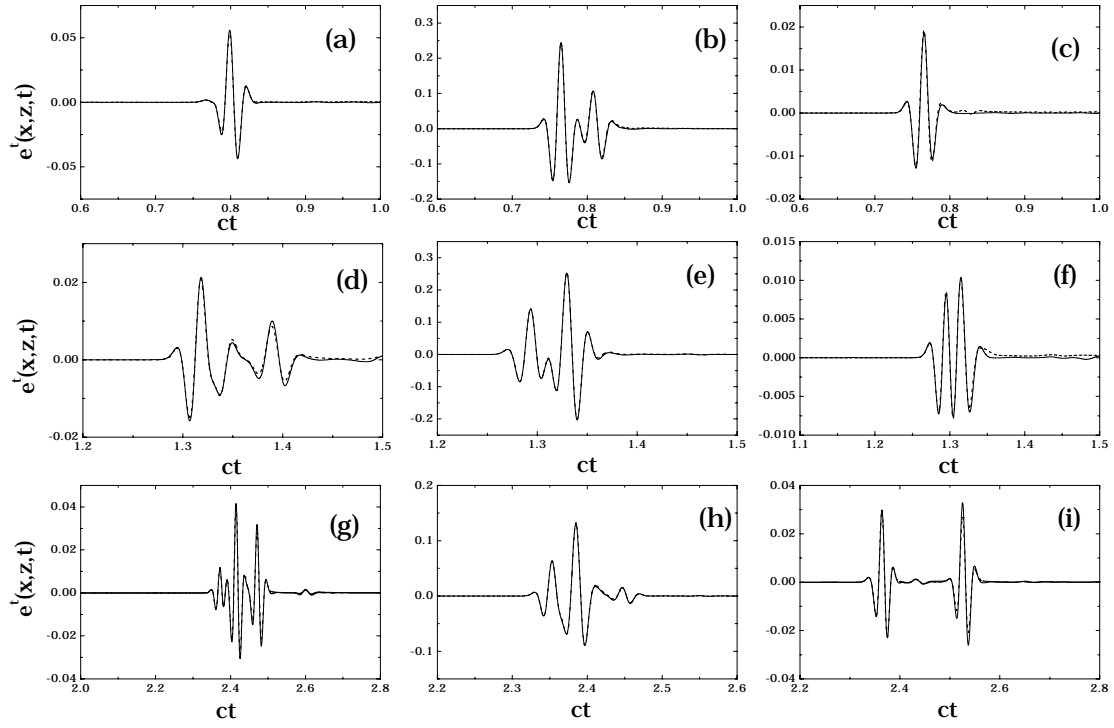


Fig. 8. Transmitted field at various observation points. Parameters as in Fig. 6. (a), (b), (c): $z = -0.25$, $x = -0.4, 0, 0.4$, respectively; (d), (e), (f): $z = -0.5$, $x = -0.4, 0, 0.4$, respectively; (g), (h), (i): $z = -1$, $x = -0.4, 0, 0.4$, respectively. — Reference solution; - - Beam synthesis ($d/L = 150$, i.e., 150 beams).

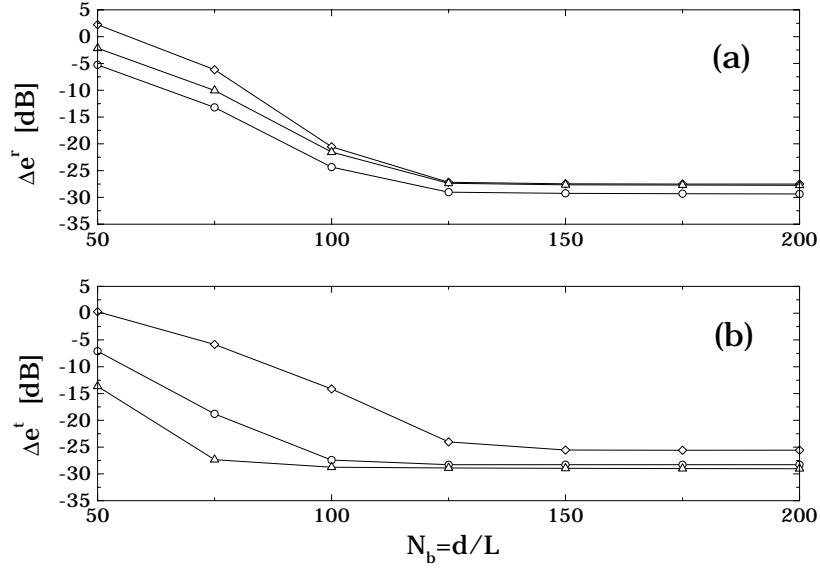


Fig. 9. r.m.s. errors $\Delta e^{r,t}$ in (66) vs. number of beams $N_b = d/L$ at various observation points. Parameters as in Fig. 6. (a): Reflected field; (b): Transmitted field; $\text{---}\diamond\text{---}$: $x = 0, z = \pm 0.25$; $\text{---}\circ\text{---}$: $x = 0, z = \pm 0.5$; $\text{---}\triangle\text{---}$: $x = 0, z = \pm 1$.

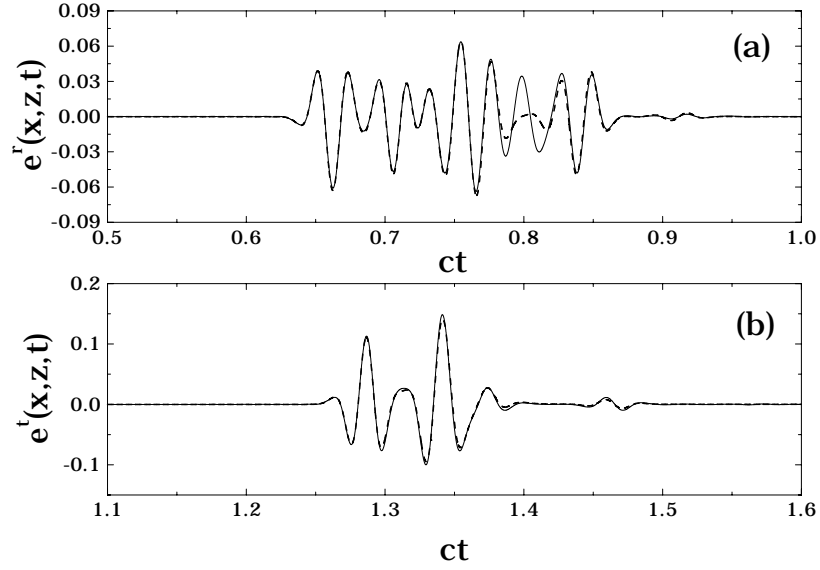


Fig. 10. Parameters as in Fig. 6, but with roughness profile scaled so that $h_{max} = 0.6cT$, $\phi_{max} = 46^\circ$, $\bar{r}_c = 4.4cT$. (a): Reflected field at $x = 0, z = 0.5$; (b): Transmitted field at $x = 0, z = -0.5$. --- Reference solution; $\text{---}\text{---}$ Beam synthesis ($d/L = 200$, i.e., 200 beams). r.m.s. errors: $\Delta e^r = -10$ dB, $\Delta e^t = -20$ dB.

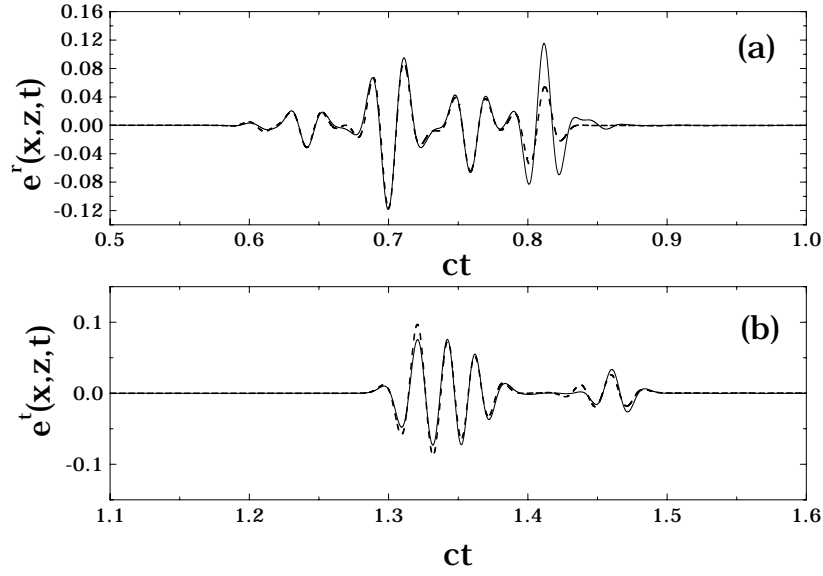


Fig. 11. As in Fig. 10, but oblique incidence with $\theta_A = 40^\circ$. — Reference solution; - - Beam synthesis ($d/L = 200$, i.e., 200 beams). r.m.s. errors: $\Delta e^r = -8$ dB, $\Delta e^t = -9$ dB.

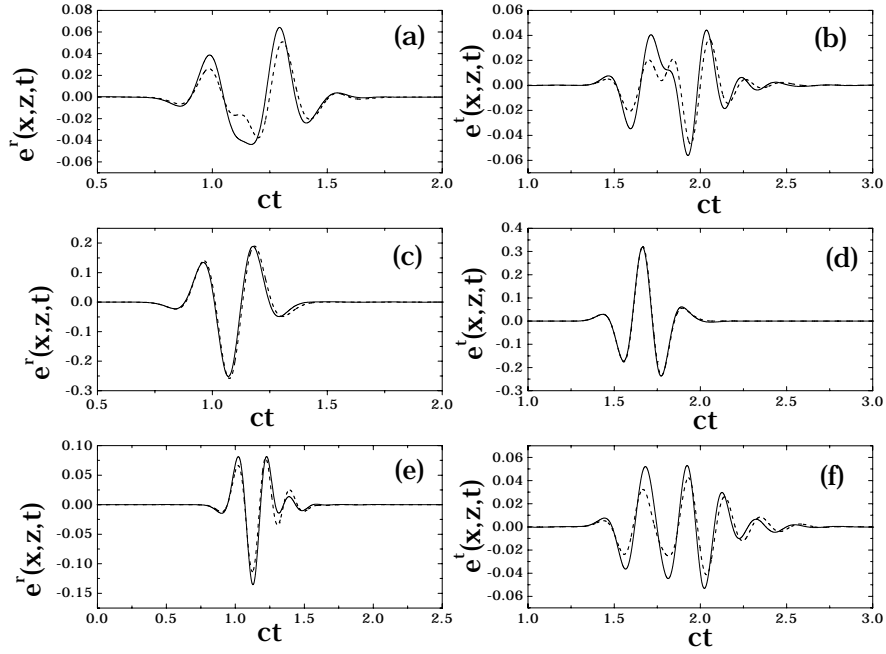


Fig. 12. Parameters as in Fig. 6, but with longer pulse, $cT = 0.8$ ($h_{max} = 0.06cT$, $\bar{r}_c = 0.44cT$). (a), (c), (e): Reflected field at $x = 0$ and $z = 0.25, 0.5, 1$, respectively (r.m.s. errors $\Delta e^r = -6, -16, -11$ dB, respectively); (b), (d), (f): Transmitted field at $x = 0$ and $z = -0.25, -0.5, -1$, respectively (r.m.s. error $\Delta e^t = -5, -25, -7$ dB, respectively). — Reference solution; - - Beam synthesis ($d/L = 30$, i.e., 30 beams).

Quasi-Ray Gaussian Beam Algorithm for Time-Harmonic Two-Dimensional Scattering by Moderately Rough Interfaces

Vincenzo Galdi, *Member, IEEE*, Leopold B. Felsen, *Life Fellow, IEEE*, and David A. Castañon, *Senior Member, IEEE*

Abstract—Gabor-based Gaussian beam (GB) algorithms, in conjunction with the complex source point (CSP) method for generating beam-like wave objects, have found application in a variety of high-frequency wave propagation and diffraction scenarios. Of special interest for efficient numerical implementation is the noncollimated narrow-waisted species of GB, which reduces the computationally intensive complex ray tracing for collimated GB propagation and scattering to quasi-real ray tracing, without the failure of strictly real ray field algorithms in caustic and other transition regions. The Gabor-based narrow-waisted CSP-GB method has been applied previously [1]–[3] to two-dimensional (2-D) propagation from extended nonfocused and focused aperture distributions through arbitrarily curved 2-D layered environments. In this 2-D study, the method is applied to aperture-excited field scattering from, and transmission through, a moderately rough interface between two dielectric media. It is shown that the algorithm produces accurate and computationally efficient solutions for this complex propagation environment, over a range of calibrated combinations of the problem parameters. One of the potential uses of the algorithm is as an efficient forward solver for inverse problems concerned with profile and object reconstruction [4].

Index Terms—Gabor lattice representations, Gaussian beams, rough surface scattering.

I. INTRODUCTION

REFLECTION from, and transmission through, a rough (irregular) boundary separating two different material media is of interest in many applications. One such application is the detection and classification of buried objects using ground-penetrating radar. This application involves above-ground transmitters and receivers. The twice-traversed air–ground interface is a principal corruptor of the signal on its way to and from the

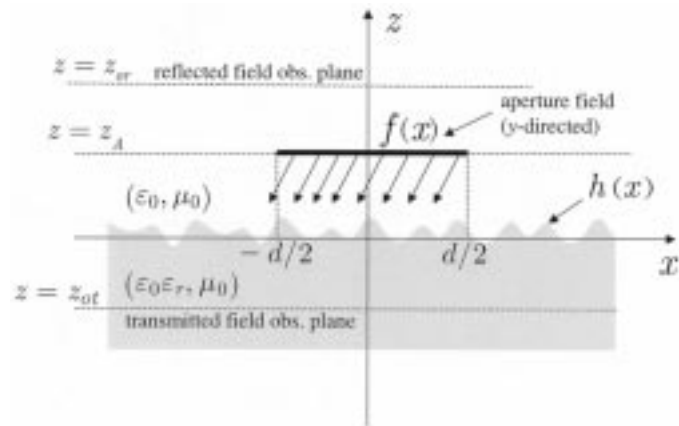


Fig. 1. Problem geometry. An incident field from an extended tapered aperture field distribution at $z = z_A$ is assumed to impinge on a moderately rough interface, described by the continuous function $h(x)$, separating free space and a dielectric half-space. Reflected and transmitted fields are observed at $z = z_{or}$ and $z = z_{ot}$, respectively.

targets of interest. To enhance the detection and classification processing, it is important to model the distortion introduced by the air–ground interface.

As a first step toward this goal, we investigate in this paper the reflection from, and transmission through, a moderately rough (coarse-scale) interface between air and a homogeneous dielectric (soil) half-space (see Fig. 1); this forward problem constitutes an electromagnetic (EM) scattering problem, and inclusion of the transmitted field anticipates the subsequent interest in computing the scattering from buried objects. In a following paper [4], we shall be concerned with the reconstruction (inverse problem) of the interface profile under the realistic constraints of spatially sampled data. A key ingredient in the reconstruction problem is a fast forward model that can relate descriptions of potential interface profiles to the measured signals at the different receiver locations.

The analytic and numerical modeling of wave scattering from rough surfaces constitutes a problem of longstanding interest (see [5]–[13] for a sparse sampling). Our objective in this paper is to develop a fast forward solver, using discretized Gabor-based, high-frequency asymptotic, narrow-waisted Gaussian beam (GB) basis fields in conjunction with the complex source point (CSP) method for generating ray-like GB wave objects [1]–[3]. The Gabor-based narrow-waisted CSP-GB method has been applied previously [1]–[3] to two-dimensional (2-D) field propagation from extended nonfocused and focused one-dimensional (1-D) planar aperture distributions through layered

Manuscript received September 6, 2000; revised February 12, 2001. This work was supported by ODDR&E under MURI Grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under Award EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. The work of L. B. Felsen was supported in part by the US-Israel Binational Science Foundation, Jerusalem, Israel, under Grant 9900448 and by Polytechnic University, Brooklyn, NY, USA.

V. Galdi is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA and the University of Sannio, Benevento, Italy (e-mail: vgaldi@bu.edu).

L. B. Felsen is with the Department of Aerospace and Mechanical Engineering and the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA and the Polytechnic University, Brooklyn, NY 11201 USA (e-mail: lfelsen@bu.edu).

D. A. Castañon is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA (e-mail: dac@bu.edu).

Publisher Item Identifier S 0018-926X(01)06371-2.

planar and cylindrical environments, and has been found to produce accurate, robust, and computationally efficient solutions over a broad range of problem parameters, provided that these do not violate the constraints imposed by the high-frequency asymptotic assumptions. Briefly, the robustness and reliability criteria are *pragmatic*, being based on the insensitivity of the result to “scramblings” (i.e., to different combinations) of the beam and lattice parameters within the constraint domain. Here, we extend this method to 1-D-aperture-excited 2-D field scattering from, and transmission through, a 1-D moderately rough arbitrary interface between two dielectric media. The resulting algorithm for radiated and scattered field synthesis by GB recombination is found to perform as in [1]–[3], but now subject to new restrictions that are stated, as necessary, in the appropriate sections of this paper.

The rest of this paper is organized as follows. In Section II, we summarize the rigorous, self-consistent Gabor-based Gaussian beam algorithm for a general aperture field distribution and the ensemble of paraxially approximated narrow-waisted CSP beams that this excitation generates [1]. Section III deals with the preliminary (canonical) problem of beam reflection from, and transmission through, a curved interface between two homogeneous dielectrics [14]. These constituents have been used previously for beam tracking through planar and curved layered dielectric configurations [2], [3]. In Section IV, we extend the algorithm to the new and more challenging problem of reflection and transmission of the plane-aperture-excited field in the presence of a moderately rough interface separating two homogeneous semi-infinite dielectric media. Extensive numerical simulations calibrate the algorithm within clearly stated constraints and highlight the role of certain critical parameters. Conclusions follow in Section V.

II. INCIDENT FIELD FROM AN EXTENDED APERTURE

A. Problem Statement

Consider a two-dimensional problem where a y -directed electric field with implicit time-harmonic dependence $\exp(-i\omega t)$ and spatial distribution $f(x)$ is assumed to occupy the aperture region $|x| \leq d/2$ at $z = z_A$ in free space, as depicted in Fig. 1

$$\mathbf{E}^{\text{inc}}(x, z_A) = f(x)\mathbf{u}_y, \quad |x| \leq d/2, \quad z = z_A. \quad (1)$$

Boldface quantities denote vectors and \mathbf{u} denotes a unit vector. The resulting (TM polarized) EM field radiated into the half-space $z < z_A$ can be expressed as a superposition of line-source generated fields (Kirchhoff integration) [15]

$$\mathbf{E}^{\text{inc}}(x, z) = \frac{i}{2} \int_{-d/2}^{d/2} f(x') \frac{\partial}{\partial z} H_0^{(1)}(k_0 R) dx' \mathbf{u}_y \quad (2)$$

$$\mathbf{H}^{\text{inc}} = \frac{1}{i\omega\mu_0} \nabla \times \mathbf{E}^{\text{inc}} = H_x^{\text{inc}} \mathbf{u}_x + H_z^{\text{inc}} \mathbf{u}_z \quad (3)$$

where $k_0 = \omega\sqrt{\epsilon_0\mu_0} = 2\pi/\lambda_0$ is the free-space wavenumber, λ_0 is the free-space wavelength, $H_0^{(1)}(\cdot)$ is the zeroth-order Hankel function of the first kind (line source Green's function), and

$$R = \sqrt{(x - x')^2 + (z - z_A)^2}. \quad (4)$$

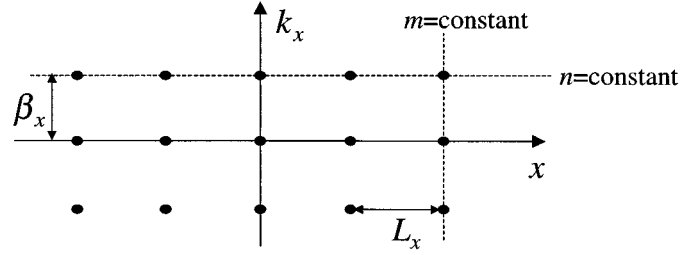


Fig. 2. Discretized phase space lattice. Spatial shift indexes m identify GB launch points at $x = mL_x$. Spectral shift indexes n identify linearly phased GB tilts at $k_x = n\beta_x$.

Because of the polarization assumed in (1), all fields can be generated from the scalar $E = E_y$ component, so that the vector notation will be dropped from here on. Alternatively, by spectral plane wave superposition, one obtains [15]

$$E^{\text{inc}}(x, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(k_x) \exp[i(k_x x + k_z z)] dk_x \quad (5)$$

where

$$\hat{f}(k_x) = \int_{-\infty}^{\infty} f(x) \exp(-ik_x x) dx \quad (6)$$

is the Fourier spectrum of $f(x)$, k_x is the x -domain wavenumber, and

$$k_z = \sqrt{k_0^2 - k_x^2}, \quad \text{Im}(k_z) \geq 0 \quad (7)$$

is the longitudinal (z -domain) wavenumber.

B. Gabor Beam Discretization

1) *Aperture Field*: The aperture field $f(x)$ is to be parameterized in terms of Gaussian basis functions via the rigorous self-consistent Gabor series representation [1], [15]–[19]

$$f(x) = \sum_{m,n=-\infty}^{\infty} A_{mn} w(x - mL_x) \exp(in\beta_x x) \quad (8)$$

where $w(x)$ represents the normalized Gaussian window function

$$w(x) = \left(\frac{\sqrt{2}}{L_x} \right)^{1/2} \exp[-\pi(x/L_x)^2], \quad \int_{-\infty}^{\infty} w^2(x) dx = 1. \quad (9)$$

This representation places the aperture distribution on a discretized (x, k_x) phase space lattice (see Fig. 2), with spatial and spectral shifts tagged by the indexes m and n , respectively. Spatial and spectral periods are related by the self-consistency relation (configuration-spectrum tradeoff) [16], [17]

$$L_x \beta_x = 2\pi. \quad (10)$$

As observed in [15] and [19], the Gaussian window provides the best occupation of the phase space. The expansion coefficients A_{mn} in (8) can be computed by introducing an auxiliary function $\gamma(x)$ defined through the *biorthogonality* condition [16], [17]

$$\int_{-\infty}^{\infty} w(x) \gamma^*(x - mL_x) \exp(-in\beta_x x) dx = \delta_m \delta_n \quad (11)$$

where $*$ denotes the complex conjugate and $\delta_q = 1$ for $q = 0$ and $\delta_q = 0$ for $q \neq 0$. The expansion coefficients can be written as [16], [17]

$$A_{mn} = \int_{-\infty}^{\infty} f(x) \gamma^*(x - mL_x) \exp(-in\beta_x x) dx. \quad (12)$$

For Gaussian windows, the biorthogonal function can be computed explicitly [16], [17] but will not be required in what follows later on.

2) *Radiated Field*: The initial distribution surrounding each lattice point in Fig. 2 generates a Gaussian beam that is launched from m -indexed locations and tilted according to n -indexed locations. The radiated incident field in the half-space $z < z_A$ [see (2)] therefore admits via (8) a similar discretized representation (we follow the notation in [1]; however, there are some sign changes with respect to [1] since here we assume propagation into the *negative* half-space $z < z_A$)

$$E^{\text{inc}}(x, z) = \sum_{m,n=-\infty}^{\infty} A_{mn} B_{mn}(x, z) \quad (13)$$

where the beam functions $B_{mn}(x, z)$ are expressed by Gabor-weighted line-source superposition

$$B_{mn}(x, z) = \frac{i}{2} \int_{-d/2}^{d/2} w(x' - mL_x) \cdot \exp(in\beta_x x') \frac{\partial}{\partial z} H_0^{(1)}(k_0 R) dx' \quad (14)$$

with R being defined in (4). By saddle point methods, the integral in (14) (or its spectral counterpart) can be evaluated asymptotically in the beam *paraxial far zone*, yielding the following *complex source point approximation* [2], [3]:

$$B_{mn}(x, z) \sim ik_0 2^{5/4} \left(\frac{L_x}{8\pi k_0 \tilde{R}_{mn}} \right)^{1/2} \cdot \exp \left\{ i \left[k_0 \left(\tilde{R}_{mn} + ib \right) + \pi/4 \right] \right\} \frac{(z - \tilde{z}'_{mn})}{\tilde{R}_{mn}} \quad (15)$$

with \tilde{R}_{mn} representing the *complex distance* between the observer at $P = (x, z)$ and the CSP

$$\tilde{P}'_{mn} = (\tilde{x}'_{mn}, \tilde{z}'_{mn}) = (mL_x + ib \sin \theta_n, z_A - ib \cos \theta_n) \quad (16)$$

$$\tilde{R}_{mn} = \overline{P \tilde{P}'_{mn}} = \sqrt{(x - \tilde{x}'_{mn})^2 + (z - \tilde{z}'_{mn})^2}. \quad (17)$$

In accord with the radiation condition, the square root in (17) is defined by $\text{Re}(\tilde{R}_{mn}) \geq 0$. Here and henceforth, the tilde \sim identifies CSP-generated complex quantities. The displacement parameter b (equal to the Fresnel length) is related to the beam lattice period L_x and the beam axis angle θ_n via [2]

$$b = (L_x \cos \theta_n)^2 / \lambda_0, \quad \theta_n = \sin^{-1}(n\lambda_0 / L_x). \quad (18)$$

Equation (15) is valid in the paraxial far-zone of each beam, $|\tilde{R}_{mn}| \gg b$. As the tilt index n increases, the beam tilt angle θ_n can become complex ($|n| > L_x / \lambda_0$), whence the corresponding beams become *evanescent*.

C. Narrow-Waisted Beams

In the following, we shall focus on nontilted ($n = 0$) *narrow-waisted* beams ($L_x \lesssim \lambda_0 \ll d$), which, as demonstrated in [1]–[3], have several attractive features. First, the Gabor coefficients A_{mn} in (13) can be effectively estimated by *sampling* the aperture field distribution at the lattice points $x = mL_x$, thus avoiding the time-consuming integration in (12) [1]

$$A_{mn} \approx \begin{cases} (L_x / \sqrt{2})^{1/2} f(mL_x), & n = 0 \\ 0, & n \neq 0. \end{cases} \quad (19)$$

Under this approximation, the tilted ($n \neq 0$) beams in the Gabor expansion, which here generate evanescent “far fields” [complex θ_n in (18)], are ignored. Second, for narrow-waisted beams, the CSP paraxial far-zone approximation (15) can be invoked at moderate distance; thus their *superposition* can furnish accurate results even in the *near zone* of the *aperture*. Third, as has been shown in [2] and [3], interaction of narrow-waisted beams with an environment can be implemented effectively by tracking the *complex ray* fields and recombining them at the observer, subject to constraints that are elucidated in the examples below.

D. Illustrative Examples

A thorough analysis of the accuracy and computational features of the *narrow-waisted* beam algorithm can be found in [1]. Here, we merely summarize relevant results. Referring to Fig. 1, we consider a linearly phased cosine-tapered aperture field distribution

$$f(x) = \begin{cases} \cos(\pi x / d) \exp(ik_0 \nu x), & |x| \leq d/2 \\ 0, & |x| > d/2 \end{cases} \quad (20)$$

where d is the aperture width and $\nu = \sin^{-1}(\varphi)$, with φ denoting the tilt angle of the main radiation lobe with respect to the z -axis. A special case (nonphased cosine, $\nu = 0$) is used in the rough interface simulation in Section IV-B. Fig. 3 shows the normalized magnitude of the exact Gabor coefficients [computed through numerical integration of (12)] as a function of the shift and tilt indexes (m, n) for narrow-waisted beams ($L_x = 0.0125d$). As one can see, the essential contribution comes from the nontilted beams, whose coefficient distribution closely matches the aperture field profile.

The beam-computed and reference near-zone *radiated* fields, which are obtained from (13) [with (15) and (19)], and via numerical evaluation of the Kirchhoff integral in (2), respectively, are shown in Fig. 4(a) for $\nu = 0.5$. Although based on the paraxial far-zone approximation in (15) for the beam propagators, the accuracy is quite good even in the near zone of the aperture (80 beams were used in this simulation). In the absence of phasing ($\nu = 0$), considerably coarser sampling is adequate. As shown in Fig. 4(b), even a beam lattice period $L_x = 0.1d$ (i.e., ten beams) yields accurate synthesis. Coarse sampling also works for linearly phased apertures if propagation-matched *tilted* basis beams are used [20]. The stated number of beams in these simulations was arrived at via the pragmatic “scramblings” test, i.e., when the result remains insensitive to variations in the beam/lattice combinations.

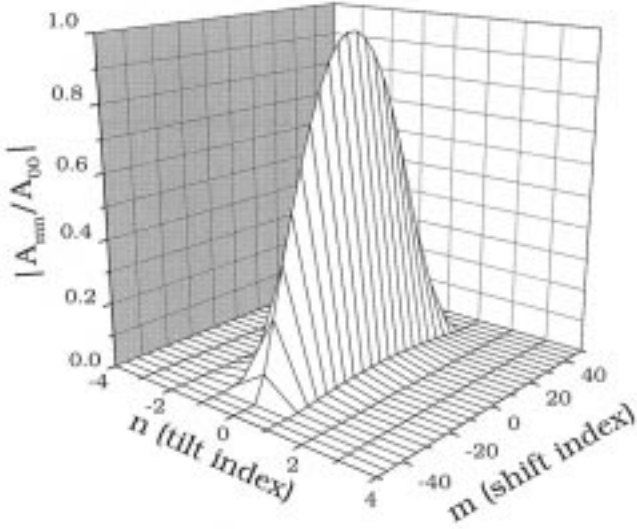


Fig. 3. Linearly phased cosine-tapered aperture field in (20) ($d = 10\lambda_0$, $\nu = 0.5$). Exact normalized Gabor coefficient magnitudes $|A_{mn}/A_{00}|$ in (12) evaluated numerically ($L_x = 0.0125d = 0.125\lambda_0$). Tilted ($n \neq 0$) beams are evanescent.

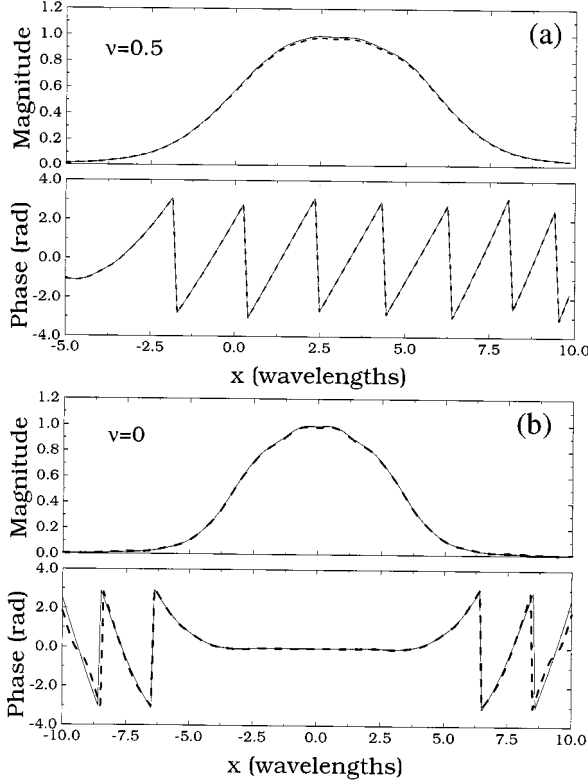


Fig. 4. Cosine-tapered aperture field in (20) ($d = 10\lambda_0$): Radiated near-zone field with $z_A = 10\lambda_0$, $z_{obs} = 5\lambda_0$. (a) $\nu = 0.5$, $L_x = 0.0125d = 0.125\lambda_0$ (i.e., 80 beams). (b) $\nu = 0$, $L_x = 0.1d = \lambda_0$ (i.e., ten beams). (—) Reference solution [Kirchhoff integration in (2)]. (---) Beam-computed from (13), with (15) and (19).

III. REFLECTION FROM AND TRANSMISSIONS THROUGH A SMOOTHLY CURVED DIELECTRIC INTERFACE: CANONICAL PROBLEM

A. Problem Strategy

Before considering the interaction of the beam-based incident field in Section II with the rough surface profile sketched in

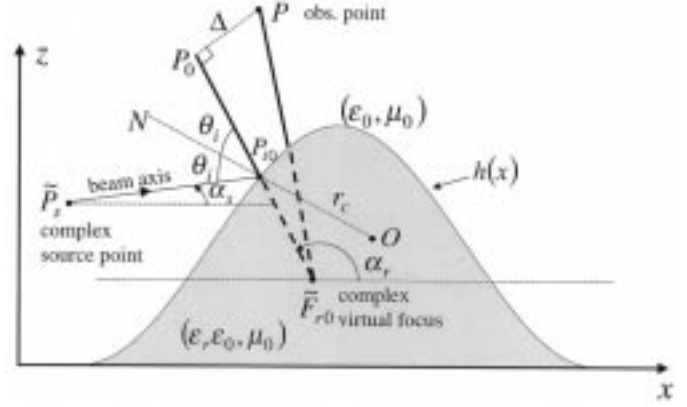


Fig. 5. Beam reflection from a curved dielectric interface. α_s = real departure angle of incident beam axis with respect to the x -axis; α_r = real departure angle of reflected beam axis with respect to the x -axis; and θ_i = real incidence/reflection angle with respect to the surface normal N at P_{i0} ; r_c = surface radius of curvature at P_{i0} .

Fig. 1, we treat the canonical problem of interaction with a smoothly curved portion of that profile as shown in Figs. 5 and 6.

The beam-based synthesis of the reflected and transmitted fields can be constructed by propagating each of the incident narrow-waisted basis beams in Sections II-B and -C through the environment and recombining thereafter. Because the asymptotic basis beams are parameterized in terms of a source point in the complex coordinate space, they propagate in complex space along *complex* ray trajectories, which yield only a *single* physical point at the real space observer. This renders the field evaluation computation intensive. However, for narrow-waisted beams, one may utilize a paraxial approximation scheme, which leads to *almost real* ray tracing [3], [14] that is only slightly different from (and even computationally cheaper than) standard ray tracing from the Kirchhoff integration (2) in *real* configuration space.

B. Reflected Field

The problem geometry is illustrated in Fig. 5. An incident Gaussian beam is generated by a CSP at

$$\tilde{P}_s = (x_s + ib \cos \alpha_s, z_s + ib \sin \alpha_s) \quad (21)$$

with α_s being the *real* departure angle of the incident beam axis with respect to the x -axis. As shown in [14], for electrically large and smooth scatterers, and when the observation point $P = (x, z)$ lies in the paraxial region of the reflected beam [$\Delta \ll (|\tilde{F}_{r0}P_0|^2 + b^2)^{1/2}$ in Fig. 5], the reflected field can be approximated in terms of the on-axis field (at P_0) and a complex phase correction. Denoting the on-axis parameters by the subscript zero, one finds for the y -directed electric field [14]

$$\begin{aligned} E^{refl}(P) &\approx E^{refl}(P_0)|_{\tilde{P}_i=P_{i0}} \exp(i k_0 \tilde{\delta}_p) \\ &= \Gamma E^{inc}(P_{i0}) \left(\frac{\tilde{f}_{r0}}{\tilde{R}_0} \right)^{1/2} \exp[i k_0 (L_{r0} + \tilde{\delta}_p)] \end{aligned} \quad (22)$$

where (see Fig. 5)

$$\tilde{R}_0 = \overline{\tilde{F}_{r0}P_0}, \quad L_{r0} = \overline{P_{i0}P_0} \quad (23)$$

and $\tilde{\delta}_p$ is the complex phase correction

$$\tilde{\delta}_p = \tilde{R} - \tilde{R}_0, \quad \tilde{R} = \overline{\tilde{F}_{r0}P}. \quad (24)$$

Moreover, $\tilde{F}_{r0} = (\tilde{x}_{fr0}, \tilde{z}_{fr0})$ is the *complex* virtual focus obtained via analytic continuation of the standard ray-optical formulas (see [21, p. 169]), but approximating the *complex* incidence point \tilde{P}_i by the *real* beam-axis incidence point $P_{i0} = (x_{i0}, z_{i0})$

$$\tilde{x}_{fr0} = x_{i0} - \tilde{f}_{r0} \cos \alpha_r, \quad \tilde{z}_{fr0} = z_{i0} - \tilde{f}_{r0} \sin \alpha_r \quad (25)$$

$$\tilde{f}_{r0} = \frac{\tilde{L}_{i0} r_c \cos \theta_i}{2\tilde{L}_{i0} + r_c \cos \theta_i}, \quad \tilde{L}_{i0} = \overline{\tilde{P}_s P_{i0}}. \quad (26)$$

Here, α_r is the real departure angle of the reflected beam axis with respect to the x -axis, r_c is the curvature radius at P_{i0}

$$r_c = -\frac{[1 + (h'(x_{i0}))^2]^{3/2}}{h''(x_{i0})}, \quad ' \equiv d/dx \quad (27)$$

and Γ is the TM plane-wave reflection coefficient

$$\Gamma = \frac{\cos \theta_i - \sqrt{\epsilon_r - \sin^2 \theta_i}}{\cos \theta_i + \sqrt{\epsilon_r - \sin^2 \theta_i}} \quad (28)$$

where ϵ_r is the dielectric relative permittivity and θ_i is the real incidence angle with respect to the surface normal N at P_{i0} .

As shown in [14], this corresponds to tracing a ray along a *complex* trajectory from the CSP at \tilde{P}_s to the intersection of the *real* beam axis with the *real* surface; from there, the path to the observer proceeds entirely in *real* configuration space, along the beam axis. Also presented in [14] are further corrections, obtained by expanding the analytic continuation of the reflection coefficient Γ in (28) and of the divergence coefficient

$$\tilde{D}_r = \left[\frac{(\overline{\tilde{P}_s \tilde{P}_i} + \overline{\tilde{P}_i P}) r_c \cos \tilde{\theta}_i}{2\overline{\tilde{P}_s \tilde{P}_i} \overline{\tilde{P}_i P} + (\overline{\tilde{P}_s \tilde{P}_i} + \overline{\tilde{P}_i P}) r_c \cos \tilde{\theta}_i} \right]^{1/2} \quad (29)$$

$\tilde{\theta}_i \equiv \theta_i|_{P_{i0}=\tilde{P}_i}$

in Taylor series about their on-axis values. Actually, these corrections are more cumbersome to generate; we obtained acceptable results by using only the phase correction in (24). Multiple reflections can be incorporated by iterating (22) whereby [with proper definition of the square root in (17)] the complex focus determined at each iteration becomes the phase reference for the next iteration (see Appendix A). Again, apart from the complex ray connecting the CSP to the first real incidence point, the multihop path to the observer proceeds entirely in *real* configuration space along the beam axes, and the phase correction is applied only to the last beam segment that reaches the observer.

C. Transmitted Field

The same considerations applied to the transmitted field (see Fig. 6) yield [14]

$$\begin{aligned} E^{tr}(P) &\approx E^{tr}(P_0) \big|_{\tilde{P}_i=P_{i0}} \exp(ik\tilde{\delta}_p) \\ &= TE^{\text{inc}}(P_{i0}) \left(\frac{\tilde{f}_{t0}}{\tilde{R}_0} \right)^{1/2} \exp \left[ik_0 (L_{t0} + \sqrt{\epsilon_r} \tilde{\delta}_p) \right] \end{aligned} \quad (30)$$

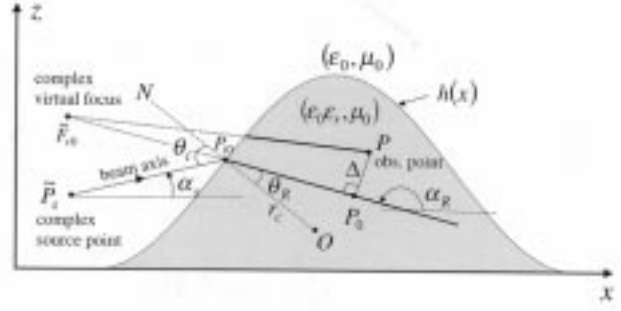


Fig. 6. Beam transmission through a curved dielectric interface. α_R and θ_R are the refracted (transmitted) real departure angles with respect to the x -axis and the surface normal N at P_{i0} , respectively.

where $k = k_0 \sqrt{\epsilon_r}$ and

$$\tilde{R}_0 = \overline{\tilde{F}_{t0} P_0}, \quad L_{t0} = \overline{P_{i0} P_0} \quad (31)$$

$$\tilde{\delta}_p = \tilde{R} - \tilde{R}_0, \quad \tilde{R} = \overline{\tilde{F}_{t0} P} \quad (32)$$

$$\tilde{F}_{t0} = (x_{i0} + \tilde{f}_{t0} \cos \alpha_R, z_{i0} + \tilde{f}_{t0} \sin \alpha_R) \quad (33)$$

$$\tilde{f}_{t0} = \frac{\tilde{L}_{i0} r_c \cos \theta_R}{(\Omega - 1)\tilde{L}_{i0} + \Omega r_c \cos \theta_i}, \quad \Omega = \frac{\cos \theta_i}{\sqrt{\epsilon_r} \cos \theta_R} \quad (34)$$

$$T = 1 + \Gamma \quad (35)$$

with α_R denoting the real departure angle of the transmitted beam axis with respect to the x -axis and θ_R denoting the refraction angle, with respect to the surface normal N at P_{i0} , according to Snell's law

$$\sin \theta_i = \sqrt{\epsilon_r} \sin \theta_R. \quad (36)$$

As before, this approximation corresponds to tracing a complex ray from the CSP to the real incidence point and then a real ray to the observer [14]. Note that in the presence of a slightly lossy dielectric as in Fig. 8, the refraction angle θ_R becomes complex. In that case, we still use a real refracted ray, propagating along the direction $\text{Re}(\theta_R)$. Multiple reflections/transmissions can be handled by iterating and combining (22) and (30), following the guidelines given in Section III-B.

IV. REFLECTION FROM AND TRANSMISSION THROUGH A MODERATELY ROUGH DIELECTRIC INTERFACE

We now address the “real problem” geometry in Fig. 1, using the building blocks of Sections II-B, II-C, and III.

A. The Multiply Reflected/Transmitted Beam Algorithm

The main steps of the proposed algorithm can be summarized as follows.

- 1) *Aperture Field Discretization.* The narrow-waisted CSP algorithm (15)–(19) in Sections II-B and -C is applied to the given aperture field distribution and yields the Gabor-weighted beam amplitudes.
- 2) *Beam-Axes Tracing.* The beam axes follow real-ray trajectories in real configuration space. For each beam

launched at the aperture plane, the axis trajectories after intersection at an interface are governed by Snell's laws of reflection and transmission (see Figs. 5 and 6). The hierarchy of possible multiple (reflection/transmission) bifurcations at each encounter with the *rough* interface can be organized formally through use of a binary-tree data structure [22], whose *nodes* contain the relevant data (i.e., incidence point, departure angle, etc.). However, in our applications here, we shall rely primarily on the first, and possibly also the second, bifurcation.

- 3) *Beam Parameter Computation.* Having determined the beam axis topology, the other relevant parameters for each beam segment (e.g., reflection/transmission coefficients, complex foci, phase lag) are computed recursively, starting from the initial complex source point and progressively scanning the binary tree (see Appendix A).
- 4) *Beam Field Computation.* Each reflected/transmitted Gabor-weighted beam field contribution at the observer is computed via the quasiray paraxial approximations (22)–(28) and (30)–(36) described in Section III. As already stated, in the presence of multiple interactions, the complex phase correction is applied only to the last beam segment that reaches the observer (see Appendix A).
- 5) *Beam Summation for Total Field.* The individual reflected/transmitted beam contributions are recombined to yield the total field at the observer.

In principle, the above multiple interaction algorithm is able to account for possible evanescent contributions (total reflection). We have not explored this aspect so far and have, in fact, restricted the problem conditions in the numerical tests of Section IV-B so that total reflection does not occur. Also excluded are near-grazing incidence contributions, which require more sophisticated propagation models than those in Section III. Note also that in [2], [3], the narrow-waisted beam algorithm is applied to propagation through planar and circular cylindrical dielectric layers where the beam tracing can be performed analytically; in our problem, due to the irregular form of the interface profile $h(x)$ (see Fig. 1), a numerical procedure is required.

B. Example Problems and Numerical Results

In the problem geometry of Fig. 1, the TM-polarized field generated by the nonphased cosine-tapered aperture field distribution in (20) (with $\nu = 0$) is assumed to impinge from free space onto a moderately rough interface described by the continuous function $h(x)$, which in the examples below is parameterized by a quartic-spline [23]

$$h(x) = \sum_{n=-4}^{N_h-1} c_n s_4(x - x_n), \quad x_{\min} \leq x \leq x_{\max} \quad (37)$$

$$x_n = x_{\min} + n\Delta x, \quad \Delta x = (x_{\max} - x_{\min})/N_h \quad (38)$$

with $s_4(\cdot)$ representing the standard quartic B-spline basis function (Fig. 7) [23]. Because of this parameterization, most of the beam tracing computational tasks (solution of fourth-degree algebraic equations) can be performed *analytically*. The aperture width is adjusted so as to illuminate most of the re-

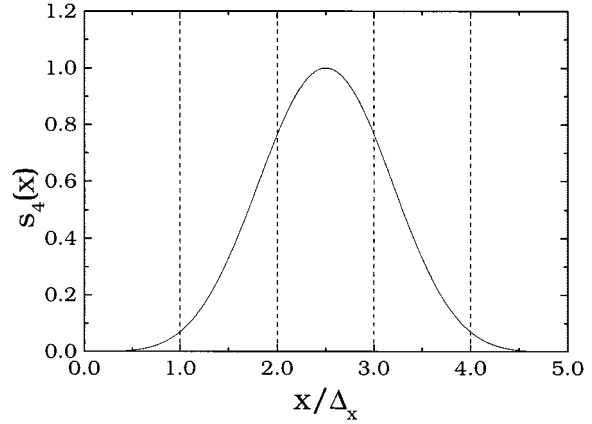


Fig. 7. Quartic B-spline basis function $s_4(x)$.

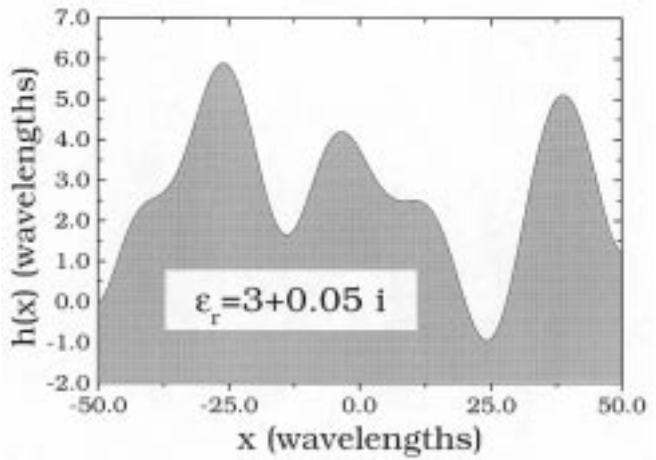


Fig. 8. Rough surface geometry and parameters. Relative permittivity: $\epsilon_r = 3 + i0.05$; maximum slope: 34° ; minimum curvature radius: $4.2\lambda_0$; maximum height: $5.88\lambda_0$. The chosen permittivity models a class of realistic soil conditions.

gion of interest, with edge effects deemphasized due to the aperture profile tapering. Since we intend to apply the beam method to detection and identification of buried objects, the dielectric half-space parameters have been chosen to simulate soil conditions [24]. The reflected and transmitted fields for this configuration have been computed via the beam algorithm detailed in Sections II and III and compared with a reference solution obtained by the Leviatan–Boag multifilament current method [25]. Although this method was introduced in [25] for cylindrical scatterers, its application to the rough surface geometry is justified in Appendix B. Since beam superposition is a high-frequency (HF) method, the rough surface parameters were adjusted so as to guarantee a minimum curvature radius of about 4.2 free-space wavelengths and a maximum slope of about 34° (see Fig. 8). This places the wave dynamics well within the range for HF asymptotic localization. Note that the dielectric is slightly lossy so that the remarks in Section III-C apply. It should be mentioned that we have previously applied the algorithm to reflection from a perfectly conducting sinusoidal profile and have calibrated its range of validity there [26]. The results from this test case have guided our choice of problem parameters here.

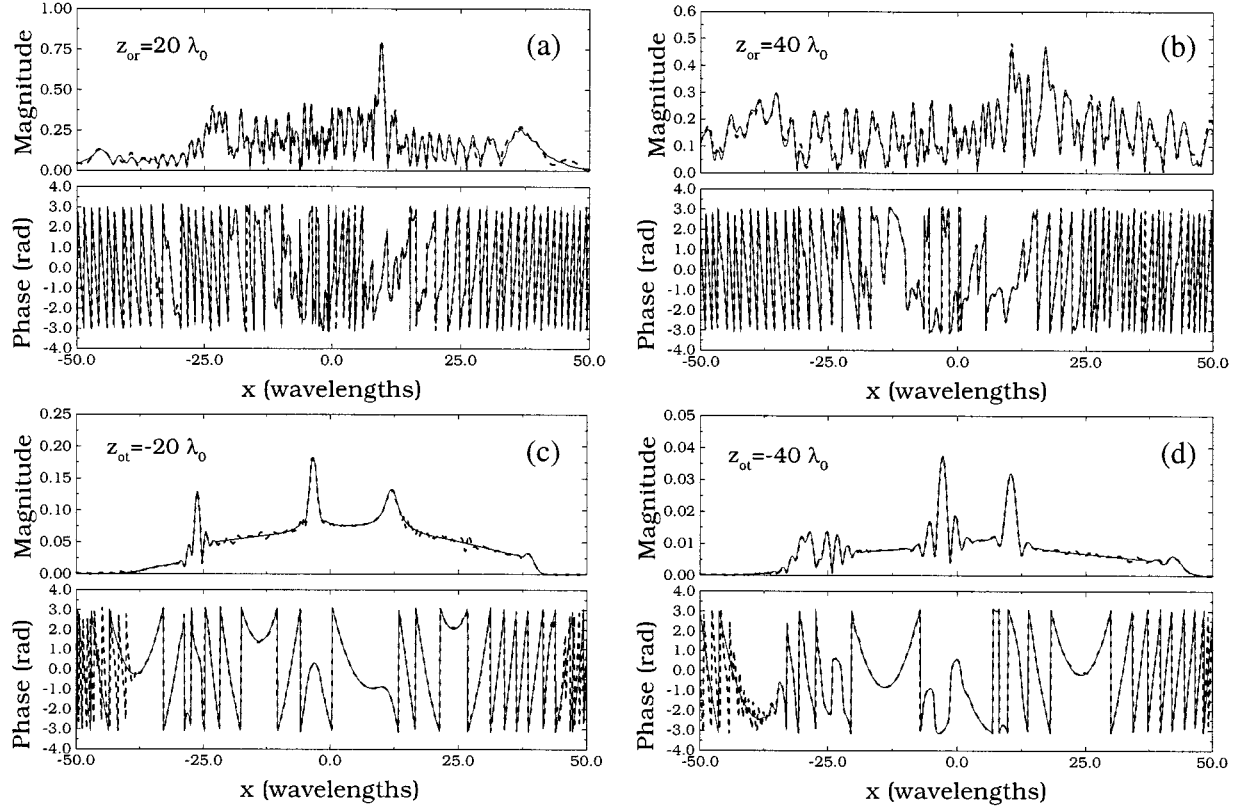


Fig. 9. Beam-computed reflected/transmitted fields and Leviatan-Boag reference solution in various observation planes. Parameters as in Fig. 8. Aperture width: $d = 80\lambda_0$; aperture height: $z_A = 6.4\lambda_0$ (i.e., $0.52\lambda_0$ above maximum profile height); beam lattice period: $L_x = 0.0125d = \lambda_0$ (i.e., 80 beams). (a) Reflected field at $z_{or} = 20\lambda_0$. (b) Reflected field at $z_{or} = 40\lambda_0$. (c) Transmitted field at $z_{ot} = -20\lambda_0$. (d) Transmitted field at $z_{ot} = -40\lambda_0$. (—) Reference solution and (---) beams.

In Fig. 9, for a fixed beam-lattice configuration and the problem parameters in Fig. 8, the beam-computed reflected and transmitted fields at various observation planes are compared with the reference solution. Here, the aperture is located $0.52\lambda_0$ away from the nearest portion of the interface profile, thereby ensuring that the *individual* narrow-waisted paraxial GBs are collimated sufficiently to have them interact *locally* with the irregular surface. We found that in order to obtain robust and accurate predictions, the aperture distance to the profile should be no larger than a wavelength. As in [26], good agreement is verified for the beam/lattice parameters listed in the figure captions. The accuracy of the scattered/transmitted field synthesis improves for *greater observation distance* because of the *collective* behavior of the beams in the far-zone paraxial approximation. The agreement is satisfactory even at moderate observation distances, despite some slight local deviations. As in [1]–[3], numerical insensitivity to beam/lattice parameter scramblings is taken as a pragmatic performance criterion for robustness of the algorithm (see Fig. 10). This feature permits *a priori* accuracy assessments when reference solutions are not available; note that the pragmatic scrambling criterion used here aims at an *adequate* number of beams, without any attempt to minimize that number. For the nonphased aperture example, we found the best tradeoff between accuracy and computational cost for $L_x \sim \lambda_0$. A finer sampling may, however, be required for phased aperture field distributions. It should be emphasized that the accuracy deteriorates with increasing distance of the aperture from the nearest point on the surface profile. As noted

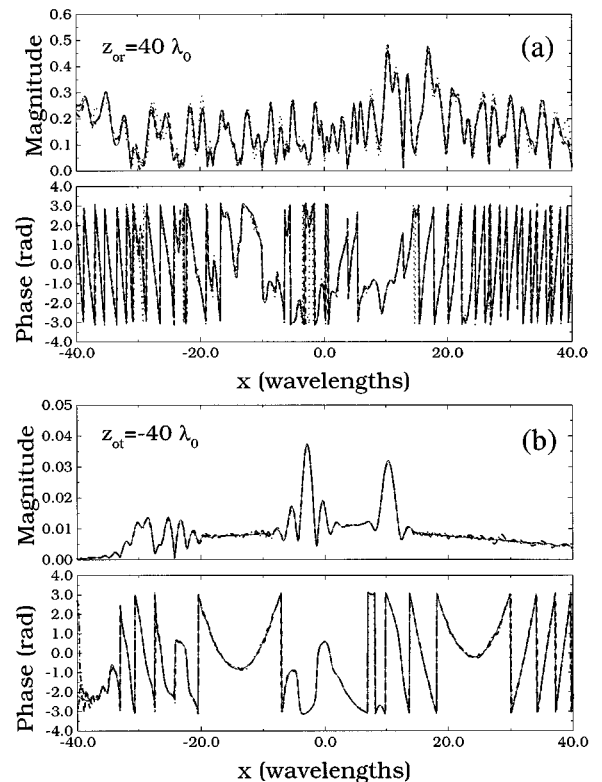


Fig. 10. (a) As in Fig. 9(b), but with various beam lattice periods. (b) As in Fig. 9(d), but with various beam lattice periods. (—) Reference solution; (·····) 134 beams ($L_x = 0.0075d$); (---) 80 beams ($L_x = 0.0125d$); and (-·-·-) 58 beams ($L_x = 0.0175d$).

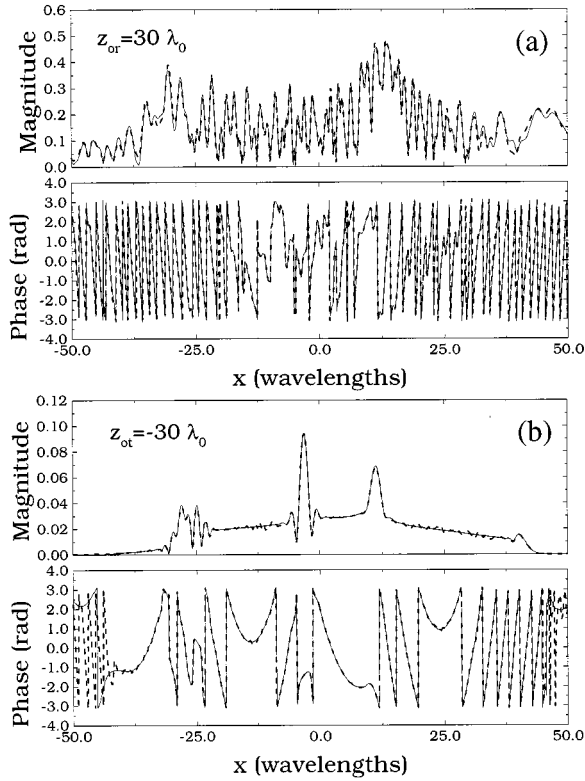


Fig. 11. As in Fig. 9, but with aperture plane at $z_A = 35.9\lambda_0$. Multistep beam algorithm: i) Narrow-waisted beam discretization at $z = z_A$ with $L_x = 0.01d$ (i.e., 100 beams). ii) Beam-computed incident field projection onto a virtual aperture at $z_V = 6.4\lambda_0$ as in Fig. 9, to restore the local sampling capability of the surface profile by the individual basis beams. iii) Narrow-waisted beam decomposition at $z = z_V$ with $L_x = 0.0125d = \lambda_0$ (i.e., 80 beams) as in Fig. 9. iv) Beam-tracing paraxial-approximation scheme. (a) Reflected field at $z_{or} = 30\lambda_0$. (b) Transmitted field at $z_{ot} = -30\lambda_0$. The maximum height of the profile in Fig. 8 is $5.88\lambda_0$. (—) Reference solution and (---) beams.

above, to establish the scattered/transmitted field by local beam sampling of the interface profile, the aperture must be close enough to the surface so that the rapid spread of the beams at greater distances has not yet taken place. However, this is not a severe limitation, since, as shown in Fig. 11, it is possible to perform a two-step Gabor decomposition, i.e., projecting the beam-computed radiated field onto a virtual aperture (within a wavelength above the surface so as to restore the local sampling capability) and then again applying the (narrow-waisted) beam algorithm. The additional cost is computationally minimal because of the highly efficient evaluation of the Gabor coefficients for narrow-waisted beams [see (19)]. Even for relatively “low-frequency” geometries with critical dimensions on the order of a wavelength or less, the beam algorithm, though no longer *highly* accurate, may still be able to provide reasonably good predictions. For *small* radius of curvature portions on the roughness profile, this is demonstrated in Fig. 12, with the problem parameters given in the figure caption. Of course, each such excursion into the “low-frequency” range must be validated independently by the scrambling criterion, but it is worth noting that such attempts need not to be doomed to failure *a priori*. Although the beam algorithm performance assessments here are based on the single example depicted in Figs. 7 and 8, we emphasize that confidence in these assessments has been established by an extensive sequence of relevant prior

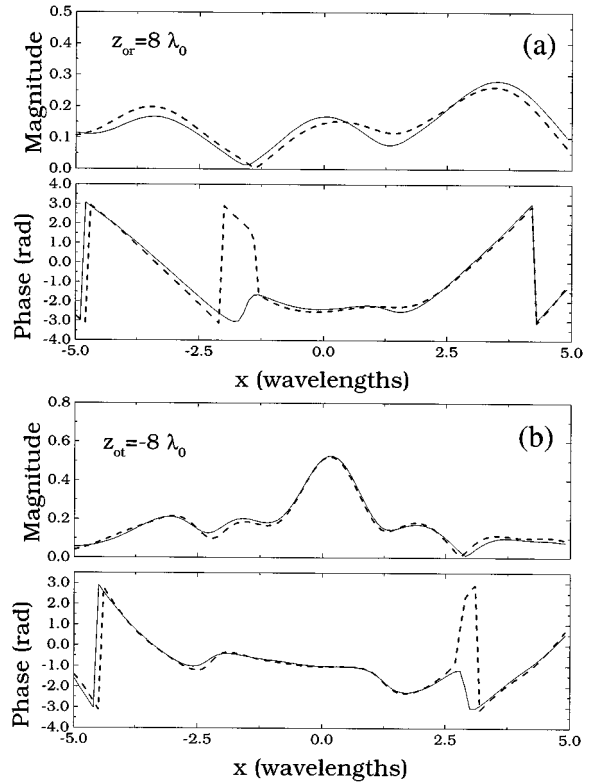


Fig. 12. Rough interface profile as in Fig. 8, but scaled so as to have maximum slope: 34° ; minimum curvature radius: $0.42\lambda_0$; maximum height: $0.588\lambda_0$. Beam-computed reflected/transmitted fields and Leviatan–Boag reference solutions. Aperture width: $d = 8\lambda_0$; aperture height: $z_A = 0.8\lambda_0$; beam lattice period: $L_x = 0.1d = 0.8\lambda_0$ (i.e., ten beams). (a) Reflected field at $z_{or} = 8\lambda_0$. (b) Transmitted field at $z_{ot} = -8\lambda_0$. (—) Reference solution and (---) beams.

numerical calibration experiments. In particular, we have found accurate results for moderate roughness with maximum slopes $\lesssim 40^\circ$ and (average) curvature radii larger than a wavelength, and for incidence directions far from grazing [$\nu \lesssim \sin 30^\circ$ in (20)]. We have also performed simulations for very high and low dielectric contrasts with $\text{Re}(\epsilon_r)$ ranging from 1.2 to 10 and $\text{Im}(\epsilon_r)$ up to 0.5; the quality of results was found to be similar to those for the case of Fig. 8. Concerning computing times, for the most expensive (two-step) implementation described in Fig. 11, we found about 9 ms to compute the field at a single position on a 500-MHz laptop; no particular effort was made to optimize the numerical code.

V. CONCLUDING REMARKS

A previously developed, rigorous, self-consistent, quasi-ray, narrow-waisted beam algorithm [1]–[3] has been reexamined and applied to the new problem of EM transmission and reflection in the presence of moderately rough surfaces.

The narrow-waisted beams are wave objects with very short collimation lengths and therefore act like *almost real* ray fields. The rigorous complex ray machinery that is required for arbitrary complex source point beams can thus be reduced to *almost real* ray asymptotics, which, however, avoids failures near caustics and other ray-field transition regions. Due to its rapid spreading, the paraxial approximations described in Section III are of little use for tracking a *single* narrow-waisted beam away

from an aperture into the far zone. However, as observed in [2] and [3], when treated *collectively* as an *ensemble*, each paraxial beam can be tracked locally to and from dielectric interfaces with large radius of curvature. Apart from the paraxial beam-tracing procedure, the computational complexity of the beam algorithm is $\mathcal{O}(N_b)$, with N_b being the number of beams in the expansion. In contrast to standard ray methods, time-consuming eigenray search for evaluating the field at a given observation point is avoided, and the beam tracing needs to be performed *only once* (phrased in another way, we are interested in reliably evaluating the actual field at the observer without worrying about which appropriately weighted individual beam contributions establish that field). The algorithm has been calibrated against a reference (full-wave) solution, and will be exploited as a fast forward solver in the interface profile estimation problem [4]. A two-step beam decomposition may be required when the aperture is “sufficiently far” from the scatterer. In this case, the computational complexity is of the same order as the physical optics (PO) Kirchhoff approximation for a dielectric interface. However, the beam algorithm usually allows a coarser discretization, so that N_b is typically smaller than the corresponding number of PO integration points in the Kirchhoff integral [see (2), but applied to integration along the interface profile] for specified accuracy. In addition, the beam algorithm is potentially able to accommodate multiple interactions in a simpler fashion. The algorithm can be extended to pulsed excitations [27].

APPENDIX A MULTIPLE REFLECTIONS

Consider an incident Gaussian beam generated by the CSP in (21). Denoting by $P_{i0}^{(j)} = (x_{i0}^{(j)}, z_{i0}^{(j)})$, $j = 1, \dots, N_r$, the real incidence points from which emerge the (real ray) reflected beam axes, and by $P_{i0}^{(N_r+1)} = P_0$ the orthogonal projection of the observation point P onto the axis of the $(N_r + 1)$ th beam that reaches the observer, the far-zone paraxial approximation (22) can be generalized as follows (implementing the procedure outlined at the end of Section III-B):

$$E^{refl}(P) \approx E^{inc}(P_{i0}^{(1)}) \exp\left(ik_0 \tilde{\delta}_p^{(N_r)}\right) \cdot \prod_{j=1}^{N_r} \Gamma^{(j)} \left(\frac{\tilde{f}_{r0}^{(j)}}{\tilde{R}_0^{(j)}} \right)^{1/2} \exp\left(ik_0 L_{r0}^{(j)}\right) \quad (39)$$

where

$$\tilde{R}_0^{(j)} = \overline{\tilde{f}_{r0}^{(j)} P_{i0}^{(j+1)}}, \quad L_{r0}^{(j)} = \overline{P_{i0}^{(j)} P_{i0}^{(j+1)}} \quad (40)$$

$$\tilde{f}_{r0}^{(j)} = \left(x_{i0}^{(j)} - \tilde{f}_{r0}^{(j)} \cos \alpha_r^{(j)}, z_{i0}^{(j)} - \tilde{f}_{r0}^{(j)} \sin \alpha_r^{(j)} \right) \quad (41)$$

$$\tilde{f}_{r0}^{(j)} = \frac{\tilde{L}_{i0}^{(j)} r_c^{(j)} \cos \theta_i^{(j)}}{2\tilde{L}_{i0}^{(j)} + r_c^{(j)} \cos \theta_i^{(j)}} \quad (42)$$

$$\tilde{L}_{i0}^{(1)} = \overline{\tilde{P}_s P_{i0}^{(1)}}, \quad \tilde{L}_{i0}^{(j)} = \overline{\tilde{f}_{r0}^{(j-1)} P_{i0}^{(j)}}, \quad j \geq 2 \quad (43)$$

$$r_c^{(j)} = - \frac{\left[1 + (h'(x_{i0}^{(j)}))^2 \right]^{3/2}}{h''(x_{i0}^{(j)})} \quad (44)$$

$$\Gamma^{(j)} = \frac{\cos \theta_i^{(j)} - \sqrt{\epsilon_r \sin^2 \theta_r^{(j)}}}{\cos \theta_i^{(j)} + \sqrt{\epsilon_r - \sin^2 \theta_r^{(j)}}} \quad (45)$$

$$\tilde{\delta}_p^{(N_r)} = \tilde{R}^{(N_r)} - \tilde{R}_0^{(N_r)}, \quad \tilde{R}^{(N_r)} = \overline{\tilde{f}_{r0}^{(N_r)} P}. \quad (46)$$

In the above expressions, $\theta_i^{(j)}$ and $\alpha_r^{(j)}$ represent the real incidence angle [with respect to the surface normal at $P_{i0}^{(j)}$] and the real reflected beam axis departure angle (with respect to the x -axis) at the j th reflection, respectively.

APPENDIX B REFERENCE SOLUTION

The multifilament–current method in [25] has been applied originally to scattering by, and transmission through, dielectric cylinders, but it can be easily adapted to smoothly irregular scatterers or moderately rough interfaces. The extension is based on the expansion of the internal and external unknown fields in terms of basis functions generated by suitably placed line sources. As in many numerical techniques, the unknown expansion coefficients are determined by solving a linear system obtained by enforcing the continuity boundary conditions of the tangential fields at selected points on the dielectric interface. In the resulting algorithm, the obtained numerical solution satisfies Maxwell’s equations by construction, and its convergence and consistency are controlled by monitoring the error (discontinuity) in the boundary conditions at the dielectric interface between the matching points [25]. In our simulations, we chose the relevant parameters so as to ensure a 0.01% maximum error in the boundary conditions.

ACKNOWLEDGMENT

The authors gratefully acknowledge helpful comments by the anonymous referees.

REFERENCES

- [1] J. J. Maciel and L. B. Felsen, “Systematic study of fields due to extended apertures by Gaussian beam discretization,” *IEEE Trans. Antennas Propagat.*, vol. 37, pp. 884–892, July 1989.
- [2] —, “Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part I—Plane layer,” *IEEE Trans. Antennas Propagat.*, vol. 38, pp. 1607–1617, Oct. 1990.
- [3] —, “Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part II—Circular cylindrical layer,” *IEEE Trans. Antennas Propagat.*, vol. 38, pp. 1618–1624, Oct. 1990.
- [4] V. Galdi, D. A. Castañón, and L. B. Felsen, “Multifrequency reconstruction of moderately rough interfaces via quasi-ray Gaussian beams,” *IEEE Trans. Geosci. Remote Sensing*, submitted for publication.
- [5] A. Ishimaru, *Wave Propagation and Scattering in Random Media*. New York: Academic, 1978.
- [6] F. G. Bass and I. M. Fucks, *Wave Scattering from Statistically Rough Surfaces*. New York: Pergamon, 1979.
- [7] P. Beckman and A. Spizzichino, *Scattering of Electromagnetic Waves from Rough Surfaces*. New York: Artech House, 1987.
- [8] E. I. Thorsos, “The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum,” *J. Acoust. Soc. Amer.*, vol. 83, pp. 78–82, Aug. 1988.
- [9] F. D. Hastings, J. B. Schneider, and S. L. Broschat, “A Monte-Carlo FDTD technique for rough surface scattering,” *IEEE Trans. Antennas Propagat.*, vol. 43, pp. 1183–1191, Nov. 1995.

- [10] K. Sarabandi, Y. Oh, and F. T. Ulaby, "Numerical simulation of scattering from one-dimensional inhomogeneous dielectric random surfaces," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 425–432, Mar. 1996.
- [11] A. Collaro, G. Franceschetti, M. Migliaccio, and D. Riccio, "Gaussian rough surfaces and Kirchhoff approximation," *IEEE Trans. Antennas Propagat.*, vol. 47, pp. 392–398, Feb. 1999.
- [12] Q. Li, C. H. Chan, and L. Tsang, "Monte Carlo simulations of wave scattering from lossy dielectric random rough surfaces using the physics-based two-grid method and the canonical-grid method," *IEEE Trans. Antennas Propagat.*, vol. 47, pp. 752–763, Apr. 1999.
- [13] T. Dogaru and L. Carin, "Multiresolution time domain analysis of scattering from a rough dielectric surface," *Radio Sci.*, vol. 35, no. 6, pp. 1279–1292, Nov./Dec. 2000.
- [14] Y. Z. Ruan and L. B. Felsen, "Reflection and transmission of beams at a curved interface," *J. Opt. Soc. Amer. A*, vol. 3, no. 4, pp. 566–579, Apr. 1986.
- [15] B. Z. Steinberg, H. Heyman, and L. B. Felsen, "Phase-space beam summation for time-harmonic radiation from large apertures," *J. Opt. Soc. Amer. A*, vol. 8, pp. 41–59, Jan. 1991.
- [16] M. J. Bastiaans, "Gabor's expansion of a signal into Gaussian elementary signals," *Proc. IEEE*, vol. 68, pp. 538–539, 1980.
- [17] —, "A sampling theorem for the complex spectrogram, and Gabor's expansion of a signal in Gaussian elementary signals," *Opt. Eng.*, vol. 20, pp. 594–598, 1981.
- [18] P. D. Einziger and M. Shapira, "Gabor representation and aperture theory," *J. Opt. Soc. Amer. A*, vol. 3, no. 4, pp. 508–522, Apr. 1986.
- [19] B. Z. Steinberg, H. Heyman, and L. B. Felsen, "Phase-space methods for radiation from large apertures," *Radio Sci.*, vol. 26, no. 1, pp. 219–227, Jan./Feb. 1991.
- [20] V. Galdi, L. B. Felsen, and D. A. Castañón, "Narrow-waisted Gaussian beam discretization for short-pulse radiation from one-dimensional large apertures," *IEEE Trans. Antennas Propagat.*, vol. 49, pp. 1322–1332, Sept. 2000.
- [21] L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves*. Piscataway, NJ: IEEE Press, 1994.
- [22] J. L. Gross and J. Yelley, *Graph Theory and its Applications*. Boca Raton, FL: CRC Press, 1998.
- [23] L. L. Shumaker, *Spline Functions: Basic Theory*. New York: Wiley, 1981.
- [24] J. E. Hipp, "Soil electromagnetic parameters as functions of frequency, soil density, and soil moisture," *Proc. IEEE*, vol. 62, pp. 98–103, Jan. 1974.
- [25] Y. Leviatan and A. Boag, "Analysis of electromagnetic scattering from dielectric cylinders using a multifilament current model," *IEEE Trans. Antennas Propagat.*, vol. AP-35, pp. 1119–1127, Oct. 1987.
- [26] L. B. Felsen and V. Galdi, "Complex-source-point narrow-waisted ray-like Gaussian beams for frequency and time domain radiation and scattering," in *Ultra-Wideband, Short Pulse Electromagnetics 5*, S. R. Cloude and P. D. Smith, Eds. New York: Kluwer/Plenum, 2001.
- [27] V. Galdi, L. B. Felsen, and D. A. Castañón, "Quasi-ray Gaussian beam algorithm for short-pulse two-dimensional scattering by moderately rough interfaces," *IEEE Trans. Antennas Propagat.*, vol. 49, submitted for publication.



Vincenzo Galdi (M'98) was born in Salerno, Italy, on July 28, 1970. He received the Laurea degree (*summa cum laude*) in electrical engineering and the Ph.D. degree in applied electromagnetics from the University of Salerno, Italy, in 1995 and 1999, respectively.

From April to December 1997, he held a Visiting Position in the Radio Frequency Division of the European Space Research & Technology Centre (ESTEC-ESA), Noordwijk, The Netherlands, where he was involved in developing CAD tools for microwave filters and phased-array antennas with coaxial excitation. In

September 1999, he received a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. In October 1999, he became a Research Associate in the Department of Electrical and Computer Engineering, Boston University, Boston, MA, where he is currently working on wave-oriented imaging algorithms for landmine detection and classification. His research interests include analytical and numerical techniques for wave propagation in complex environments, path integrals, and stochastic resonance.

Dr. Galdi is the recipient of the 2001 International Union of Radio Science (URSI) "Young Scientist Award." He is a member of Sigma Xi.



Leopold B. Felsen (S'47–M'54–SM'55–F'62–LF'90) was born in Munich, Germany, on May 7, 1924. He received the B.E.E., M.E.E., and D.E.E. degrees from the Polytechnic Institute of Brooklyn, Brooklyn, NY, in 1948, 1950, and, 1952, respectively.

He emigrated to the United States in 1939 and served in the U.S. Army from 1943 to 1946. After 1952, he remained with the Polytechnic Institute of Brooklyn (now Polytechnic University), becoming a Professor in 1962 and University Professor in 1978. From 1974 to 1978, he was Dean of Engineering.

In 1994, he resigned from the full-time Polytechnic Faculty and became University Professor Emeritus. He is now a part-time Professor of Aerospace and Mechanical Engineering and Professor of Electrical and Computer Engineering at Boston University, Boston, MA. He is the author or coauthor of more than 300 papers and several books, including *Radiation and Scattering of Waves* (Englewood Cliffs, NJ: Prentice-Hall, 1973; Piscataway, NJ: IEEE Press, 1994). He is an Associate Editor of several professional journals and an editor of the *Wave Phenomena Series* (New York: Springer-Verlag). His research interests encompass wave propagation and diffraction in complex environments and in various disciplines, high-frequency asymptotic and short-pulse techniques, and phase-space methods with an emphasis on wave-oriented data processing and imaging.

Dr. Felsen is a member of Sigma Xi and a Fellow of the Optical Society of America and the Acoustical Society of America. He has held named Visiting Professorships and Fellowships at universities in the United States and abroad, including the Guggenheim in 1973 and the Humboldt Foundation Senior Scientist Award in 1981. In 1974, he was an IEEE Antennas and Propagation Society (APS) Distinguished Lecturer. He received the Balthasar van der Pol Gold Medal from the International Union of Radio Science (URSI) in 1975, an honorary doctorate from the Technical University of Denmark in 1979, the IEEE Heinrich Hertz Gold Medal for 1991, the APS Distinguished Achievement Award for 1998, the IEEE Third Millennium Medal in 2000 (nomination by APS), three Distinguished Faculty Alumnus Awards from Polytechnic University, and an IEEE Centennial Medal in 1984. He also has received Best Paper awards for several papers in which he was an author or coauthor. In 1977, he was elected to the National Academy of Engineering. He served on the APS Administrative Committee from 1963 to 1966 and as Vice Chairman and Chairman for both the United States (1966–1973) and the International (1978–1984) URSI Commission B.

David A. Castañón (S'68–M'79–SM'98) received the B.S. degree in electrical engineering from Tulane University, New Orleans, LA, in 1971 and the Ph.D. degree in applied mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1976.

From 1976 to 1981, he was a Research Associate with the Laboratory for Information and Decision Systems, MIT. From 1982 to 1990, he was Senior and Chief Research Scientist at Alphatech, Inc., Burlington, MA. Since 1990, he has been a Professor in the Department of Electrical and Computer Engineering, Boston University, Boston, MA. His research interests include stochastic control and estimation, optimization, and image processing.

Dr. Castañón served as a member of the Board of Governors of the IEEE Control Systems Society. He is also a member of AMS, SIAM, and INFORMS.

Narrow-Waisted Gaussian Beam Discretization for Short-Pulse Radiation From One-Dimensional Large Apertures

Vincenzo Galdi, *Member, IEEE*, Leopold B. Felsen, *Life Fellow, IEEE*, and David A. Castañón, *Senior Member, IEEE*

Abstract—In this paper, we develop a Gabor-based Gaussian Beam (GB) algorithm for representing two-dimensional (2-D) radiation from finite aperture distributions with short-pulse excitation in the time domain (TD). The work extends previous results using 2-D frequency-domain (FD) narrow-waisted Gaussian beams [1]. The FD algorithm evolves from the rigorous Kirchhoff integration over the aperture distribution, which is then parameterized via the discrete Gabor basis and evaluated asymptotically for high frequencies to furnish the GB propagators to the observer. The TD results are obtained by Fourier inversion from the FD and yield pulsed beams (PB). We describe the resulting TD algorithm for several aperture distributions, ranging from simple linearly phased (linear delay) to arbitrary time delay profiles; the latter accommodate the important case of focusing TD aperture fields. For modulated pulses with Gaussian envelopes, we compute accurate closed form analytic solutions, which have been calibrated against numerical reference data. Our results confirm that the previously established utility of the Gabor-based narrow-waisted FD-GB algorithm for radiation from distributed apertures [1] remains intact in the TD.

Index Terms—Gabor lattice representations, Gaussian beams (GBs), pulsed beam wavepackets.

I. INTRODUCTION

GAUSSIAN beams (GBs) have been used *conventionally* as *highly collimated* basis elements in a variety of frequency domain (FD) and time domain (TD) radiation, propagation and scattering scenarios, especially in the asymptotic high frequency (HF) regime. The bases can be continuous or discrete, the latter being anchored to a Gabor lattice in the (physical-domain)–(spectral domain) FD or TD phase space [2]–[10]. When applied to radiation from extended

GB-parameterized aperture distributions, the radiated fields at the observer are synthesized by summation over the individual basis beam contributions. When interacting with a complex propagation or scattering environment, the (localization) resolution capabilities of the GB algorithms are essentially on the order of the collimation width of the basis beams, which generally extends over many wavelengths. This is in contrast to HF ray-based algorithms that explore local properties of the environment, but have deficiencies due to failures in ray-optical transition regions near shadow boundaries, caustics, etc. A *nonconventional* form of the Gabor algorithm, which uses *narrow-waisted poorly collimated* basis beams, simulates ray-like behavior without the transition region failures of ray fields. In the FD, the collective effect of the narrow-waisted nonconventional Gabor-stacked basis beams has been shown capable of furnishing highly accurate and numerically efficient solutions for scattering by, or transmission through, complex environments irradiated by aperture distributions [11]–[14]; this renders the algorithm useful as an efficient forward solver for inverse scattering and reconstruction [15].

In this paper, we extend the FD Gabor-based narrow-waisted beam algorithm for radiation from extended apertures [1] to the TD. The initial Kirchhoff integration over the FD aperture distribution is parameterized via the discrete Gabor basis and reduced by high frequency asymptotics to furnish the GB propagators to the observer; the TD pulsed beam (PB) propagators are obtained by Fourier inversion from the FD. For insight into the analytic implications and physical interpretation of the FD-inverted integrals, various aperture distributions are analyzed, starting from linearly phased (linear delay) to arbitrary time delay profiles; the latter are illustrated by the important example of focusing TD aperture fields. Accurate and easily computable closed-form analytic solutions are obtained for modulated pulses with Gaussian envelopes and are calibrated against independently generated numerical reference data. These results confirm that the previously established utility of the Gabor-based narrow-waisted FD-GB algorithm for radiation from distributed apertures [1] remains intact in the time domain.

In the FD, it has been established that the GB basis beams can be tracked efficiently from the aperture through interactions with complex scattering environments [11]–[14] and are therefore useful forward solvers in inversion scenarios [15]. Interaction of the PB propagators with complex environments is presently under consideration.

The rest of this paper is organized as follows. Section II has a compact summary of the FD formulation and its asymptotic re-

Manuscript received September 19, 2000; revised February 28, 2001. This work was supported by ODDR&E under MURI Grants ARO DAAG55-97-1-0013, by AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. The work of L. B. Felsen was supported in part under Grant 9900448 by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University, Brooklyn, NY, USA.

V. Galdi is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA and with the University of Sannio, Benevento, Italy (e-mail: vgaldi@bu.edu).

L. B. Felsen is with the Department of Aerospace and Mechanical Engineering and the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA and Polytechnic University, Brooklyn, NY 11201 USA (e-mail: lfelsen@bu.edu).

D. A. Castañón is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA (e-mail: dac@bu.edu).

Publisher Item Identifier S 0018-926X(01)07643-8.

duction for narrow-waisted basis beams; the results are applied to linearly phased apertures (which includes the nonphased special case). Section III is concerned with the inversion to the TD, treating in sequence space-time separable linear-delay and nonlinear-delay aperture field profiles. The analytic reduction of the GB inversion integrals for modulated Gaussian envelope pulsed beams (PB) is presented succinctly, with details relegated to several appendixes. Simplifying approximations are stated. The resulting constraints on the range of validity of the algorithms are assessed analytically and phrased concisely in terms of critical nondimensional estimators. Calibration/confirmation is implemented through a sequence of numerical tests and examples, including radiation from a cosine-tapered aperture distribution with nonlinear (quadratic) delay, which is tracked as the wavefield approaches, reaches, and propagates beyond, the focal plane. Snapshots in the examples are resolved in terms of spatial and temporal cuts that depict the spatial and temporal deformations along the propagation path of the pulsed wavefield. Conclusions are presented in Section IV.

II. FREQUENCY DOMAIN FORMULATION

We briefly summarize the frequency domain (FD) formulation, and refer the interested reader to [1]–[4], [11], [12] for details. We restrict ourselves to the 2-D (x, z) domain, with all configurational quantities independent on y . Excitation is specified as a y -directed electric field with implicit time-harmonic dependence $\exp(-i\omega t)$ and spatial distribution $F(x, \omega)$, which occupies the aperture region $|x| \leq d/2$ at $z = 0$

$$E_y(x, z = 0, \omega) = F(x, \omega), \quad |x| \leq d/2. \quad (1)$$

The aperture geometry is shown in Fig. 1. The resulting (TM polarized) electromagnetic (EM) field radiated into the half-space $z > 0$ can be expressed by line-source superposition (Kirchhoff integration) in physical space

$$E_y(x, z, \omega) = -2 \frac{\partial}{\partial z} \int_{-d/2}^{d/2} F(x', \omega) G_{2D}(kR) dx' \quad (2)$$

$$\mathbf{H} = \frac{1}{i\omega\mu_0} \nabla \times E_y \mathbf{u}_y = H_x \mathbf{u}_x + H_z \mathbf{u}_z \quad (3)$$

where

$k = \omega\sqrt{\epsilon_0\mu_0} = 2\pi/\lambda$ free-space wavenumber, with λ as the wavelength;

$\mathbf{u}_0^{(1)}(\cdot)$ denotes a unit vector; zeroth order Hankel function of the first kind;

and (in accord with the radiation condition)

$$R = \sqrt{(x - x')^2 + z^2} \geq 0. \quad (4)$$

In the remainder of this section, the ω dependence in E_y and F will be omitted in the notation. Equation (2) is derived using the free-space 2-D Green's function $G_{2D}(R)|_{z'=0}$ to simulate an equivalent magnetic line dipole current distribution on an electric wall boundary.

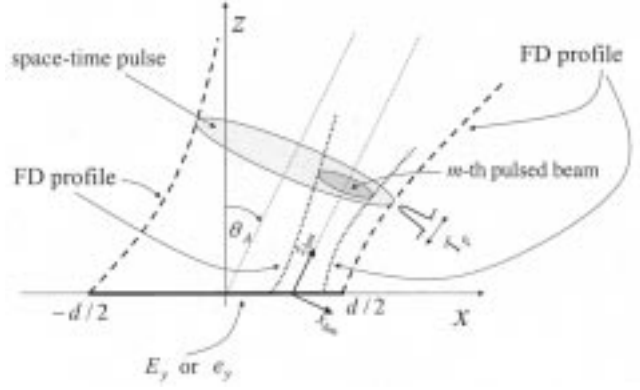


Fig. 1. Aperture geometry, global coordinates and beam coordinates for a tapered linearly phased aperture field distribution. FD: $F(x, \omega) = P(\omega)g(x) \exp(ikx \sin \theta_A)$ [see (29)]. TD: $f(x, t) = g(x)p(t - c^{-1}x \sin \theta_A)$ [see (28)].

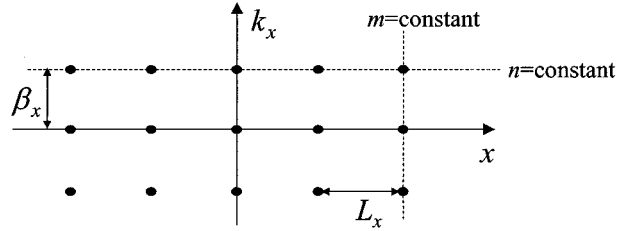


Fig. 2. Discretized Gabor lattice in the (x, k_x) phase space, where k_x is the x -domain wavenumber. $n = \text{const.}$: spatially displaced beams centered at $x_m = mL_x$, with fixed phase gradient (beam tilt) $(n\beta_x)$; $m = \text{const.}$: spectrally displaced beams with beam tilts $(n\beta_x)$ at fixed location x_m .

A. FD Beam Discretization

1) *Aperture Field*: The aperture field $F(x)$ is to be parameterized in terms of GB basis functions via the rigorous self-consistent Gabor series representation [16], [17]

$$F(x) = \sum_{m,n=-\infty}^{\infty} A_{mn} w(x - mL_x) \exp(in\beta_x x) \quad (5)$$

where $w(x)$ represents the normalized Gaussian window (initial beam profile)

$$w(x) = \left(\frac{\sqrt{2}}{L_x} \right)^{1/2} \exp[-\pi(x/L_x)^2] \quad (6)$$

$$\int_{-\infty}^{\infty} w^2(x) dx = 1.$$

This representation places the beam basis functions on a discretized (x, k_x) phase-space lattice (see Fig. 2), on which each lattice point gives rise to a GB, whose spatial and spectral (beam tilt) shifts are tagged by the indexes m and n , respectively. Spatial and spectral periods are related by the self-consistency relation (configuration-spectrum tradeoff) $L_x\beta_x = 2\pi$ [16], [17]. The expansion coefficients in (5) can be computed by introducing an auxiliary *bi-orthogonal* function $\gamma(x)$ [16], [17]

$$A_{mn} = \int_{-\infty}^{\infty} F(x) \gamma^*(x - mL_x) \exp(-in\beta_x x) dx \quad (7)$$

but as shown below, our subsequent use of narrow-waisted beams avoids the cumbersome computation of γ and numerical implementation of (7). Combining (2) and (5) yields the Gabor lattice representation of the aperture field.

2) *Radiated Field*: The corresponding Gabor lattice representation for the field radiated into the half-space $z > 0$ is given by [1]

$$E_y(x, z) = \sum_{m, n=-\infty}^{\infty} A_{mn} B_{mn}(x, z) \quad (8)$$

where the beam functions $B_{mn}(x, z)$ are expressed by Gabor-weighted line-source superposition

$$B_{mn}(x, z) = -\frac{i}{2} \int_{-d/2}^{d/2} w(x' - mL_x) \times \exp(in\beta_x x') \frac{\partial}{\partial z} H_0^{(1)}(kR) dx' \quad (9)$$

R being defined in (4). By saddle point methods, the integral in (9) (or its spectral counterpart) can be evaluated asymptotically in the beam *paraxial far zone*, yielding the following *complex source point* (CSP) approximation [11], [12]

$$B_{mn}(x, z) \sim -ik2^{5/4} \left(\frac{L_x}{8\pi k} \right)^{1/2} \times \exp\left\{i\left[k\left(\tilde{R}_{mn} + ib_n\right) + \pi/4\right]\right\} \frac{(z - \tilde{z}'_{mn})}{\tilde{R}_{mn}^{3/2}} \quad (10)$$

with \tilde{R}_{mn} representing the *complex distance*

$$\tilde{R}_{mn} = \overline{P_o \tilde{P}'_{mn}} = \sqrt{(x - \tilde{x}'_{mn})^2 + (z - \tilde{z}'_{mn})^2} \quad (11)$$

between the observer at $P_o = (x, z)$ and the *complex source point* (CSP), $\tilde{P}'_{mn} = (\tilde{x}'_{mn}, \tilde{z}'_{mn}) = (mL_x + ib_n \sin \theta_n, ib_n \cos \theta_n)$, where, in accord with the radiation condition [see (4)], the square root is defined by $\text{Re}(\tilde{R}_{mn}) \geq 0$. Here and henceforth, the tilde \sim denotes a complex quantity. The CSP displacement parameter (Fresnel length) b_n is related to the beam lattice period L_x and the beam axis angle $\theta_n = \sin^{-1}(n\lambda/L_x)$ via $b_n = (L_x \cos \theta_n)^2 / \lambda$ [11], and (10) is valid in the paraxial far-zone of each beam, $|\tilde{R}_{mn}| \gg b_n$. As the tilt index n increases to $|n| > L_x/\lambda$, the beam tilt angle θ_n becomes complex, and the corresponding beams become *evanescent*.

B. Narrow-Waisted Beams

As emphasized in [1], [11], and [12], *narrow-waisted* beams ($L_x \lesssim \lambda \ll d$) have several attractive features. First, the Gabor coefficients can be estimated with good approximation by *sampling* the aperture field distribution, avoiding the time-consuming integration (7) [1]

$$A_{mn} \approx \begin{cases} (L_x/\sqrt{2})^{1/2} F(mL_x), & n = 0 \\ 0, & n \neq 0 \end{cases} \quad (12)$$

so that

$$E_y(x, z) \approx \sum_{|m| \leq (d/2L_x)} A_{m0} B_{m0}(x, z) \quad (13)$$

$$B_{m0}(x, z) \sim -ik2^{5/4} \left(\frac{L_x}{8\pi k} \right)^{1/2} \times \exp\left\{i\left[k\left(\tilde{R}_{m0} + ib\right) + \pi/4\right]\right\} \frac{(z - ib)}{\tilde{R}_{m0}^{3/2}} \quad (14)$$

$$\tilde{R}_{m0} = \sqrt{(x - mL_x)^2 + (z - ib)^2} \quad (15)$$

$$b = b_0 = L_x^2/\lambda. \quad (16)$$

In this approximation, the tilted ($n \neq 0$) beams in the Gabor expansion, which generate evanescent “far fields” for $L_x/\lambda < |n|$, are ignored. Second, for narrow-waisted beams, the CSP paraxial *far-zone* approximation (10) can be invoked at *moderate* distance and, therefore, the paraxial beam superposition gives accurate results even in the *near zone* of the *aperture*. Third, as shown in [11]–[14], interaction of narrow-waisted beams with an environment can be effectively tracked by *complex ray asymptotics*.

C. Linearly Phased Aperture

As shown in [1] and [11], narrow-waisted beams work very well for nonphased apertures, but usually require finer aperture sampling in the presence of phasing. There are, however, special cases where more efficient implementations are possible. In particular, we consider a linearly phased aperture, which will efficiently parameterize the general case of nonlinear phasing later on

$$F(x) = g(x) \exp(ikx \sin \theta_A) \quad (17)$$

where $g(x)$ is a real function and θ_A denotes the real tilt angle of the main radiation lobe with respect to the z axis. In this case, a more effective discretization can be obtained by Gabor expanding the real function $g(x)$ only and including the linear phasing in the beam integral (9) for the B_{m0} beam propagator. Accordingly, the narrow-waisted ($L_x \lesssim \lambda \ll d$) beam expansion can be recast as (for simplicity, the subscript “ $m0$ ” is, henceforth, replaced by “ m ”)

$$E_y(x, z) \approx \sum_{|m| \leq (d/2L_x)} C_m \mathcal{B}_m(x, z) \quad (18)$$

where

$$C_m = \left(L_x/\sqrt{2} \right)^{1/2} g(mL_x), \quad |m| \leq (d/2L_x) \quad (19)$$

$$\mathcal{B}_m(x, z) = -ik2^{5/4} \left(\frac{L_x}{8\pi k} \right)^{1/2} \times \exp\left\{i\left[k\left(\tilde{R}_m + mL_x \sin \theta_A + ib\right) + \pi/4\right]\right\} \times \frac{(z - ib \cos \theta_A)}{\tilde{R}_m^{3/2}} \quad (20)$$

$$\tilde{R}_m = \sqrt{(x - mL_x - ib \sin \theta_A)^2 + (z - ib \cos \theta_A)^2} \quad (21)$$

$$b = (L_x \cos \theta_A)^2/\lambda. \quad (22)$$

The beam propagator (20) differs from B_{m0} in (14) by the phase shift ($ikmL_x \sin \theta_A$) and by the different definitions of

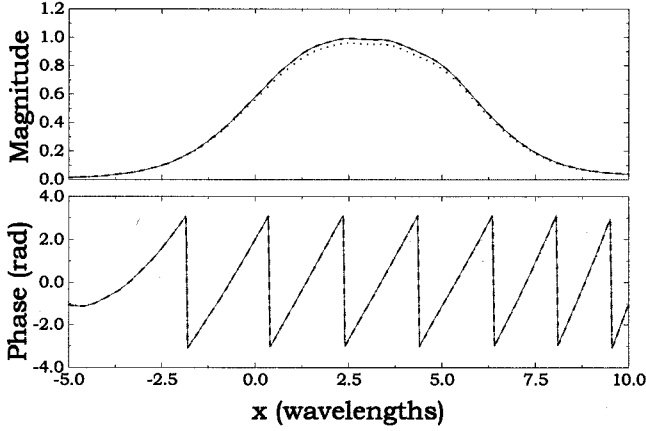


Fig. 3. Linearly phased cosine-tapered aperture distribution in (17), (23) ($d = 10\lambda$, $\theta_A = 30^\circ$) (see Fig. 1). Near-zone ($z = 5\lambda$) radiated field synthesized using narrow-waisted nontilted and tilted beams is compared with the reference solution [Kirchhoff integration in (2)]. — Reference solution; --- Tilted beam synthesis ($L_x = 0.02d$); Nontilted beam synthesis ($L_x = 0.02d$).

\tilde{R}_m and b in (21), (22), respectively, which produce the propagation-matched tilt θ_A in the beam direction (see Fig. 1). For comparison of the tilted and nontilted formulations, we consider the linearly phased distribution (17) with cosine tapering

$$g(x) = \begin{cases} \cos(\pi x/d), & |x| \leq d/2 \\ 0, & |x| > d/2 \end{cases} \quad (23)$$

$\theta_A = 30^\circ$, and a fixed beam lattice period ($L_x = 0.02d$). In Fig. 3, the near-zone fields synthesized under the same conditions with the same number of narrow-waisted nontilted (13) and tilted (18) beams are compared with the reference solution [brute force Kirchhoff integration in (2)]. The tilted beam synthesis is hardly distinguishable from the reference solution, whereas the nontilted synthesis is somewhat less accurate in magnitude.

III. TD FORMULATION

We shall now explore the extension of the FD results for aperture radiation in Section II to time-dependent excitation, in particular to short pulses. To this end, we consider a space-time aperture field distribution at $z = 0$

$$e_y(x, z = 0, t) = f(x, t), \quad |x| \leq d/2. \quad (24)$$

Using the 2D TD Green's function [18, Sec. 7.3]

$$g_{2D}(x, z, t; x', z' = 0, t') = \frac{H(t - t' - R/c)}{2\pi\sqrt{(t - t')^2 - (R/c)^2}} \quad (25)$$

with R defined in (4), c representing the speed of light, and $H(\cdot)$ representing the Heaviside step function, the field radiated into the half-space $z > 0$ can be expressed via the space-time Kirchhoff integration [18], [19]

$$\begin{aligned} e_y(x, z, t) &= \frac{z}{\pi} \int_{-d/2}^{d/2} dx' \frac{1}{R^2} \int_{-\infty}^{t-R/c} (t - t') \frac{\partial f}{\partial t'}(x', t') \\ &\quad \frac{dt'}{\sqrt{(t - t')^2 - (R/c)^2}} \end{aligned} \quad (26)$$

It is shown in Appendix A that the FD and TD fields $E_y(x, z, \omega)$ and $e_y(x, z, t)$ in (2) and (26), respectively, are related through the Fourier transform pair

$$\begin{aligned} e_y(x, z, t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} E_y(x, z, \omega) \exp(-i\omega t) d\omega \\ E_y(x, z, \omega) &= \int_{-\infty}^{\infty} e_y(x, z, t) \exp(i\omega t) dt. \end{aligned} \quad (27)$$

Direct numerical integration of (26) is complicated by the (integrable) square-root singularity at the upper limit $t' = t - R/c$ and requires care in its execution. Here, we use the Newton-Cotes scheme proposed in [19] for the numerical integration of (26), which represents our reference solution.

A. TD Beam Discretization: Linear-Delay Aperture Fields

The GB discretization for the time-harmonic aperture distribution in Section II can be generalized to arbitrary time-dependent excitation by use of a four-index Gabor series, which is set on a discretized lattice in the eight-dimensional (space-wavenumber, time-frequency) phase space (see [6] for analytic and computational issues). A condensed summary of the analytic results of section III-A has been given in [13], together with one simple numerical example. In our presentation below we give detailed derivations that were omitted in [13] but are required for complete understanding of the results. Moreover, we have here a much more comprehensive set of numerical examples as well as calibration of accuracy, which are not contained in [13]. The results in this section are used for the important generalization in Section III-B to nonlinear-delay profiles which is entirely new.

For assessment of the extent to which the FD narrow-waisted beam approach can be generalized to TD (short-pulse) excitation, we begin by considering the case of a space-time separable aperture field with linear time delay (assuming θ_A real)

$$f(x, t) = g(x)p(t - c^{-1}x \sin \theta_A) \quad (28)$$

and its FD counterpart via (27)

$$F(x, \omega) = P(\omega)g(x) \exp(ikx \sin \theta_A) \quad (29)$$

where $p(t)$ is a time-pulse with characteristic width T_p , and $P(\omega)$ is the Fourier transform of $p(t)$,

$$P(\omega) = \int_{-\infty}^{\infty} p(t) \exp(i\omega t) dt. \quad (30)$$

Equation (29) differs from (17) by the (spatially independent) frequency-dependent weight factor $P(\omega)$. This TD counterpart of the linearly phased aperture in Section II-C generates a space-time-resolved pulse propagating in the θ_A direction (see Fig. 1), with an assumed normalized pulse length much shorter than the aperture dimension d , i.e., $cT_p \ll d$.

Fourier inversion of the narrow-waisted tilted beam expansion (18) for a FD linearly phased aperture field yields a pulsed beam (PB) expansion for the corresponding TD aperture field in (28). However, the evanescent spectrum content in the FD beam propagator (20) motivates use of the *analytic signal* formulation

instead of the standard Fourier transform [5], [8]. The analytic signal $f^+(t)$ is defined via the one-sided Fourier transform

$$f^+(t) = \frac{1}{\pi} \int_0^\infty F(\omega) \exp(-i\omega t) d\omega, \quad \text{Im}(t) \leq 0 \quad (31)$$

where $F(\omega)$ is the conventional Fourier spectrum of the real signal $f(t)$ [see (27)]. The real signal for real t is recovered via

$$f(t) = \text{Re} \left[f^+(t) \right]. \quad (32)$$

The analytic TD counterpart of the narrow-waisted FD beam expansion (18) for the aperture field distribution (29) may thus be written as

$$e_y^+(x, z, t) \approx \sum_{|m| \leq (d/2L_x)} c_m b_m^+(x, z, t) \quad (33)$$

where the analytic PB propagator b_m^+ is given in terms of the paraxial, far-zone FD propagator \mathcal{B}_m in (20)

$$b_m^+(x, z, t) = \frac{1}{\pi} \int_0^\infty \mathcal{B}_m(x, z, \omega) P(\omega) \times \exp(-i\omega t) d\omega \quad \text{Im}(t) \leq 0. \quad (34)$$

For implementation, one can choose a frequency-independent beam lattice period L_x (resulting in a frequency (i.e., wavelength)-dependent beam parameter b [see (22)], or a frequency-independent beam parameter b (resulting in a frequency-dependent L_x). We consider frequency-independent L_x preferable because this yields frequency-independent Gabor coefficients [see (19)]. The real TD field e_y then follows from (33) and (34)

$$e_y(x, z, t) = \text{Re} \left[e_y^+(x, z, t) \right]$$

with

$$b_m(x, z, t) = \text{Re} \left[b_m^+(x, z, t) \right] \quad (35)$$

where the Gabor coefficients c_m can be estimated through aperture sampling when $L_x \lesssim \lambda \ll d$ [see (19)]

$$c_m = \left(L_x / \sqrt{2} \right)^{1/2} g(mL_x), \quad |m| \leq (d/2L_x). \quad (36)$$

For the important class of Gaussian time pulses, we have found useful closed-form approximations for the integral in (34). In particular, we use a Rayleigh (four-times-differentiated Gaussian) pulse (see Fig. 4)

$$p(t) = \exp \left[-\frac{50(t - T_p/2)^2}{T_p^2} \right] \times \left[1 + \frac{10000(t - T_p/2)^4 - 600(t - T_p/2)^2 T_p^2}{3T_p^4} \right] \quad (37)$$

with spectrum

$$P(\omega) = \sqrt{\frac{\pi}{2}} \frac{\omega^4 T_p^5}{150000} \exp \left(-\frac{\omega^2 T_p^2}{200} + i \frac{\omega T_p}{2} \right) \quad (38)$$

which has desirable wideband properties, being sufficiently bounded away from $\omega = 0$ and ∞ . However, the procedure below can be applied to any kind of modulated or differentiated Gaussian pulse.

The beam parameter b and hence, the complex distance \tilde{R}_m (21) in (20) are *frequency dependent* because the beam lattice period L_x has been chosen frequency independent. For $z \gg |b \cos \theta_A|$, the amplitude factor in the FD beam propagator (20) can be approximated by

$$\frac{(z - ib \cos \theta_A)}{\tilde{R}_m^{3/2}} \approx \frac{z}{R_m^{3/2}} \quad R_m = \sqrt{(x - mL_x)^2 + z^2} \quad (39)$$

where the distance $R_m \approx \tilde{R}_m$ is *real* and frequency independent. In the phase, under the same conditions, we retain the first order paraxial correction

$$\begin{aligned} \tilde{R}_m &\approx z_{bm} - ib + \frac{x_{bm}^2}{2(z_{bm} - ib)} \\ &\approx z_{bm} - ib + \frac{x_{bm}^2(z_{bm} + ib)}{2z_{bm}^2}, \quad b \ll z_{bm} \end{aligned} \quad (40)$$

where (x_{bm}, z_{bm}) are the beam coordinates (see Fig. 1)

$$\begin{bmatrix} x_{bm} \\ z_{bm} \end{bmatrix} = \begin{bmatrix} \cos \theta_A & -\sin \theta_A \\ \sin \theta_A & \cos \theta_A \end{bmatrix} \begin{bmatrix} x - mL_x \\ z \end{bmatrix}. \quad (41)$$

The phase approximation (40) is valid in the paraxial region of each beam

$$|z_{bm} - ib| \gg |x_{bm}| \quad (42)$$

together with the constraint that the beam parameter b remains small over the bandwidth Ω_p of the pulse spectrum $P(\omega)$

$$b = \frac{(L_x \cos \theta_A)^2 \omega}{2\pi c} \ll z_{bm}, \quad \omega \leq \Omega_p. \quad (43)$$

With these approximations, the integral in (34) can be evaluated in closed form, and the analytic PB propagator can be written explicitly as (see Appendix B for details)

$$\begin{aligned} b_m^+(x, z, t) &\approx \eta_m \left[\alpha_m \Gamma \left(\frac{11}{4} \right) M_1 \left(5 \sqrt{\frac{\pi}{2}} \frac{\beta_m}{\alpha_m} \right) \right. \\ &\quad \left. + 5 \sqrt{2\pi} i \beta_m \Gamma \left(\frac{13}{4} \right) M_2 \left(5 \sqrt{\frac{\pi}{2}} \frac{\beta_m}{\alpha_m} \right) \right] \end{aligned} \quad (44)$$

with

$$\alpha_m(x, z) = \sqrt{\pi(cT_p z_{bm})^2 + 50(L_x x_{bm} \cos \theta_A)^2} \quad (45)$$

$$\begin{aligned} \beta_m(x, z, t) &= x_{bm}^2 + z_{bm}(2z_{bm} + cT_p - 2ct) \\ &\quad + 2mL_x z_{bm} \sin \theta_A \end{aligned} \quad (46)$$

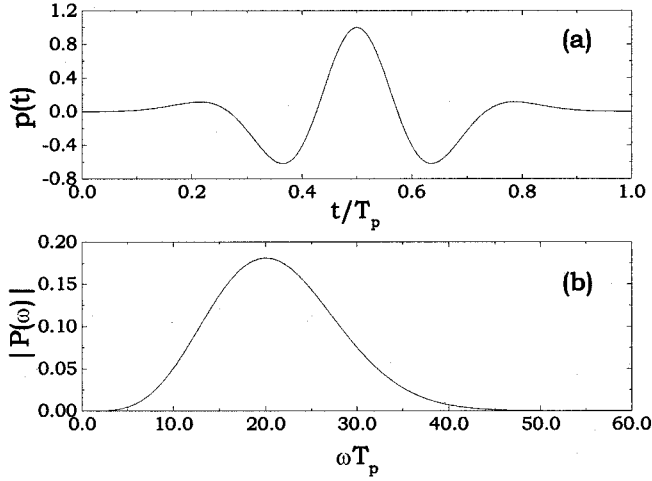


Fig. 4. Rayleigh pulse. (a) Temporal profile in (37). (b) Spectrum (magnitude) in (38).

$$\eta_m(x, z) = \frac{4z\sqrt{10L_x}\pi^{7/4}(cT_p)^5 z_{bm}^{11/2} \exp(-i\pi/4)}{3R_m^{3/2} \alpha_m(x, z)^{13/2}} \quad (47)$$

$$M_1(y) = {}_1F_1\left(\frac{11}{4}, \frac{1}{2}, -y^2\right) \quad (48)$$

$$M_2(y) = {}_1F_1\left(\frac{13}{4}, \frac{3}{2}, -y^2\right) \quad (49)$$

where R_m , x_{bm} and z_{bm} are defined in (39) and (41), $\Gamma(\cdot)$ is the gamma function [20, Sec. 6], and ${}_1F_1(u, v, y)$ is the Kummer confluent hypergeometric function [20, Sec. 13]. Since modulation or differentiation operations affect only the arguments of ${}_1F_1(u, v, y)$, the above reduction can be applied to any Gaussian-envelope pulse.

The following simple approximations can be exploited for the functions M_1 and M_2 (see Appendix C):

$$M_1(y) \approx \exp(-y^2) \left(1 - \frac{9}{2}y^2 + \frac{15}{8}y^4 - \frac{y^6}{16} - \frac{3}{896}y^8 - \frac{y^{10}}{3840} \right) \quad (50)$$

$$M_2(y) \approx \exp(-y^2) \left(1 - \frac{7}{6}y^2 + \frac{7}{40}y^4 + \frac{y^6}{240} + \frac{y^8}{3456} + \frac{y^{10}}{42240} \right). \quad (51)$$

Fig. 5 shows comparisons between the exact functions (computed via reliable numerical routines [21]) and the proposed approximations (50), (51). The results essentially overlap on the scale of the plots. Since the functions $M_{1,2}$ resemble the form and shape of the Rayleigh pulse in (37) and Fig. 4, then using (50) and (51), the TD beam propagator (44) can be computed very efficiently. Fig. 6 shows a number of instantaneous snapshots of $b_m(x, z, t)$. Note the rapid transverse spatial spreading of the ray-like narrow-waisted PB, but its retention of the wave-front behavior in the longitudinal (radial) direction.

B. TD Beam Discretization: Nonlinear-Delay Aperture Fields

Extension of the TD Gabor algorithm in Section III-A to nonlinear delay profiles is an important new generalization. Instead

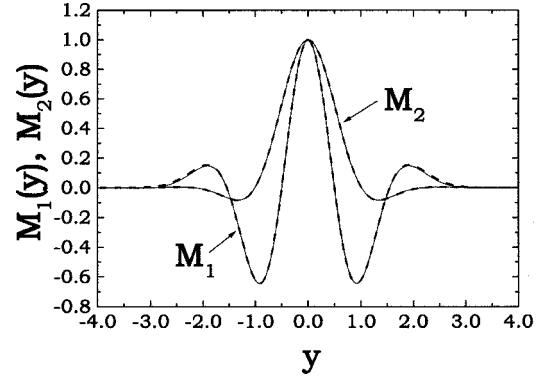


Fig. 5. Comparison between the exact and approximated functions $M_{1,2}$ in (48)–(51). — Exact; - - - Approximated.

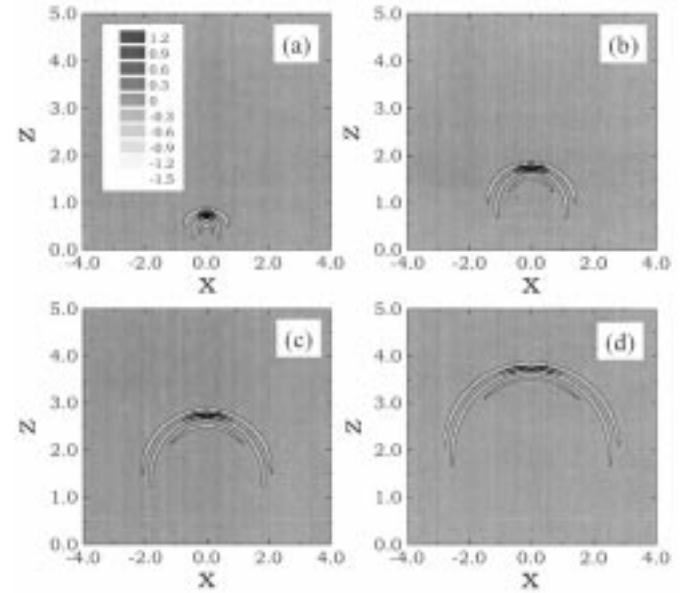


Fig. 6. Instantaneous snapshots of the TD beam propagator $b_m(x, z, t)$ ($m = 0$, $L_x = d/20$, $\theta_A = 0$, $cT_p = 0.5$). (a) $ct = 1$. (b) $ct = 2$. (c) $ct = 3$. (d) $ct = 4$.

of the linear delay in (28), we now consider the aperture field distribution

$$f(x, t) = h(x)p(t - c^{-1}\phi(x)) \quad (52)$$

with its FD counterpart

$$F(x, \omega) = P(\omega)h(x) \exp\left[i\frac{\omega}{c}\phi(x)\right] \quad (53)$$

where $h(x)$ and $\phi(x)$ are real functions. Recalling the efficient treatment of the linear delay fields in Section III-A, we split the phase (delay) function $\phi(x)$ into a linear part plus a nonlinear remainder

$$\phi(x) = x \sin \theta_A + \phi_{NL}(x) \quad (54)$$

where $\phi_{NL}(x)$ does not contain linear terms. The FD narrow-waisted tilted beam decomposition (18) of Section II-C can, thus, be applied, with $g(x)$ in (17) given by

$$g(x) = h(x) \exp[ik\phi_{NL}(x)]. \quad (55)$$

The corresponding FD Gabor coefficients [see (19)] are complex

$$C_m = \left(L_x / \sqrt{2} \right)^{1/2} h(mL_x) \exp[ik\phi_{NL}(mL_x)] \quad (56)$$

$$|m| \leq (d/2L_x)$$

and the corresponding TD beam expansion differs only slightly from (33)

$$e_y(x, z, t) \approx \sum_{|m| \leq (d/2L_x)} h_m b_m(x, z, t - \tau_m) \quad (57)$$

where b_m is the real part of the analytic PB propagator in (44), and

$$h_m = \left(L_x / \sqrt{2} \right)^{1/2} h(mL_x)$$

$$\tau_m = c^{-1} \phi_{NL}(mL_x), \quad |m| \leq (d/2L_x). \quad (58)$$

C. Assessment of Accuracy

The explicit analytic results in (44) are based on the approximations (39) and (40), which are valid in the paraxial *far-zone* (42). An additional condition is the *smallness* of the beam parameter b over the whole bandwidth of interest [see (43)]. This latter restriction, in turn, determines the maximum allowable lattice period (i.e., the minimum number of beams) for specified Ω_p , θ_A and observation point. A rough theoretical estimate can be obtained from (43). For specified observation plane at $z = z_{obs}$ and recalling that $z_{bm} = z_{obs} / \cos \theta_A$ (see Fig. 1) the overall constraint can be expressed by the following inequality in terms of the nondimensional estimator Q

$$Q \equiv N_b^{-1} \sqrt{\frac{\kappa(\cos \theta_A)^3}{\chi}} \ll 1 \quad (59)$$

where the integer $N_b = d/L_x$ represents the number of beams in the expansion (33), $\kappa = \Omega_p T_p / 2\pi$ is the normalized bandwidth of the pulse $p(t)$, and $\chi = z_{obs} / F_d$ is the distance to the observation plane scaled by the Fresnel distance of the aperture, $F_d = d^2 / (cT_p)$. The nondimensional estimator $Q \ll 1$ contains all relevant parameters of the problem. For example, decreasing the number of beams N_b can be compensated by a corresponding increase of $(z_{obs})^{1/2}$. However, a finer discretization may be required in the presence of nonlinear phasing/delay for the same degree of accuracy.

D. Numerical Results

The TD narrow-waisted beam expansions in (35), (57) have been calibrated against a reference solution implemented by space-time Kirchhoff integration of (26) in an extensive number of numerical simulations. We present and discuss selected typical results, starting with the linear-delay space-time aperture distribution (28) where $g(x)$ is the cosine tapering in (23) and $p(t)$ is the time dependence of the Rayleigh pulse in (37). Figs. 7 and 8 show typical results for the nonphased case ($\theta_A = 0$). Specifically, Fig. 7(a) depicts the time evolution of the y -directed scalar electric field at a fixed observation point in the

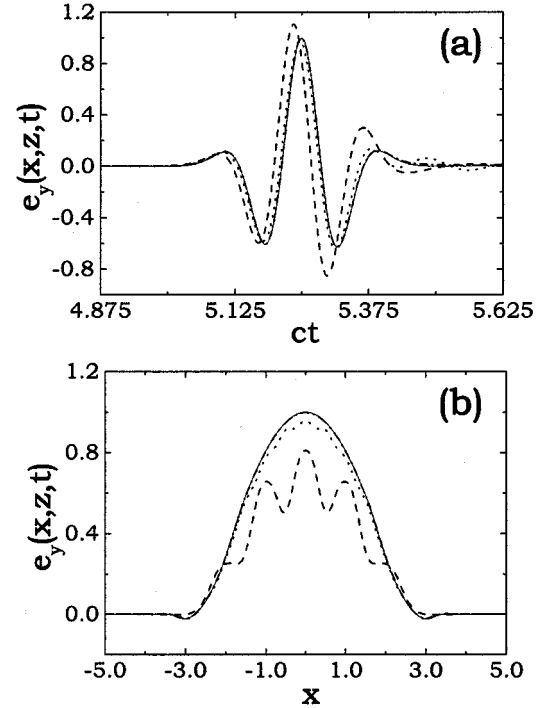


Fig. 7. Radiated field due to nonphased ($\theta_A = 0$) cosine-tapered aperture distribution. Parameters: $cT_p = 0.5$, $d = 5 = 10cT_p$ (arbitrary units). Reference solution via space-time Kirchhoff integration of (26): solid curves. TD narrow-waisted beam synthesis in (35): dashed and dotted curves. Observation points for the temporal profile are on the beam axis ($x = 0$). (a) Temporal profile at $x = 0$, $z = 5 = 0.1F_d$ (arbitrary units); (b) Spatial transverse profile at $z = 5$, $ct = 5.25$; --- 5 beams ($Q = 1.6$, $\Delta e_y = -2$ dB); 10 beams ($Q = 0.8$, $\Delta e_y = -14$ dB), - - - - 30 beams ($Q = 0.26$, $\Delta e_y = -31$ dB). The r.m.s. errors Δe_y pertain to Fig. 7(a).

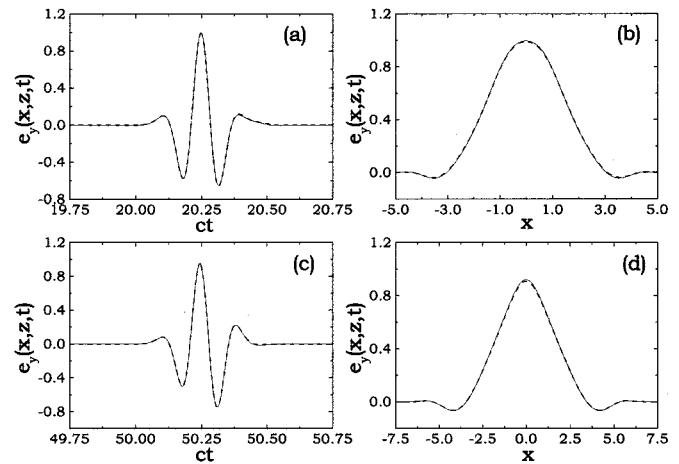


Fig. 8. Parameters as in Fig. 7. (a) Temporal profile at $x = 0$, $z = 20 = 0.4F_d$, 15 beams ($Q = 0.27$, $\Delta e_y = -32$ dB). (b) Spatial transverse profile at $z = 20$, $ct = 20.25$, 15 beams ($Q = 0.27$). (c) Temporal profile at $x = 0$, $z = 50 = F_d$, 10 beams ($Q = 0.25$, $\Delta e_y = -34$ dB). (d) Spatial transverse profile at $z = 50$, $ct = 50.25$, 10 beams ($Q = 0.25$). Reference solutions (solid curves) and beam solutions (dashed curves) coincide on the scale of the plots.

near zone ($z = 0.1F_d$) of a large aperture ($d = 10cT_p$); the reference space-time Kirchhoff integration (26) is compared with the TD beam synthesis (35) for various numbers of beams. In all simulations, we used a pulse-bandwidth value $\Omega_p T_p = 40$ (see

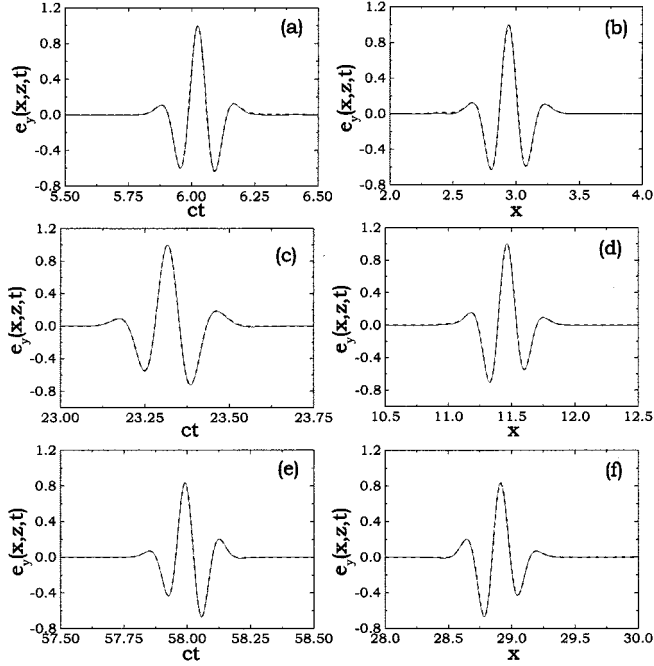


Fig. 9. Radiated field due to linearly phased ($\theta_A = 30^\circ$) cosine-tapered aperture distribution. Parameters: $cT_p = 0.5$, $d = 5 = 10cT_p$ (arbitrary units). Observation points for the temporal profiles are on the tilted beam axis θ_A . (a) Temporal profile at $x = 2.89$, $z = 5 = 0.1F_d$ (arbitrary units), 25 beams ($Q = 0.26$, $\Delta e_y = -32$ dB). (b) Spatial transverse profile at $z = 5$, $ct = 6.05$, 25 beams ($Q = 0.26$). (c) Temporal profile at $x = 11.5$, $z = 20 = 0.4F_d$, 12 beams ($Q = 0.27$, $\Delta e_y = -33$ dB). (d) Spatial transverse profile at $z = 20$, $ct = 23.3$, 12 beams ($Q = 0.27$). (e) Temporal profile at $x = 28.9$, $z = 50 = F_d$, 8 beams ($Q = 0.25$, $\Delta e_y = -33$ dB). (f) Spatial transverse profile at $z = 50$, $ct = 58$, 8 beams ($Q = 0.25$). Reference solutions (solid curves) and beam solutions (dashed curves) coincide on the scale of the plots.

Fig. 4), which sets the reference level for the nondimensional estimator Q in (59). As expected, the agreement improves with increase in the number of beams, and satisfactory accuracy is achieved for $Q \leq 0.3$. In order to better quantify the accuracy of the TD beam synthesis and the role of the nondimensional estimator Q , we computed the r.m.s. (energy) error at a fixed observation point

$$\Delta e_y := \frac{\int_{-\infty}^{\infty} |e_y^{(\text{ref})}(x, z, t) - e_y^{(\text{beam})}(x, z, t)|^2 dt}{\left[\int_{-\infty}^{\infty} |e_y^{(\text{ref})}(x, z, t)|^2 dt \int_{-\infty}^{\infty} |e_y^{(\text{beam})}(x, z, t)|^2 dt \right]^{1/2}} \quad (60)$$

whose values are explicitly indicated in the figure captions. It is observed that values of $Q \leq 0.3$ yield errors $\lesssim -30$ dB. We shall use Q as a convenient calibrator of the accuracy in the numerical experiments. It is observed from the transverse cut shown in Fig. 7(b) that despite the use of the paraxial far-zone approximation (42), the TD beam synthesis works quite well even in the near zone of the aperture ($z = 0.1F_d$) and not only around the main radiation lobe. As the observation distance increases, a coarser discretization can be used subject to (59). For

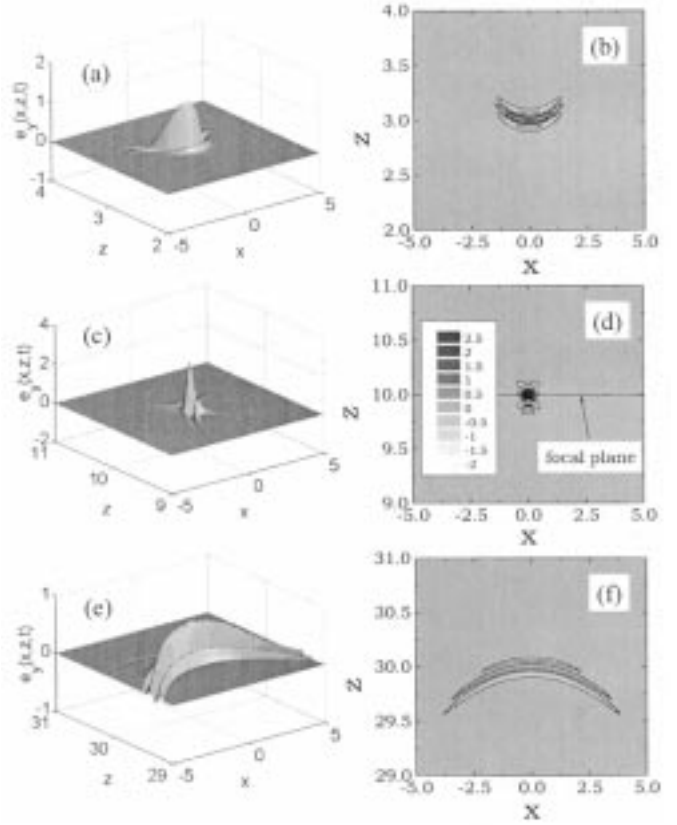


Fig. 10. Radiated field due to cosine-tapered focused aperture distribution in (61), computed via space-time Kirchhoff integration of (26). Parameters: $cT_p = 0.5$, $d = 5 = 10cT_p$, $L_f = 10 = 0.2F_d$ (arbitrary units). Instantaneous snapshots at various times. (a) $ct = 3.25$; (c) $ct = 10.235$; (e) $ct = 30.22$. (b), (d), (f) Gray-scale plots corresponding to (a), (c), (e), respectively. See also the spatial profiles in Figs. 12(b), 11(b) and 12(d) for different visualization.

instance, Fig. 8(a) and (b) show results for a moderate observation distance ($z = 0.4F_d$), whereas Fig. 8(c) and (d) show far-zone ($z = F_d$) results. In these examples, only one beam synthesis is shown, with the beam lattice period chosen so as to assure $Q \leq 0.3$, according to the above calibration; accuracy is confirmed. We conclude that even at moderate distances, acceptable results can be obtained with a relatively small number of beams (~ 20). The corresponding data for a linear-delay example ($\theta_A = 30^\circ$) are shown in Fig. 9. The same conclusions apply, except that the field maxima are tracked along the tilted beam axis θ_A .

As our final and most important example, we consider a nonlinearly phased aperture distribution. In particular, we have chosen a cosine-tapered aperture field with quadratic delay

$$f(x, t) = \begin{cases} \cos(\pi x/d)p[t + x^2/(2cL_f)], & |x| \leq d/2 \\ 0, & |x| > d/2 \end{cases} \quad (61)$$

which represents the TD counterpart of a typical time-harmonic focusing ($L_f > 0$) or defocusing ($L_f < 0$) distribution, with L_f representing the conventional focal length. The focusing case is the most challenging, since standard (nonuniform) ray asymptotics would fail due to the presence of caustic transition regions. Moreover, we have chosen numerical values so as to get focusing at moderate distance from the aperture ($L_f = 0.2F_d$). A

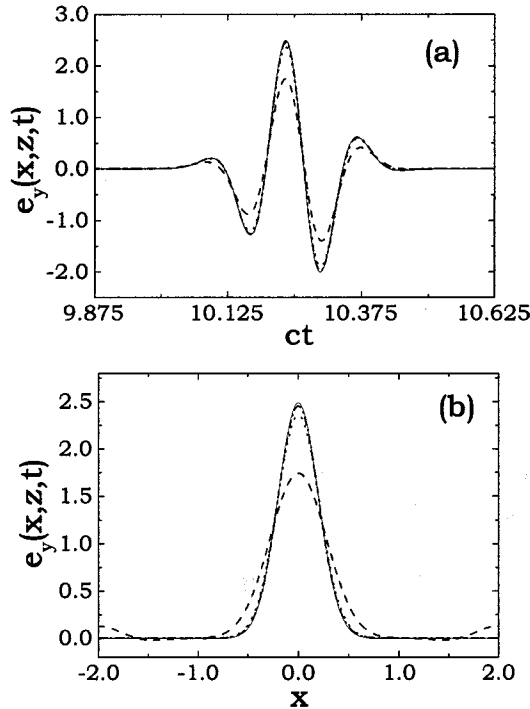


Fig. 11. Radiated field due to cosine-tapered focused aperture distribution in (61), observed in the focal plane. Parameters: $cT_p = 0.5$, $d = 5 = 10cT_p$, $L_f = 10 = 0.2F_d$ (arbitrary units). Reference solution via space-time Kirchhoff integration of (26): solid curves. TD narrow-waisted beam synthesis in (57): dashed and dotted curves. (a) Temporal profile at $x = 0$, $z = 10 = L_f = 0.2F_d$. (b) Spatial transverse profile at $z = 10$, $ct = 10.235$; --- 10 beams ($Q = 0.56$, $\Delta e_y = -9$ dB), 30 beams ($Q = 0.19$, $\Delta e_y = -25$ dB), - - - - 60 beams ($Q = 0.09$, $\Delta e_y = -33$ dB). The r.m.s. errors Δe_y pertain to Fig. 11(a).

sequence of instantaneous snapshots of the radiated field at different times, computed via the space-time Kirchhoff integration in (26), is shown in Fig. 10. One observes how the space-time wavepacket, starting from the initial distribution (61), has a concave (focusing) wavefront on its way to the focal plane, i.e., for $z < L_f$ [Fig. 10(a) and (b)]; it experiences its maximum space-time localization at the focal plane $z = L_f$ [Fig. 10(c) and (d)], and spreads out beyond the focal plane with a convex wavefront [Fig. 10(e) and (f)]. In order to assess the accuracy of the PB expansion in (57), the comparison with the reference solution is shown in Figs. 11 and 12 for representative temporal and spatial cuts at various observation distances. Fig. 11(a) and (b) show the convergence results at the focal plane $z = L_f$ (most challenging test). A slightly finer discretization is needed than in the linear-delay case, for the same degree of accuracy; accurate syntheses ($\Delta e_y < -30$ dB) are obtained for $Q \leq 0.1$, but reasonable accuracy ($\Delta e_y < -20$ dB) is still within the less stringent $Q \leq 0.3$ range. Fig. 12(a)–(d), show results for observation distances smaller ($z = 0.3L_f$) and larger ($z = 3L_f$) than the focal length, with the number of beams chosen so as to assure $Q \leq 0.1$. The spatial profiles in Fig. 12(b) and (d) show the broader spatial extent of the wavefront, with respect to the focal plane pattern in Fig. 11(b), which is analogous to the snapshots in Fig. 10(a) and (e). The temporal profile of the incident pulse in Fig. 12(a) is distorted at [Fig. 11(a)] and beyond [Fig. 12(c)] the focus due to caustic-induced phase shifts.

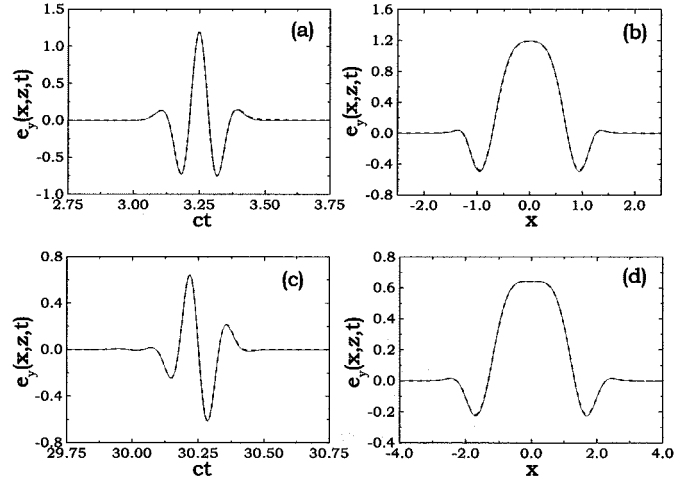


Fig. 12. Parameters as in Fig. 11, observed before and beyond the focal plane. (a) Temporal profile at $x = 0$, $z = 3 = 0.3L_f = 0.06F_d$, 100 beams ($Q = 0.1$, $\Delta e_y = -33$ dB); (b) Spatial transverse profile at $z = 3$, $ct = 3.25$, 100 beams ($Q = 0.1$); (c) Temporal profile at $x = 0$, $z = 30 = 3L_f = 0.6F_d$, 30 beams ($Q = 0.1$, $\Delta e_y = -37$ dB); (d) Spatial transverse profile at $z = 30$, $ct = 30.22$, 30 beams ($Q = 0.1$). Reference solutions (solid curves) and beam solutions (dashed curves) coincide on the scale of the plots.

IV. CONCLUSION

By extending the time-harmonic approach presented in [1], a 2D narrow-waisted pulsed beam algorithm has been presented which allows an effective discretization of short-pulse radiation from one-dimensional (1-D) large apertures. Analytic approximations in the form of readily computable functions have been obtained for aperture field distributions with separable space-time dependence, general phasing, and Gaussian pulse excitation.

Accuracy assessments of the beam algorithm have been formalized theoretically in terms of the nondimensional estimator Q , and calibrated numerically through a variety of simulations involving phased and focused pulsed aperture field profiles; independent calibration of the numerical data was achieved by comparison with a rigorously based brute force space-time Kirchhoff integration. The results are encouraging and confirm that within the stated criteria, one can make fairly reliable predictions of performance.

Overall, it appears that reasonably accurate syntheses can be obtained with a moderate number of beams, within the constraint $Q \leq 0.3$, thus making the algorithm a promising efficient parameterizer for beam tracing in the presence of propagation and scattering environments. Toward this goal, the TD extension of the previously performed FD studies in [11]–[14] is now under consideration [22]. For extension to two-dimensional (2-D) apertures which generate three-dimensional (3-D) (vector) fields, see [23].

APPENDIX A PERTAINING TO (26)

By Fourier transforming the FD radiated field in (2), recalling that G_{2D} and g_{2D} are related through the Fourier transform pair

(27) [18], and applying the convolution theorem [24], the TD field radiated into the $z > 0$ halfspace can be written as

$$e_y(x, z, t) = -\frac{1}{\pi} \int_{-d/2}^{d/2} dx' \frac{\partial}{\partial z} \int_{-\infty}^{t-R/c} \frac{f(x', t')}{\sqrt{(t-t')^2 - (R/c)^2}} dt' \quad (62)$$

with R being defined in (4), and with $f(x, t)$ and $F(x, \omega)$ being related through the Fourier transform pair (27). The differentiation of the temporal integral in (62) can be performed through differentiation under the integral sign, coupled with a limiting procedure to circumvent the singularity at the upper integration limit (see [19] for details), yielding

$$\begin{aligned} \frac{\partial}{\partial z} \int_{-\infty}^{t-R/c} \frac{f(x', t')}{\sqrt{(t-t')^2 - (R/c)^2}} dt' \\ = -\frac{z}{R^2} \int_{-\infty}^{t-R/c} \frac{(t-t')}{\sqrt{(t-t')^2 - (R/c)^2}} \frac{\partial f}{\partial t'}(x', t') dt' \end{aligned} \quad (63)$$

from which (26) follows.

APPENDIX B PERTAINING TO (44)

By substituting (20) and (38), with (22), (39), (40), into (34) one obtains

$$\begin{aligned} b_m^+(x, z, t) = -\frac{2^{-3/4} e^{i3\pi/4} \sqrt{L_x} z T_P^5}{150000\pi \sqrt{c} R_m^{3/2}} \int_0^\infty \omega^{9/2} \\ \times \exp \left[i\omega \left(\frac{i\alpha_m^2 \omega + 100\pi c z_{bm} \beta_m}{200c^2 \pi z_{bm}^2} \right) \right] d\omega \end{aligned} \quad (64)$$

with α_m, β_m defined in (45), (46), respectively. The frequency integral in (64), in its canonical form

$$I(\xi, \zeta) = \int_0^\infty \omega^{9/2} \exp(i\xi\omega - \zeta\omega^2) d\omega, \quad \zeta > 0 \quad (65)$$

can be evaluated in closed form (we used Mathematica [21]) as

$$\begin{aligned} I(\xi, \zeta) = \frac{\zeta^{-13/4}}{2} \left[\sqrt{\zeta} \Gamma \left(\frac{11}{4} \right) {}_1F_1 \left(\frac{11}{4}, \frac{1}{2}, -\frac{\xi^2}{4\zeta} \right) \right. \\ \left. + i\xi \Gamma \left(\frac{13}{4} \right) {}_1F_1 \left(\frac{13}{4}, \frac{3}{2}, -\frac{\xi^2}{4\zeta} \right) \right] \end{aligned} \quad (66)$$

where $\Gamma(\cdot)$ is the gamma function [20, Sec. 6], and ${}_1F_1(u, v, y)$ is the Kummer confluent hypergeometric function [20, Sec. 13]. Equation (44) follows from (64), (66) after straightforward algebra.

APPENDIX C PERTAINING TO (50), (51)

The Kummer confluent hypergeometric function is defined as [20, Sec. 13]

$${}_1F_1(u, v, y) = \sum_{n=0}^{\infty} \frac{(u)_n y^n}{(v)_n n!} \quad (67)$$

where $(u)_n$ is the Pochhammer symbol [20]

$$(u)_n = u(u+1)(u+2)\cdots(u+n-1) \quad (u)_0 = 1. \quad (68)$$

A rapidly convergent approximation for the functions M_1 and M_2 in (48), (49) can be obtained by exploiting the Kummer transformation [20]

$${}_1F_1(u, v, y) = \exp(y) {}_1F_1(v-u, v, -y). \quad (69)$$

Accordingly

$$\begin{aligned} M_1(y) &= {}_1F_1 \left(\frac{11}{4}, \frac{1}{2}, -y^2 \right) \\ &= \exp(-y^2) {}_1F_1 \left(-\frac{9}{4}, \frac{1}{2}, y^2 \right) \\ &= \exp(-y^2) \left(1 - \frac{9}{2} y^2 + \frac{15}{8} y^4 - \frac{y^6}{16} - \frac{3}{896} y^8 \right. \\ &\quad \left. - \frac{y^{10}}{3840} + \mathcal{O}(y^{12}) \right) \end{aligned} \quad (70)$$

$$\begin{aligned} M_2(y) &= {}_1F_1 \left(\frac{13}{4}, \frac{3}{2}, -y^2 \right) \\ &= \exp(-y^2) {}_1F_1 \left(-\frac{7}{4}, \frac{3}{2}, y^2 \right) \\ &= \exp(-y^2) \left(1 - \frac{7}{6} y^2 + \frac{7}{40} y^4 + \frac{y^6}{240} + \frac{y^8}{3456} \right. \\ &\quad \left. + \frac{y^{10}}{42240} + \mathcal{O}(y^{12}) \right) \end{aligned} \quad (71)$$

as shown in (50) and (51).

REFERENCES

- [1] J. J. Maciel and L. B. Felsen, "Systematic study of fields due to extended apertures by Gaussian beam discretization," *IEEE Trans. Antennas Propagat.*, vol. 37, pp. 884–892, July 1989.
- [2] P. D. Einziger and M. Shapira, "Gabor representation and aperture theory," *J. Opt. Soc. Amer. A*, vol. 3, no. 4, pp. 508–522, Apr. 1986.
- [3] B. Z. Steinberg, H. Heyman, and L. B. Felsen, "Phase-space beam summation for time-harmonic radiation from large apertures," *J. Opt. Soc. Amer. A*, vol. 8, no. 1, pp. 41–59, Jan. 1991.
- [4] —, "Phase-space methods for radiation from large apertures," *Radio Sci.*, vol. 26, no. 1, pp. 219–227, Jan.–Feb. 1991.
- [5] B. Z. Steinberg and E. Heyman, "Phase-space beam summation for time-dependent radiation from large apertures: Continuous parameterization," *J. Opt. Soc. Amer. A*, vol. 8, no. 6, pp. 943–958, June 1991.
- [6] —, "Phase-space beam summation for time-dependent radiation from large apertures: Discretized parameterization," *J. Opt. Soc. Am. A*, vol. 8, no. 6, pp. 959–966, June 1991.
- [7] B. Rao and L. Carin, "A hybrid (parabolic equation)-(Gaussian beam) algorithm for wave propagation through large inhomogeneous regions," *IEEE Trans. Antennas Propagat.*, vol. 46, pp. 700–709, May 1998.
- [8] E. Heyman and T. Melamed, "Space-time representation of ultra wideband signals," in *Advances in Imaging and Electron Physics*, P. W. Hawkes, Ed. New York: Academic Press, 1998, vol. 103, pp. 1–63.
- [9] T. Melamed, E. Heyman, and L. B. Felsen, "Local spectral analysis of short-pulse excited scattering from weakly inhomogeneous media. Part I: Forward scattering," *IEEE Trans. Antennas Propagat.*, vol. 47, pp. 1208–1217, July 1999.
- [10] B. Rao and L. Carin, "Beam-tracing-based inverse scattering for general aperture antennas," *J. Opt. Soc. Am. A*, vol. 16, no. 9, pp. 2219–2231, Sept. 1999.

- [11] J. J. Maciel and L. B. Felsen, "Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part I—Plane layer," *IEEE Trans. Antennas Propagat.*, vol. 38, pp. 1607–1617, Oct. 1990.
- [12] —, "Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part II—Circular cylindrical layer," *IEEE Trans. Antennas Propagat.*, vol. 38, pp. 1618–1624, Oct. 1990.
- [13] L. B. Felsen and V. Galdi, "Complex-source-point narrow-waisted ray-like Gaussian beams for frequency and time domain radiation and scattering," in *Ultra-Wideband, Short Pulse Electromagnetics 5*, S. R. Cloude and P. D. Smith, Eds, NY: Kluwer/Plenum, 2001.
- [14] V. Galdi, L. B. Felsen, and D. A. Castañón, "Quasi-ray Gaussian beam algorithm for time-harmonic two-dimensional scattering by moderately rough interfaces," *IEEE Trans. Antennas Propagat.*, vol. 49, pp. 1305–1314, Sept. 2001.
- [15] V. Galdi, D. A. Castañón, and L. B. Felsen, "Multifrequency reconstruction of moderately rough interfaces via quasi-ray Gaussian beams," *IEEE Trans. Geosci. Remote Sensing*, Sept. 2000, submitted for publication.
- [16] M. J. Bastiaans, "Gabor's expansion of a signal into Gaussian elementary signals," *Proc. IEEE*, vol. 68, pp. 538–539, Apr. 1980.
- [17] —, "A sampling theorem for the complex spectrogram, and Gabor's expansion of a signal in Gaussian elementary signals," *Opt. Eng.*, vol. 20, pp. 594–598, 1981.
- [18] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*. New York: McGraw-Hill, 1953.
- [19] M. Kragalott, M. S. Kluskens, and W. P. Pala, "Time-domain fields exterior to a two-dimensional FDTD space," *IEEE Trans. Antennas Propagat.*, vol. 45, pp. 1655–1663, Nov. 1997.
- [20] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1964.
- [21] S. Wolfram, *The Mathematica Book*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 1996.
- [22] V. Galdi, L. B. Felsen, and D. A. Castañón, "Quasi-ray Gaussian beam algorithm for short-pulse two-dimensional scattering by moderately rough interfaces," *IEEE Trans. Antennas Propagat.*, Feb. 2001, submitted for publication.
- [23] —, "Time-domain radiation from large two-dimensional apertures via narrow-waisted Gaussian beams," *IEEE Trans. Antennas Propagat.*, Jan. 2001, submitted for publication.
- [24] A. Papoulis, *Signal Analysis*. New York: McGraw-Hill, 1977.



Vincenzo Galdi (M'98)

He received the Laurea degree (*summa cum laude*) in electrical engineering and the Ph.D. degree in applied electromagnetics from the University of Salerno, Italy, in 1995 and 1999, respectively.

From April to December 1997, he held a Visiting Position in the Radio Frequency Division of the European Space Research & Technology Centre (ESTEC-ESA), Noordwijk, The Netherlands, where he was involved in developing CAD tools for microwave filters and phased-array antennas with coaxial excitation. In

September 1999, he received a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. In October 1999, he became a Research Associate in the Department of Electrical and Computer Engineering, Boston University, Boston, MA, where he is currently working on wave-oriented imaging algorithms for landmine detection and classification. His research interests include analytical and numerical techniques for wave propagation in complex environments, path integrals, and stochastic resonance.

Dr. Galdi is the recipient of the 2001 International Union of Radio Science (URSI) "Young Scientist Award." He is a member of Sigma Xi.



Leopold B. Felsen (S'47–M'54–SM'55–F'62–LF'90) was born in Munich, Germany, on May 7, 1924. He received the B.E.E., M.E.E., and D.E.E. degrees from the Polytechnic Institute of Brooklyn, Brooklyn, NY, in 1948, 1950, and, 1952, respectively.

He emigrated to the United States in 1939 and served in the U.S. Army from 1943 to 1946. After 1952, he remained with the Polytechnic Institute of Brooklyn (now Polytechnic University), becoming a Professor in 1962 and University Professor in 1978. From 1974 to 1978, he was Dean of Engineering.

In 1994, he resigned from the full-time Polytechnic Faculty and became University Professor Emeritus. He is now a part-time Professor of Aerospace and Mechanical Engineering and Professor of Electrical and Computer Engineering at Boston University, Boston, MA. He is the author or coauthor of more than 300 papers and several books, including *Radiation and Scattering of Waves* (Englewood Cliffs, NJ: Prentice-Hall, 1973; Piscataway, NJ: IEEE Press, 1994). He is an Associate Editor of several professional journals and an editor of the *Wave Phenomena Series* (New York: Springer-Verlag). His research interests encompass wave propagation and diffraction in complex environments and in various disciplines, high-frequency asymptotic and short-pulse techniques, and phase-space methods with an emphasis on wave-oriented data processing and imaging.

Dr. Felsen is a member of Sigma Xi and a Fellow of the Optical Society of America and the Acoustical Society of America. He has held named Visiting Professorships and Fellowships at universities in the United States and abroad, including the Guggenheim in 1973 and the Humboldt Foundation Senior Scientist Award in 1981. In 1974, he was an IEEE Antennas and Propagation Society (APS) Distinguished Lecturer. He received the Balthasar van der Pol Gold Medal from the International Union of Radio Science (URSI) in 1975, an honorary doctorate from the Technical University of Denmark in 1979, the IEEE Heinrich Hertz Gold Medal for 1991, the APS Distinguished Achievement Award for 1998, the IEEE Third Millennium Medal in 2000 (nomination by APS), three Distinguished Faculty Alumnus Awards from Polytechnic University, and an IEEE Centennial Medal in 1984. He also has received Best Paper awards for several papers in which he was an author or coauthor. In 1977, he was elected to the National Academy of Engineering. He served on the APS Administrative Committee from 1963 to 1966 and as Vice Chairman and Chairman for both the United States (1966–1973) and the International (1978–1984) URSI Commission B.

David A. Castañón (S'68–M'79–SM'98) received the B.S. degree in electrical engineering from Tulane University, New Orleans, LA, in 1971 and the Ph.D. degree in applied mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1976.

From 1976 to 1981, he was a Research Associate with the Laboratory for Information and Decision Systems, MIT. From 1982 to 1990, he was Senior and Chief Research Scientist at Alphatech, Inc., Burlington, MA. Since 1990, he has been a Professor in the Department of Electrical and Computer Engineering, Boston University, Boston, MA. His research interests include stochastic control and estimation, optimization, and image processing.

Dr. Castañón served as a member of the Board of Governors of the IEEE Control Systems Society. He is also a member of AMS, SIAM, and INFORMS.

Moderately Rough Dielectric Interface Profile Reconstruction via Short-Pulse Quasi-Ray Gaussian Beams

Vincenzo Galdi, *Member, IEEE*, Julia Pavlovich, W. Clem Karl, *Senior Member, IEEE*, David A. Castañon, *Senior Member, IEEE*,
and Leopold B. Felsen, *Life Fellow, IEEE*

This work was supported by ODDR&E under MURI Grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from Grant No. 9900448 by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University, Brooklyn, NY 11201 USA.

V. Galdi is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, on leave from University of Sannio, Benevento, Italy (e-mail: vgaldi@bu.edu).

J. Pavlovich, W.C. Karl and D.A. Castañon are with the Multi-Dimensional Signal Processing Laboratory, Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA (e-mail: oska@bu.edu, dac@bu.edu, wckarl@bu.edu).

L.B. Felsen is with the Department of Aerospace and Mechanical Engineering and the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, part-time. He is also University Professor Emeritus, Polytechnic University, Brooklyn, NY 11201 USA (e-mail: lfelsen@bu.edu).

Abstract

A new technique for estimating the coarse-scale profile of a moderately rough interface between air and a homogeneous dielectric half-space is presented. The proposed approach is based on space-time sparsely sampled reflected field observations and uses a quasi-ray Gaussian beam fast forward model, coupled with a compact parameterization of the surface profile in terms of B-splines, from which the profile estimation problem is posed as a nonlinear optimization problem. Numerical experiments are presented to assess accuracy, reliability and computational efficiency. The proposed approach finds application in adaptive schemes for rough surface underground imaging of shallowly buried targets via ultra-wideband ground penetrating radars.

Keywords

GPR, imaging, rough surfaces, Gaussian beams, short pulses.

I. INTRODUCTION

In ground penetrating radar (GPR) applications, the twice-traversed unknown rough interface separating air and soil acts as a major source of clutter by distorting the interrogating signal on its way to and from the targets of interest, and by generating complicated backscattered field patterns which may obscure the useful signals. Physics-based modeling of such clutter, which could significantly enhance the ultimate GPR performance, poses challenging problems from both the electromagnetic (EM) and signal processing viewpoints. Standard statistical approaches, which tend to model such distortion as additive colored Gaussian noise, perform reasonably well in *detection* problems [1], [2]. However, they have been found to yield limited accuracy and reliability in underground imaging techniques for target *localization* and *classification* (see, e.g., [3]), for which alternative approaches need to be explored. In this connection, we have been investigating a novel *adaptive* framework, based on *quasi-deterministic* compensation of the coarse-scale roughness effect. This approach is based on *prior* estimation of the coarse-scale roughness profile, which is accomplished by exploiting sparse reflected field observation data and fast forward scattering models. In this communication, we address this problem for the case of short-pulse illumination, typical of current ultra-wideband (UWB) GPR systems. The proposed approach is built on recently developed Gabor-based narrow-waisted *quasi-ray* Gaussian beam (GB) algorithms for short-pulse scattering from moderately rough dielectric interfaces [4]. By exploiting these fast forward models and a low-dimensional spline

interface parameterization, together with the (usually small) separation between the rough interface and the target, the prior surface estimation is posed as a nonlinear optimization problem by fitting the model-based prediction to the available *early-time* observation data. The subsequent problem of quasi-deterministic compensation and underground imaging via *late-time* response processing is addressed in a separate paper [5], with particular reference to shallowly buried plastic mine-like targets.

II. STATEMENT OF THE PROBLEM

We consider the two-dimensional (2D) problem geometry sketched in the (x, z) coordinate space of Fig. 1, where all quantities and fields are assumed to be y -independent. A homogeneous dielectric half-space (soil) of *known* relative permittivity ϵ_{r1} and conductivity σ_1 , bounded by a moderately rough interface with profile $z = h(x)$, is illuminated by a y -directed (TM-polarized) pulsed well-collimated electromagnetic (EM) field, generated by a large truncated aperture field distribution of width d at $z = z_A$, which is assumed to be well approximated by a pulsed truncated tapered plane wave,

$$e^i(\mathbf{r}, t) \sim g(x_B)p(t - c^{-1}z_B), \quad (1)$$

where $\mathbf{r} \equiv (x, z)$, c is the free-space wavespeed, $p(t)$ is a short pulse of length $T \ll d/c$, and (x_B, z_B) are beam centered coordinates,

$$\begin{bmatrix} x_B \\ z_B \end{bmatrix} = \begin{bmatrix} \cos \theta_A & \sin \theta_A \\ \sin \theta_A & -\cos \theta_A \end{bmatrix} \begin{bmatrix} x - x_A \\ z - z_A \end{bmatrix}. \quad (2)$$

In (1) and (2), $g(x_B)$ is a spatial taper, while θ_A and x_A denote the tilt angle of the radiated beam relative to the z -axis and its spatial displacement, respectively; parameters are chosen so that the illumination tapers to zero for $|x| \leq d/2$ and vanishes for $|x| > d/2$ (Fig. 1). At this stage of investigation, we ignore the presence of buried targets, which is dealt with elsewhere [5]. Furthermore, we also neglect the *noisy* (incoherent) contribution of finer-scale roughness, and focus on estimating the coarse scale roughness profile $h(x)$ from sparse reflected field observations. In [6], we addressed this problem for narrow-band stepped-frequency sparse observations. This approach is extended here to *pulsed* excitation. To proceed, the y -directed reflected electric field is sampled at N_t time instants

at N_r fixed receiver locations $x_1^r, \dots, x_{N_r}^r$ at the observation plane $z = z^r$; the *known term* in the problem is given by the set of $N_r \times N_t$ samples. In our numerical experiments in Sec. V, we shall use synthetic field observation data generated via a full-wave solution of the forward scattering problem (see [4] for details).

III. ROUGH SURFACE FORWARD SCATTERING MODEL

The forward scattering model, detailed in [4], is based on the Kirchhoff Physical Optics (PO) approximation in conjunction with the Gabor-based narrow-waisted pulsed beam (PB) discretization of 1D aperture field distributions in [7], and is restricted to moderate roughness (both in height and slope) and slightly lossy soils.

The PO “equivalent current”, which generates the reflected field, is first parameterized in the frequency domain in terms of x -domain discretized m -indexed Gabor basis functions with narrow width L , centered on the Gabor lattice points $x_m = mL$; these initial conditions generate narrow-waisted, quasi-ray, complex-source-point GBs propagating along the local reflection directions. For Rayleigh (i.e., differentiated Gaussian) pulses, the resulting time-domain analytic Fourier inversion integral can be approximated by rapidly computable closed form expressions, yielding the following approximate PO-PB expansion for the reflected field e^r (see [4] for details),

$$e^r(\mathbf{r}, t) \sim \sum_{|m| \leq (d/2L)} c_m^r b_m^r(\mathbf{r}, t - t_m), \quad (3)$$

where the Gabor expansion coefficients c_m^r and the time delays t_m are approximated by sampling the PO equivalent currents at the lattice points $x_m = mL$, and the PB propagators b_m^r are expressed in terms of rapidly computable confluent hypergeometric functions [4, Sec. IV-B]. The approximate forward scattering model in (3) has been validated and calibrated against a rigorous reference solution [4, Sec. V-A], and has been found to provide accurate and robust predictions for moderate roughness (both in height and slope), non-grazing incidence, and slightly lossy soils (see [4, Sec. V-C] for details). Numerical convergence is usually achieved with $d/L \sim 30$ to 100 (narrow-waisted) PBs, resulting in minimal storage requirements and typical computing times of 2 ms per space-time field sample on a 700 MHz PC. Moreover, for computing a number of field time samples at

fixed receiver locations, as required in Sec. IV, one can take advantage of the structure of the PB propagators in (3) (see [4, Sec. IV-B] for details) to compute the time-independent parts only once, with resulting typical computing times of 20 ms for a *whole* 100-sample waveform, as compared with about 240 secs. required by our full-wave reference solution. This light computational burden, is essential for the overall computational feasibility of the proposed surface estimation approach.

IV. SURFACE PROFILE ESTIMATION

Due to the inherent ill-posedness of the surface estimation problem, it is necessary to exploit *regularization* strategies. Acknowledging the implicit limits of retrievable information through inverse scattering, our regularization strategy is based on a compact parameterization of the unknown profile function in a suitable finite-dimensional space. As in [6], we model the surface profile $h(x)$ using a quartic B-spline parameterization [8] with fixed resolution matched to the coarse level of detail for the reconstruction,

$$h(x) = \sum_{n=-4}^{N_h-1} h_n s^{(4)}(x - x_{min} + n\Delta_x), \quad x_{min} \leq x \leq x_{max}, \quad (4)$$

whereby the coefficients $\{h_n\}$ become the *unknowns* of the problem. In (4), $s^{(4)}(\cdot)$ is a standard quartic B-spline basis function [8] (see also [6, Eq. (5)]) with finite spatial support $5\Delta_x$, where $\Delta_x = (x_{max} - x_{min})/N_h$ denotes a scale parameter whose choice can be guided by possible prior information and refined *adaptively* (see Sec. V). To prevent ill-conditioning, it is crucial that the number of unknown parameters to be retrieved does not substantially exceed the *essential* dimension of the observation data set. For the problem of interest in this paper, however, the intrinsic limitation in the retrievable information does not pose serious *quantitative* restrictions, since our implementation allows the number of unknowns to be kept fairly small ($N_h \lesssim 30$). By exploiting our PO-PB fast forward model in [4] and the spline interface profile parameterization in (4), the well-posedness of the problem is restored by estimating the unknown parameters via *fitting* the model-based forward scattering prediction to the available observation data, i.e., minimizing a suitable *cost functional*. We use a simple least square formulation,

$$J(\underline{h}) = \|\underline{\mathbf{e}}^r(\underline{h}) - \hat{\underline{\mathbf{e}}}^r\|^2 = \sum_{p=1}^{N_r} \sum_{q=1}^{N_t} \gamma_p \left(e_{pq}^r - \hat{e}_{pq}^r \right)^2, \quad (5)$$

where $\hat{e}_{pq}^r = \hat{e}^r(\mathbf{r}_p^r, t_{pq})$ denotes the y -directed reflected field *observed* at time $t_{pq} \in [T_p^{(on)}, T_p^{(off)}]$ at receiver locations $\mathbf{r}_p^r \equiv (x_p^r, z^r)$ (Fig. 1), $e_{pq}^r = e^r(\mathbf{r}_p^r, t_{pq}; \underline{h})$ denotes the corresponding *forward prediction* for the surface profile coefficients $\underline{h} = \{h_{-4}, \dots, h_{N_h-1}\}$, and γ_p are normalization coefficients. Anticipating the possible presence of buried targets, dealt with in [5], the observation intervals $[T_p^{(on)}, T_p^{(off)}]$ are chosen so as to gate out the *late-time* response (i.e., causal contributions from regions beyond a critical depth), in order to prevent any possible bias in the surface profile estimation.

In general, the predictive forward model $e^r(\mathbf{r}_p^r, t_{pq}; \underline{h})$ is a nonlinear function of the coefficients \underline{h} . Therefore, the cost functional in (5) is generally *non-convex* with respect to \underline{h} and may have multiple local minima which correspond to *false solutions*. Standard descent-based optimization techniques (e.g., conjugate gradient [9]) can be trapped unless an accurate initial guess is available. For the narrow-band step-frequency configuration in [6], the smoothness of the cost functional was found to be essentially dependent on the frequency content of the excitation field, and a multi-resolution *frequency-hopping* strategy [10] was devised to achieve the global optimization (see [6, Sec. III-C] for details). The same guidelines can be exploited for the pulsed excitation of interest here, with the pulse length cT now playing the key role. In particular, *short* pulses are desirable to enhance resolution and accuracy in the reconstruction, but an exceedingly wide-band excitation would most likely yield a highly non-convex cost functional with many local minima, whose global minimization could become computationally unfeasible. In our numerical investigation, we found that values of $cT \lesssim 0.2d$ tend to ensure local minima in the cost functional, whereas for $cT \gtrsim 0.5d$, the achieved resolution deteriorates. Moreover, for the eventual underground imaging problem of interest (see [5]), it is also essential to achieve adequate soil penetration, and therefore operate at sufficiently low frequencies. In principle, one could gain more flexibility by using different sets of pulses for the surface estimation (e.g., implementing multi-resolution optimization strategies as in [6]) and for the underground imaging, at the expense of hardware complexity. However, in our numerical experiments, we found that a *single* pulse in the typical UWB GPR frequency range (e.g., $cT \sim 0.4d$,

i.e., $T \sim 1.3\text{ns}$ for $d = 1\text{m}$) can be used for *both* surface estimation *and* underground imaging of shallowly buried targets (see [5]), thus having the potential of yielding a reasonable trade-off between the above contrasting requirements.

V. NUMERICAL RESULTS

As mentioned earlier, the needed reflected field observation data in (5) were simulated via a reliable full-wave solution of the forward scattering problem, based on the time-harmonic multifilament current method in [11] and the fast Fourier transform (see [4, Sec. V-A] for details). Forward predictions in (5) were generated via the PO-PB model in (3) with $d/L = 40$ beams, for which the accuracy was preliminarily verified.

We begin with a simplified configuration where we use as a template for inversion *the same* spline model that was used for *generating* the actual profile. In other words, we assume *a priori* knowledge of the scale parameter Δ_x in (4) (i.e., the number of B-spline basis functions), and focus on retrieving the unknown coefficients h_n only. This somewhat unrealistic assumption will be removed subsequently via an iterative *adaptive* framework. In all simulations below, a cosine-tapered normally-incident excitation was used with $g(x) = \cos(\pi x/d)$, $\theta_A = x_A = 0$, and a single fourth-order Rayleigh pulse with $cT = 0.4d$ (Fig. 2a), which was found to provide a good compromise between resolution and smoothness in the cost functional. An observation time window $[T_p^{(on)}, T_p^{(off)}]$ with $cT_p^{(on)} = 0.3d$ and $T_p^{(off)} = 0.8d$ was used, so as to roughly gate out scattering contributions from possible targets buried deeper than $\sim 8\text{cm}$ below nominal ground ($z = 0$). The resulting cost functional in (5) was minimized via the Polak-Ribiere version of the conjugate gradient (CG) algorithm (particularly suited for non-quadratic functions [9]). The needed gradient of J was computed using a central difference formula, resulting in $2N_h + 8$ functional evaluations (i.e., $2N_h + 8$ solutions of a forward scattering problem), $N_h + 4$ being the number of unknown spline coefficients in (4). Loose *a priori* knowledge was exploited by using as an initial guess a flat interface at $z = 0$ (i.e., $\underline{h} = \underline{0}$) and restricting the surface profile search to $\pm 8\text{cm}$ around it.

A typical reconstruction example is shown in Fig. 2b. The surface profile realization was generated using the quartic spline model in (4) with random coefficients. Although no specific roughness model (e.g., Gaussian) was simulated, geometric and constitutive

parameters were selected so as to mimic natural moderate roughness with maximum-to-minimum height ~ 4 cm and maximum slope $\sim 32^\circ$ for a class of realistic soils [12]. The reconstruction is reasonably accurate throughout most of the interval, except near the edges of the illuminated region. The likely explanation for this loss of accuracy, also observed in [6], is due to the weak illumination in these regions, attributed to the aperture field (cosine) tapering which was, however, required to avoid numerical artifacts (edge effects). This kind of accuracy was observed in many numerical experiments, with numerical convergence of the minimization algorithm typically achieved within ~ 30 CG iterations (see Fig. 2c), resulting in computing times of ~ 55 secs.

In order to assess the reliability of the surface profile estimation algorithm, we performed a number of sensitivity tests with respect to possible uncertainty in the prior knowledge as well as corruption in the observed data. Figure 3 displays typical reconstructions obtained by introducing a $\pm 10\%$ error in the values of the soil parameters ϵ_{r1} and σ_1 used in the forward scattering model, and by using observation data corrupted by a $\pm 10\%$ uniform noise. As one can see, the surface estimation algorithm turns out to be remarkably robust.

Finally, to remove the unrealistic perfectly-matched-template assumption, we now invoke an adaptive iterative framework for estimating the surface profile in Fig. 2b, which was generated via the B-spline model in (4) with $N_h = 16$, by postulating a *mismatched* spline template (i.e., $N_h \neq 16$). Referring to Fig. 4, the procedure is started with a tentative initial coarse parameterization ($N_h = 6$) and the usual flat-interface ($z = 0$) initial guess for the profile, which leads to the corresponding gross-scale reconstruction in Fig. 4a. The resolution is gradually increased, utilizing at each stage the previous stage reconstruction as the initial guess. Figures 4b-d, for instance, show the iterative improvements obtained via the sequence $N_h = 12, 18, 24$ (deliberately chosen so as to avoid the perfectly matched case $N_h = 16$), with the spline parameterization progressively tuned so as to capture the various details in the surface profile. The accuracy in the last stage reconstruction (Fig. 4d) is comparable to that achieved with a perfectly matched template (Fig. 2b). In this example, a pragmatic stopping criterion was used, based on the (in)sensitivity of the cost functional with respect to further increases in N_h (see Fig. 4 caption). More systematic approaches based, e.g., on the *minimum description length*

principle, can be found in [13]. Also, more flexible *multi-scale* parameterizations can be exploited in principle to deal with more complex surface profiles, but this is outside the scope of the present paper.

VI. CONCLUSIONS

In this communication, we have presented a novel inversion algorithm for the reconstruction of moderately rough dielectric interfaces using space-time sparsely sampled reflected field data. The proposed algorithm was found to provide fast, accurate and robust estimations for moderate roughness ($\sim 4\text{cm}$ maximum-to-minimum, maximum slopes $\lesssim 30^\circ$), even for noisy data and with imperfect knowledge of soil parameters. These results lay the foundation for the adaptive techniques for subsurface GPR image reconstruction of shallowly buried plastic mine-like targets in the presence of rough air-soil interfaces addressed in [5]. Extensions presently under investigation include generalization to fully 3D geometries.

REFERENCES

- [1] T. Dogaru, L. Collins, and L. Carin, "Optimal time-domain detection of a deterministic target buried under a randomly rough interface," *IEEE Trans. Antennas Propagat.*, vol. 49, No. 3, pp. 313-326, Mar. 2001.
- [2] H. Zhan, C.M. Rappaport, M. El-Shenawee, and E.L. Miller, "Mine detection under rough ground surfaces using 2-D FDTD modeling and hypothesis testing," Proc. 2001 IEEE Antennas Propagat. Int. Symposium, Boston, MA, USA, July 8-13, 2001, vol. 3, p. 756.
- [3] H. Feng, D. A. Castañon, W. C. Karl, and E. L. Miller, "GPR imaging approaches for buried plastic landmine detection," in *Detection and Remediation Technologies for Mines and Minelike Targets V*, A.C. Dubey, J.F. Harvey, J.T. Broach, E.R. Dugan, Eds., *Proc. SPIE*, vol. 4038, pp. 1485-1496, Aug. 2000.
- [4] V. Galdi, L. B. Felsen, and D. A. Castañon, "Quasi-ray Gaussian beam algorithm for short-pulse two-dimensional scattering by moderately rough dielectric interfaces," submitted to *IEEE Trans. Antennas Propagat.*, Apr. 2001.
- [5] V. Galdi, H. Feng, D. A. Castañon, W.C. Karl, and L. B. Felsen, "Moderately rough surface underground imaging via short-pulse quasi-ray Gaussian beams," submitted to *IEEE Trans. Antennas Propagat.*, Oct. 2001.
- [6] V. Galdi, D.A. Castañon, and L.B. Felsen, "Multifrequency reconstruction of moderately rough interfaces via quasi-ray Gaussian beams," to be published in *IEEE Trans. Geosci. Remote Sensing*, 2002.
- [7] V. Galdi, L. B. Felsen, and D. A. Castañon, "Narrow-waisted Gaussian beam discretization for short-pulse radiation from one-dimensional large apertures," *IEEE Trans. Antennas Propagat.*, Vol. 49, No. 9, pp. 1322-1332, Sept. 2001.
- [8] L.L. Shumaker, *Spline Functions: Basic Theory*. New York (NY): Wiley, 1981.

- [9] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge (UK): Cambridge Univ. Press, 1992.
- [10] W.C. Chew and J.H. Lin, "A frequency-hopping approach for microwave imaging of large inhomogeneous bodies," *IEEE Microwave Guided Wave Lett.*, vol. 5, pp. 439-441, 1995.
- [11] Y. Leviatan and A. Boag, "Analysis of electromagnetic scattering from dielectric cylinders using a multifilament current model," *IEEE Trans. Antennas Propagat.*, vol. 35, No. 10, pp. 1119-1127, Oct. 1987.
- [12] J.E. Hipp, "Soil electromagnetic parameters as functions of frequency, soil density, and soil moisture," *Proc. IEEE*, vol. 62, pp. 98-103, Jan. 1974.
- [13] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, Series in Computer Science, vol. 15. Singapore: World Scientific, 1989.

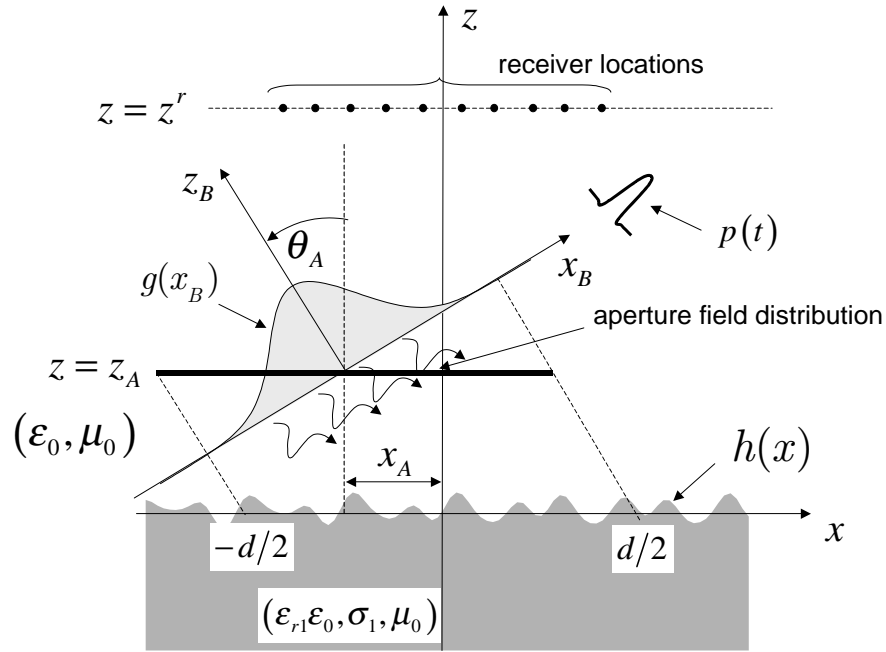


Fig. 1. Problem geometry. An aperture-generated, quasi-plane-wave, TM-polarized pulsed field impinges from free-space onto a dielectric half-space with known relative permittivity ϵ_{r1} and conductivity σ_1 , bounded by a moderately rough interface profile $z = h(x)$. The reflected field is sampled at N_t time instants at N_r fixed receiver locations $x_1^r, \dots, x_{N_r}^r$ at the observation plane $z = z^r$.

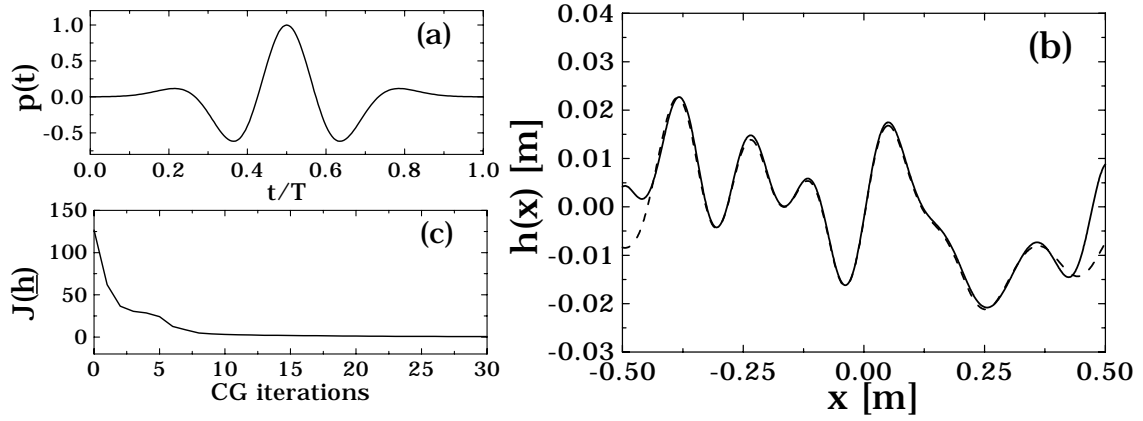


Fig. 2. Rough surface profile reconstruction. Geometry as in Fig. 1, with $g(x) = \cos(\pi x/d)$, $d = 1\text{m}$, $\theta_A = 0$, $x_A = 0$, $z_A = 0.1\text{m}$. The rough surface profile realization was randomly generated via the spline model in (4) with $x_{min} = -0.55\text{m}$, $x_{max} = 0.55\text{m}$, $N_h = 16$, so as to simulate typical moderate roughness (maximum height $\sim 4\text{cm}$ maximum-to-minimum, maximum slope $\sim 32^\circ$) for a class of realistic soils ($\epsilon_{r1} = 4$, $\sigma_1 = 0.01\text{ S/m}$). For the surface profile estimation, the reflected field is sampled at $N_r = 11$ receivers and $N_t = 50$ time instants at $z^r = 0.3\text{m}$ and $x_p^r = -0.5\text{m}, -0.4\text{m}, \dots, 0.5\text{m}$, with $cT_p^{(on)} = 0.3d$ and $cT_p^{(off)} = 0.8d$, $p = 1, \dots, N_r$. (a): Fourth-order Rayleigh pulsed excitation $p(t)$ ($cT = 0.4d$, i.e., $T \sim 1.3\text{ns}$). (b): — Actual profile; - - - Reconstruction; (c): Corresponding cost function in (5) vs. number of conjugate gradient (CG) iterations.

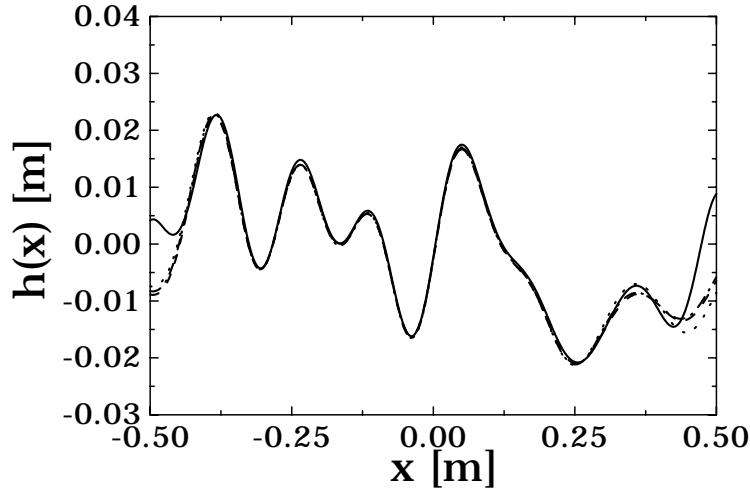


Fig. 3. Parameters as in Fig. 2. Surface profile reconstruction examples with uncertain or corrupted data. — Actual profile; - - - Reconstruction with +10% error in ϵ_{r1} , σ_1 ; Reconstruction with -10% error in ϵ_{r1} , σ_1 ; -.-.- Reconstruction with observation data corrupted by a $\pm 10\%$ uniform noise.

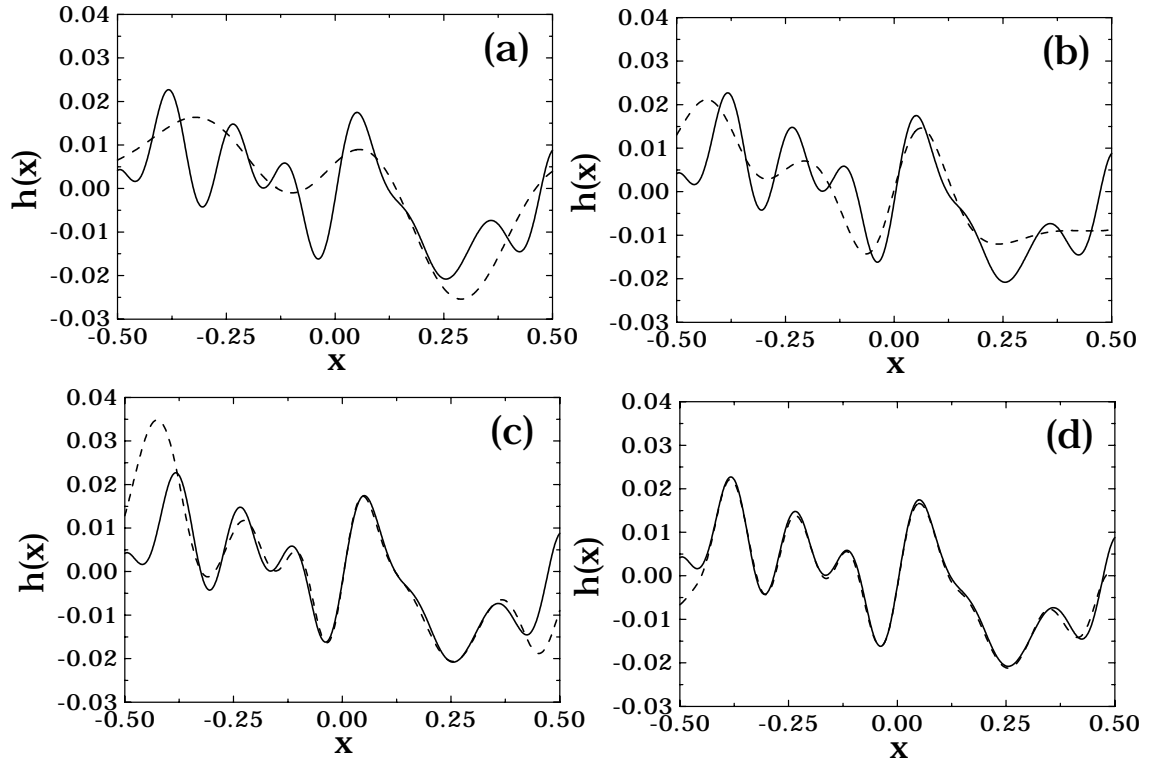


Fig. 4. Parameters as in Fig. 2. Example of iterative adaptive spline parameterization. (a): Reconstruction using a tentative initial coarse parameterization ($N_h = 6$) and a flat interface initial guess (number of CG iterations: $N_{CG} = 8$; final value of cost function: $J^{(min)} = 70.4$); (b), (c), (d): Refinements obtained by progressively increasing the number of basis functions, $N_h = 12, 18, 24$, respectively, and using the previous stage reconstruction as initial guess ($N_{CG} = 12, 12, 23$, respectively; $J^{(min)} = 26, 3.7, 0.53$, respectively). — Actual profile; - - - Reconstruction.

Moderately Rough Surface Underground Imaging via Short-Pulse Quasi-Ray Gaussian Beams

Vincenzo Galdi, *Member, IEEE*, Haihua Feng, David A. Castañón, *Senior Member, IEEE*, W. Clem Karl, *Senior Member, IEEE*,
and Leopold B. Felsen, *Life Fellow, IEEE*

This work was supported by ODDR&E under MURI Grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from Grant No. 9900448 by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University, Brooklyn, NY 11201 USA.

V. Galdi is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, on leave from University of Sannio, Benevento, Italy (e-mail: vgaldi@bu.edu).

H. Feng, D.A. Castañón and W.C. Karl are with the Multi-Dimensional Signal Processing Laboratory, Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA (e-mail: hfeng@bu.edu, dac@bu.edu, wckarl@bu.edu).

L.B. Felsen is with the Department of Aerospace and Mechanical Engineering and the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA, part-time. He is also University Professor Emeritus, Polytechnic University, Brooklyn, NY 11201 USA (e-mail: lfelsen@bu.edu).

Abstract

An adaptive framework is presented for ultra-wideband ground penetrating radar imaging of shallowly-buried low-contrast dielectric objects in the presence of a moderately rough air-soil interface. The proposed approach works with sparse data and relies on recently developed Gabor-based narrow-waisted Gaussian beam algorithms as fast forward scattering predictive models. First, a nonlinear inverse scattering problem is solved to estimate the unknown coarse-scale roughness profile. This sets the stage for adaptive compensation of clutter-induced distortion in the underground imaging problem, which is linearized via Born approximation and subsequently solved via various *pixel-based* and *object-based* techniques. Numerical simulations are presented to assess accuracy, robustness and computational efficiency. The proposed approach has potential applications to antipersonnel land mine remediation.

Keywords

Ground penetrating radars, rough surfaces, Gaussian beams, short pulses, total variation, curve evolution.

I. INTRODUCTION

The problem of detecting and localizing buried objects via ground penetrating radar (GPR) has received considerable attention in recent years, with potential applications ranging from non-destructive testing to environmental remediation. In this context, one of the most important and difficult applications is related to the clearance of buried unexploded ordinance, such as plastic antipersonnel land mines [1]. Attacking such a problem requires a judicious blend of efficient physics-based modeling and advanced signal processing. The associated electromagnetic (EM) model involves wave propagation in inhomogeneous, lossy, dispersive media, and “near-field” scattering from irregularly-shaped objects and interfaces. From a signal processing perspective, the main challenges are related to the limited-viewing geometry, the low permittivity contrast of the target with respect to the background soil (i.e., low signal-to-noise ratio), the imperfect knowledge of geometric and constitutive properties of soil and targets, and the possible presence of clutter which obscures the useful signals. In particular, the effect of the rough air-ground interface, in terms of backscattered clutter and distortion introduced into the interrogating signal on its way to and from the targets of interest, is a key issue from both the EM and signal processing perspectives. In this connection, standard approaches tend to be *fully statistical*,

modeling such effects as additive colored Gaussian noise whose features are typically estimated via full-wave Monte Carlo simulations with many random soil realizations. These approaches have been explored in [2]–[5], yielding reasonably good results for the *detection* problem in the presence of small roughness. However, their accuracy and reliability turn out to be not completely satisfactory in underground imaging algorithms aimed at *localizing* and possibly *classifying* a target (see, e.g., [6]).

In an ongoing series of recent investigations, so far restricted to two-dimensional (2D) geometries and moderate roughness, we have been exploring a novel *adaptive* framework, based on prior estimation of the unknown coarse scale roughness profile and subsequent compensation for the deterministic features of the related clutter. In a stepwise approach toward constructing the necessary tools, in both the frequency (FD) and time (TD) domains, we first developed physics-based fast forward algorithms for scattering from, and transmission through, moderately rough dielectric interfaces, generalizing previously developed Gabor-based narrow-waisted Gaussian beam (GB) methods [7]–[9]. The resulting models [10], [11] have subsequently been incorporated in inverse scattering scenarios aimed at enhancing the underground imaging by compensating for the coarse scale deterministic features of the roughness profile. In particular, FD and TD surface estimation algorithms working with sparse data have been explored in [12] and [13], respectively. These techniques were found to provide reasonably accurate and robust estimations, even with noisy data and imperfect knowledge of soil parameters, requiring modest computational effort. Applications to frequency-stepped GPR subsurface imaging have been explored in [14].

The present paper is concerned with the application of the Gabor-based Physical-Optics Pulsed-Beam (PO-PB) forward scattering model in [11] and the surface estimation technique in [13] to the problem of imaging of shallowly buried plastic mine-like targets in the presence of rough air-soil interfaces, working with TD sparse data. Many GPR systems operate with short-pulse excitation. Although one could still process the data in the FD via Fourier transform, a *direct* approach in the TD offers potential advantages, including the possibility of carrying out *selective* imaging via suitable *time-windowing* of the data. Accordingly, in our problem, we shall exploit the (usually small) separation between the rough interface and the target by performing first the surface estimation via early-time

response processing (as discussed in [13]), with subsequent compensation for the roughness effect and eventual target imaging via late-time response processing. In this connection, we shall utilize the weak-scattering Born approximation to linearize the inversion problem, and explore various regularization techniques to cope with its inherent ill-posedness.

The layout of the paper is as follows. Section II introduces the rough surface underground imaging problem with TD GPR sparse data. Section III gives a short review of the surface estimation algorithm in [13]. Section IV describes the linearized forward scattering model in the presence of a *known* rough interface, and its narrow-waisted GB implementation. Section V addresses the underground imaging inversion problem and the various regularization techniques explored. Section VI details the outcomes of numerical experiments to assess the accuracy and reliability of the proposed approach, as well as its computational efficiency. Conclusions follow in Section VII. Figure 1 shows a schematic flow-chart of the proposed framework, coordinated with the paper layout, which illustrates how the various models and techniques described in the paper are combined to attack the end-to-end problem.

II. STATEMENT OF THE PROBLEM

The problem geometry is sketched in Fig. 2: in an (x, z) two-dimensional (2D) coordinate space, a target with dielectric permittivity ϵ_{r2} and negligible conductivity σ_2 is buried in a lossy homogeneous dielectric halfspace (soil) of known relative permittivity ϵ_{r1} and weak conductivity σ_1 , bounded by a moderately rough interface with profile $z = h(x)$. The target region in the halfspace $z < h(x)$ is denoted by \mathcal{D} . As stated previously, we are mainly interested in shallowly buried plastic mine-like targets, which are essentially lossless ($\sigma_2 \approx 0$) and may have dielectric properties very close to those of the background soil, i.e., $|\epsilon_{r2} - \epsilon_{r1}|/\epsilon_{r1} \ll 1$. In what follows, the wavenumbers in free space and soil will be denoted by $k_0 = \omega\sqrt{\epsilon_0\mu_0}$ and $k_1 = k_0\sqrt{\epsilon_{r1} + i\sigma_1/(\omega\epsilon_0)}$, respectively, with ω representing the angular frequency, and ϵ_0, μ_0 denoting the free-space permittivity and permeability, respectively.

The soil is illuminated by a y -directed (TM-polarized) pulsed field generated by a (projected) large truncated aperture field distribution of width d at $z = z_A$. As typical of many ultra-wideband (UWB) GPR systems, we assume that the incident field is sufficiently well-

collimated so that it can be approximated by a pulsed truncated tapered plane wave,

$$e^i(\mathbf{r}, t) \sim g(x_B)p(t - c^{-1}z_B), \quad (1)$$

with FD spectrum

$$E^i(\mathbf{r}, \omega) \sim g(x_B)P(\omega)\exp(ik_0z_B). \quad (2)$$

In (1) and (2), $\mathbf{r} \equiv (x, z)$, $p(t)$ is a short pulse of length $T \ll d/c$ with FD spectrum $P(\omega)$, $c = (\epsilon_0\mu_0)^{-1/2}$ is the free-space wavespeed, $g(x_B)$ is a spatial taper and (x_B, z_B) are beam centered coordinates,

$$\begin{bmatrix} x_B \\ z_B \end{bmatrix} = \begin{bmatrix} \cos \theta_A & \sin \theta_A \\ \sin \theta_A & -\cos \theta_A \end{bmatrix} \begin{bmatrix} x - x_A \\ z - z_A \end{bmatrix}, \quad (3)$$

with θ_A and x_A denoting the tilt angle of the radiated beam relative to the z -axis and its spatial displacement, respectively. Parameters are chosen so that the illumination tapers to zero for $|x| \leq d/2$ and vanishes for $|x| > d/2$ (Fig. 2). Here and henceforth, capital letters identify FD quantities, whereas lower case letters are used for TD quantities; FD and TD quantities are related through the following Fourier transform pair

$$f(\mathbf{r}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\mathbf{r}, \omega) \exp(-i\omega t) d\omega, \quad F(\mathbf{r}, \omega) = \int_{-\infty}^{\infty} f(\mathbf{r}, t) \exp(i\omega t) dt. \quad (4)$$

Apart from the presence of the target, the geometry and notation are the same as in [11] and [13], to which we shall refer frequently throughout the paper. Notationally, we shall use $(\alpha.\beta)$ to denote eq. (β) in ref. $[\alpha]$; for instance, (3.8) means eq. (8) in ref. [3]. As in [11] and [13], the soil is assumed to be slightly lossy, i.e., $\sigma_1 \ll \epsilon_{r1}/(\Omega_0\epsilon_0)$, with Ω_0 denoting the center angular frequency of the pulse $P(\omega)$.

The actual goal of this investigation is the *imaging* of an underground test domain (e.g., $\mathcal{D}^{(test)}$ in Fig. 2), i.e., the estimation, from sparse TD scattered field observations, of its dielectric properties in order to localize and classify possible anomalies. To this end, the y -directed scattered electric field is sampled at N_t time instants at N_r fixed receiver locations $x_1^r, \dots, x_{N_r}^r$ on the plane $z = z^r$ (Fig. 2) to obtain a set of observations. The *known term* in the problem is this set of $N_r \times N_t$ samples. In our numerical experiments

in Sec. VI, we shall use synthetic field observation data generated via a rigorous full-wave solution of the forward scattering problem (see Sec. VI-A).

III. ROUGH SURFACE PROFILE ESTIMATION

In principle, the problem of estimating the coarse-scale roughness profile of the air-ground interface can be addressed with various technologies (e.g., acoustics, laser, SAR, etc.). In [13], we proposed a simple strategy based on the processing of the *early-time* response of a UWB GPR system sampled at a limited number of receivers. Our approach is based on a compact low-dimensional spline parameterization of the roughness profile, which provides implicit regularization mitigating the inherent ill-posedness of the problem, and on the Physical-Optics Pulsed-Beam (PO-PB) forward scattering model in [11] to generate predictions of the reflected field at the receivers. The estimation problem is thus converted into a nonlinear optimization problem aimed at retrieving the unknown spline coefficients via minimization of a least-square error functional that involves the PO-PB forward scattering prediction and the available observed data. The observation time windows are chosen so as to gate out the *late-time* response (i.e., causal contributions from regions beyond a critical depth) due to the possible presence there of buried targets which may produce a bias in the surface estimation. The resulting optimization problem is generally *non-convex*, and therefore the possible presence of local minima renders its numerical implementation non-trivial. Optimization strategies and computational issues are discussed in [13]. The proposed approach was found to provide accurate and robust estimations (even with noisy data and imperfect knowledge of soil parameters) with mild computational effort (typical computing time ~ 55 secs. on a 700 MHz PC).

Having established the possibility of estimating the coarse shape of the rough air-ground interface via *early-time* response processing, we now turn to the problem of underground imaging in the presence of a *known* moderately rough interface. In our investigation, incoherent scattering contributions from finer-scale roughness are not included. These finer-scale contributions are effectively dealt with through the use of additive noise models.

IV. FORWARD SCATTERING MODEL

A. Formulation

Referring to Fig. 2, the TM-polarized *total* FD field observed in free space at $\mathbf{r} = (x, z)$ can be conveniently expressed as a sum of two components – the target scattering contribution E^s and the *background field* E^b (i.e., the field in the absence of the target),

$$E(\mathbf{r}, \omega) = E^b(\mathbf{r}, \omega) + E^s(\mathbf{r}, \omega), \quad (5)$$

with E^s given by [15]

$$E^s(\mathbf{r}, \omega) = \frac{\omega^2}{c^2} \iint_{\mathcal{D}} E(\mathbf{r}', \omega) G_b(\mathbf{r}, \mathbf{r}', \omega) O(\mathbf{r}', \omega) d\mathbf{r}'. \quad (6)$$

In (6), E is the *total* field in the target region, G_b represents the FD Green's function of the rough-interface dielectric halfspace, and

$$O(\mathbf{r}', \omega) = [\epsilon_r(\mathbf{r}') - \epsilon_{r1}] + i \frac{[\sigma(\mathbf{r}') - \sigma_1]}{\omega \epsilon_0} = \Delta\epsilon_r(\mathbf{r}') + i \frac{\Delta\sigma(\mathbf{r}')}{\omega \epsilon_0} \quad (7)$$

is usually referred to as the *object function*, with $\epsilon_r(\mathbf{r}')$ and $\sigma(\mathbf{r}')$ denoting the local relative dielectric permittivity and conductivity, respectively. Since the object function in (7) vanishes outside the target region \mathcal{D} , the integration in (6) is limited accordingly. The corresponding TD results, of interest in this paper, are obtained by Fourier inversion of (5) and (6),

$$e(\mathbf{r}, t) = e^b(\mathbf{r}, t) + e^s(\mathbf{r}, t), \quad (8)$$

where

$$e^b(\mathbf{r}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} E^b(\mathbf{r}, \omega) \exp(-i\omega t) d\omega, \quad (9)$$

$$e^s(\mathbf{r}, t) = \iint_{\mathcal{D}} \Delta\epsilon_r(\mathbf{r}') u(\mathbf{r}, \mathbf{r}', t) d\mathbf{r}' + \frac{1}{c\epsilon_0} \iint_{\mathcal{D}} \Delta\sigma(\mathbf{r}') v(\mathbf{r}, \mathbf{r}', t) d\mathbf{r}'. \quad (10)$$

The kernels u and v in (10) are given by

$$u(\mathbf{r}, \mathbf{r}', t) = \frac{1}{2\pi c^2} \int_{-\infty}^{\infty} \omega^2 E(\mathbf{r}', \omega) G_b(\mathbf{r}, \mathbf{r}', \omega) \exp(-i\omega t) d\omega, \quad (11)$$

$$v(\mathbf{r}, \mathbf{r}', t) = \frac{1}{2\pi c} \int_{-\infty}^{\infty} i\omega E(\mathbf{r}', \omega) G_b(\mathbf{r}, \mathbf{r}', \omega) \exp(-i\omega t) d\omega. \quad (12)$$

Thus, in the total backscattered field (8) observed at the receivers, e^s is the *useful* signal carrying the information needed for imaging the target, whereas e^b is the interface-generated clutter. Further distortion introduced in the useful signal by the twice-traversed rough air-ground interface is accounted for in the fields E and G_b in (11) and (12). Assuming that an estimation of the rough interface profile is somehow available, one can use a full-wave forward solver, or the more efficient PO-PB forward scattering model in [11], to generate a prediction of the clutter e^b in (8) for a given pulsed excitation, and therefore *isolate* the target contribution e^s . Even removing the background clutter e^b , the relation in (10) between the scattered field at the receivers and the dielectric and conductivity inhomogeneities $\Delta\epsilon_r$ and $\Delta\sigma$ is not easily invertible, due to the presence in (11) and (12) of the total field E which is itself dependent on $\Delta\epsilon_r$ and $\Delta\sigma$ and consequently renders the problem *nonlinear*. However, for the plastic anti-personnel land mines of interest here, with dielectric properties close to those of the background soil, one can invoke the weak scattering limit via use of the Born approximation [15], where the total field E inside the target is replaced by the transmitted field E^t in \mathcal{D} , in the absence of the target. Moreover, we shall neglect the conductivity contrast contribution in (10), thus arriving at a *linear* model relating the scattered field at the receivers to the dielectric inhomogeneity $\Delta\epsilon_r$,

$$e^s(\mathbf{r}, t) \approx \iint_{\mathcal{D}} \Delta\epsilon_r(\mathbf{r}') \bar{u}(\mathbf{r}, \mathbf{r}', t) d\mathbf{r}', \quad (13)$$

$$\bar{u}(\mathbf{r}, \mathbf{r}', t) = \frac{1}{2\pi c^2} \int_{-\infty}^{\infty} \omega^2 E^t(\mathbf{r}', \omega) G_b(\mathbf{r}, \mathbf{r}', \omega) \exp(-i\omega t) d\omega. \quad (14)$$

The robust inversion of this model will be discussed in Sec. V. The limitations of the Born approximation, which neglects multibounce interactions inside the target, have been thoroughly investigated and are well documented in the technical literature (see, e.g., [16]). More sophisticated and accurate nonlinear [17], [18], iterative [19] and distorted

[20] variants could be exploited in principle, but this is outside the scope of the present investigation. For moderately lossy soils, once a first estimate of $\Delta\epsilon_r$ (and hence of \mathcal{D}) is obtained by inverting the linear model in (13) (see Sec. V), one can iteratively refine this estimate by re-introducing into (13) the neglected (Born-approximated) conductivity contribution in (10) as a *known term*.

B. Narrow-Waisted Gaussian Beam Implementation

In order to implement, and eventually invert, the TD linear model in (13) for the pulsed excitation in (1) and (2), one needs to compute the TD kernel \bar{u} in (14) at several observation points for a large number of source positions and time instants, with typical total number $\sim 10^6$ (see Sec. VI-D). As a consequence, the availability of a fast forward solver is a key ingredient for the overall computational feasibility of the proposed approach. In this scenario, the computational burdens of typical full-wave solvers (e.g., moment methods or finite differences) tend to be prohibitive, and it is therefore suggestive to resort to approximate approaches. We found that the Gabor-based narrow-waisted Gaussian beam algorithms in [11] are remarkably well-suited for this problem, and that they provide a reasonable trade-off between accuracy and computational burden. The PO-PB synthesis of plane-wave-excited fields transmitted into a rough-surface homogeneous halfspace detailed in [11] can readily be extended to deal with the subsequent irradiation from an induced line-source in the dielectric halfspace, thereby yielding a *closed form* TD expression for the kernel in (14) similar to those in [11]. However, for reasons that will become clear later on (Sec. IV-B.3), it is computationally cheaper to synthesize the transmitted field E^t and the rough-surface halfspace Green's function G_b in the FD, and then compute the TD kernel in (14) via fast Fourier transform (FFT) algorithms [21].

B.1 Transmitted Field

The computation of the field transmitted into the halfspace $z < h(x)$ in the absence of the target, under the quasi-plane-wave pulsed illumination in (2), is a problem already addressed in [11]. We therefore merely cite the final results, referring the reader to [11] for theoretical and computational details. The FD PO “equivalent current”, which generates the transmitted field, is parameterized in terms of x -domain discretized m_1 -indexed Gabor

basis functions with narrow width $L_1 \lesssim |\lambda_1| \ll d$, centered on the Gabor lattice points $x_{m_1} = m_1 L_1$, with $\lambda_1 = 2\pi/k_1$ denoting the wavelength in soil. These initial conditions generate narrow-waisted quasi-ray GBs propagating along the local refraction directions (Fig. 3). One obtains the following FD synthesis (cf. (11.21) and (11.22))

$$E^t(\mathbf{r}, \omega) \sim \frac{1}{2} \sum_{|m_1| \leq (d/2L_1)} C_{m_1}^t(\omega) \tilde{\mathcal{B}}_{m_1}^t(\mathbf{r}, \omega), \quad (15)$$

where the Gabor coefficients $C_{m_1}^t$ can be estimated efficiently by sampling the PO “equivalent current” at the lattice points,

$$C_{m_1}^t(\omega) \approx 2(L_1/\sqrt{2})^{1/2} g(x_{Bm_1}) [1 + \mathcal{R}_{m_1}^{(1)}] P(\omega) \exp[ik_0(z_{Bm_1} - x_{m_1} \sin \theta_{m_1}^i \cos \alpha_{m_1})]. \quad (16)$$

In (16), (x_{Bm_1}, z_{Bm_1}) denote the lattice points $(x_{m_1}, h(x_{m_1}))$ in the incident-beam coordinates (3), $\alpha_{m_1} = \tan^{-1}[dh(x_{m_1})/dx]$ is the local surface slope, $\theta_{m_1}^i = \theta_A - \alpha_{m_1}$ is the local incidence angle (see Fig. 3), and $\mathcal{R}_{m_1}^{(1)}$ is the local Fresnel TM plane-wave reflection coefficient for incidence from free space,

$$\mathcal{R}_{m_1}^{(1)} = \frac{\cos \theta_{m_1}^i - (\epsilon_{r1} - \sin^2 \theta_{m_1}^i)^{1/2}}{\cos \theta_{m_1}^i + (\epsilon_{r1} - \sin^2 \theta_{m_1}^i)^{1/2}}. \quad (17)$$

The CSP GB propagator $\tilde{\mathcal{B}}_{m_1}^t$ in (15) is computed efficiently via complex-source-point (CSP) paraxial asymptotics (cf. (11.30)-(11.34)),

$$\begin{aligned} \tilde{\mathcal{B}}_{m_1}^t(\mathbf{r}, \omega) &\sim ik_1 2^{5/4} \left(\frac{L_1}{8\pi k_1} \right)^{1/2} \frac{(\zeta_{m_1} - ib_{m_1}^t \cos \theta_{m_1}^t)}{(\tilde{R}_{m_1}^t)^{3/2}} \\ &\times \exp \left\{ i \left[k_1 \left(\tilde{R}_{m_1}^t + x_{m_1} \sin \theta_{m_1}^t \cos \alpha_{m_1} + ib_{m_1}^t \right) + \pi/4 \right] \right\}, \quad (18) \end{aligned}$$

where $\theta_{m_1}^t$ is the local refraction angle (see Fig. 3), related to $\theta_{m_1}^i$ and ϵ_{r1} via Snell’s law $\sin \theta_{m_1}^i = \sqrt{\epsilon_{r1}} \sin \theta_{m_1}^t$, $\zeta_{m_1} = -(x - x_{m_1}) \sin \alpha_{m_1} + [z - h(x_{m_1})] \cos \alpha_{m_1}$, and $\tilde{R}_{m_1}^t$ denotes the *complex distance*

$$\tilde{R}_{m_1}^t = \sqrt{(x_{bm_1}^t)^2 + (z_{bm_1}^t - ib_{m_1}^t)^2}, \quad \text{Re}(\tilde{R}_{m_1}^t) \geq 0, \quad (19)$$

with

$$\begin{bmatrix} x_{bm_1}^t \\ z_{bm_1}^t \end{bmatrix} = \begin{bmatrix} \cos \gamma_{m_1}^t & \sin \gamma_{m_1}^t \\ \sin \gamma_{m_1}^t & -\cos \gamma_{m_1}^t \end{bmatrix} \begin{bmatrix} x - x_{m_1} \\ z - h(x_{m_1}) \end{bmatrix}, \quad \gamma_{m_1}^t = \theta_{m_1}^t + \alpha_{m_1}, \quad (20)$$

$$b_{m_1}^t = \sqrt{\epsilon_{r1}} (L_1 \cos \alpha_{m_1} \cos \theta_{m_1}^t)^2 / \lambda_0. \quad (21)$$

The tilde \sim denotes quantities extended into complex coordinate space via the CSP method. In our numerical implementation, we found reasonably accurate syntheses with beam spacing $L_1 \sim 0.5|\lambda_1|$, where $\lambda_1 = 2\pi/k_1$ denotes the wavelength in soil. The resulting computational burden (see Sec. VI-E) was found to be about 2 times lower than the quasi-real ray-tracing scheme in [10], and about 4 times lower than standard Kirchhoff-PO implementations.

B.2 Rough-Surface Halfspace Green's Function

The next step is the Gabor-based narrow-waisted GB synthesis of the rough-surface halfspace Green's function G_b , which, in our TD-Born model in (13) and (14), represents the scattering contribution from a (infinitesimal) pixel in the target region of Fig. 2. Thus, instead of a quasi-plane-wave pulsed illumination from free space as in Sec. IV-B.1, we now have cylindrical waves impinging from underground induced source points \mathbf{r}' ,

$$G_1(\mathbf{r}, \mathbf{r}', \omega) = -\frac{i}{4} H_0^{(1)}(k_1 |\mathbf{r} - \mathbf{r}'|), \quad \mathbf{r}' \in \mathcal{D}, \quad (22)$$

with $H_0^{(1)}(\cdot)$ denoting the zeroth order Hankel function of the first kind. As in [11, Sec. III-A], the field radiated into the halfspace $z > h(x)$ is generated by an induced PO “surface current” (cf. (11.16))

$$J_{PO}^g(x, \mathbf{r}', \omega) = 2G_1(\mathbf{r}_h, \mathbf{r}', \omega) [1 + \mathcal{R}^{(2)}(x)], \quad \mathbf{r}_h \equiv (x, h(x)), \quad (23)$$

where $\mathcal{R}^{(2)}(x)$ denotes the local Fresnel TM plane-wave reflection coefficient for incidence from soil,

$$\mathcal{R}^{(2)}(x) = \frac{\sqrt{\epsilon_{r1}} \cos \alpha(x) - [1 - \epsilon_{r1} \sin^2 \alpha(x)]^{1/2}}{\sqrt{\epsilon_{r1}} \cos \alpha(x) + [1 - \epsilon_{r1} \sin^2 \alpha(x)]^{1/2}}, \quad (24)$$

with $\alpha(x) = \tan^{-1}[dh(x)/dx]$ being the local slope (Fig. 4). As in [11, Sec. III-B], it is expedient to rewrite the PO current by separating out the locally linear phase term that an incident plane wave propagating along the positive z -direction would induce on the locally tangent plane (cf. (11.17))

$$J_{PO}^g(x, \mathbf{r}', \omega) = \mathcal{J}_{PO}^g(x, \mathbf{r}', \omega) \exp[-ik_1 x \sin \alpha(x) \cos \alpha(x)]. \quad (25)$$

By Gabor-expanding the weakly phased *reduced* PO current (cf. (11.18)-(11.22))

$$\mathcal{J}_{PO}^g(x, \mathbf{r}', \omega) = J_{PO}^g(x, \mathbf{r}', \omega) \exp[ik_1 x \sin \alpha(x) \cos \alpha(x)], \quad (26)$$

the resulting transmitted field is parameterized in terms of narrow-waisted GBs launched from lattice points $(x_{m_2}, h(x_{m_2}))$, $x_{m_2} = m_2 L_2$, into the halfspace $z > h(x)$, analogous to what is done in (15). We obtain (see Fig. 4)

$$G_b(\mathbf{r}, \mathbf{r}', \omega) \sim \frac{1}{2} \sum_{|m_2| \leq (d/2L_2)} C_{m_2}^g(\mathbf{r}', \omega) \tilde{\mathcal{B}}_{m_2}^g(\mathbf{r}, \omega), \quad (27)$$

$$C_{m_2}^g(\mathbf{r}', \omega) \approx (L_2/\sqrt{2})^{1/2} J_{PO}^g(x_{m_2}, \mathbf{r}', \omega) \exp(ik_1 x_{m_2} \sin \alpha_{m_2} \cos \alpha_{m_2}), \quad (28)$$

$$\begin{aligned} \tilde{\mathcal{B}}_{m_2}^g(\mathbf{r}, \omega) &\sim ik_0 2^{5/4} \left(\frac{L_2}{8\pi k_0} \right)^{1/2} \frac{(\zeta_{m_2} - ib_{m_2}^g \cos \theta_{m_2}^g)}{(\tilde{R}_{m_2}^g)^{3/2}} \\ &\times \exp \left\{ i \left[k_0 \left(\tilde{R}_{m_2}^g + x_{m_2} \sin \theta_{m_2}^g \cos \alpha_{m_2} + ib_{m_2}^g \right) + \pi/4 \right] \right\}, \end{aligned} \quad (29)$$

$$\theta_{m_2}^g = \sin^{-1}(\sqrt{\epsilon_{r1}} \sin \alpha_{m_2}), \quad \alpha_{m_2} = \alpha(x_{m_2}), \quad (30)$$

$$\zeta_{m_2} = -(x - x_{m_2}) \sin \alpha_{m_2} + [z - h(x_{m_2})] \cos \alpha_{m_2}, \quad (31)$$

$$\tilde{R}_{m_2}^g = \sqrt{(x_{bm_2}^g)^2 + (z_{bm_2}^g - ib_{m_2}^g)^2}, \quad \text{Re}(\tilde{R}_{m_2}^g) \geq 0, \quad (32)$$

$$\begin{bmatrix} x_{bm_2}^g \\ z_{bm_2}^g \end{bmatrix} = \begin{bmatrix} \cos \gamma_{m_2}^g & -\sin \gamma_{m_2}^g \\ \sin \gamma_{m_2}^g & \cos \gamma_{m_2}^g \end{bmatrix} \begin{bmatrix} x - x_{m_2} \\ z - h(x_{m_2}) \end{bmatrix}, \quad \gamma_{m_2}^g = \theta_{m_2}^g - \alpha_{m_2}, \quad (33)$$

$$b_{m_2}^g = (L_2 \cos \alpha_{m_2} \cos \theta_{m_2}^g)^2 / \lambda_0. \quad (34)$$

Note that possible evanescent contributions arising from total reflection are not accounted for in this simple model; in case $\sin \theta_{m_2}^g > 1$, the corresponding GB is simply ignored. Also note that, as in (15), the discretization in (27) is limited to the region $|x| \leq d/2$, although now the line-source illumination does not necessarily vanish outside. However, in all numerical simulations presented in Sec. VI, it was verified that possible truncation-induced (late-time) effects were essentially negligible in the useful observation time-window. Concerning computational burden, the same considerations as in Sec. IV-B.1 apply.

B.3 Target-Scattered Field

Using the narrow-waisted GB syntheses in Secs. IV-B.1 and IV-B.2, the FD spectra of the transmitted field E^t and the rough-surface halfspace Green's function G_b are evaluated over the bandwidth of interest, and the TD Born kernel in (14) is computed via FFT [21] (see Sec. VI-B for details). Note that, in order to accommodate the evanescent spectra in the CSP GB propagators (18) and (29), the analytic signal formulation (one-side Fourier transform) should be used

$$\bar{u}(\mathbf{r}, \mathbf{r}', t) = \frac{1}{2\pi c^2} \text{Re} \left[\int_0^\infty \omega^2 E^t(\mathbf{r}', \omega) G_b(\mathbf{r}^r, \mathbf{r}', \omega) \exp(-i\omega t) d\omega \right], \quad \text{Im}(t) \leq 0. \quad (35)$$

Denoting by $N_{b1} = d/L_1$ and $N_{b2} = d/L_2$ the number of beams used in the GB syntheses (15) and (27), respectively, the computational complexity for evaluating the kernel waveform in (35), for fixed frequency and fixed observation and source locations, is thus $\mathcal{O}(N_{b1} + N_{b2})$. As mentioned earlier, a *direct* TD closed-form synthesis could also be obtained in terms of rapidly computable confluent hypergeometric functions, paralleling the procedure in [11, Sec. IV]. However, in order to keep track of the frequency dependence, this would require accounting for the element-by-element interaction between the beams in (15) and those in (27), resulting in a *double* beam summation for each space-time sample with a consequent computational complexity $\mathcal{O}(N_{b1} \cdot N_{b2})$, which would be considerably more expensive.

V. ROUGH SURFACE UNDERGROUND IMAGING

Based on the linear forward scattering model in (13) and on a number of observations of the scattered field waveforms at the receivers, an inversion scheme can be constructed to retrieve the unknown dielectric contrast $\Delta\epsilon_r$ in the test domain $\mathcal{D}^{(test)}$ to be imaged (Fig. 2). We recall that this problem is inherently ill-posed [22] and that only limited-viewing observations and approximate forward modelings are available, in addition to the unavoidable noise and measurement uncertainty of any practically conceivable implementation. As a consequence, classical linear approaches, such as diffraction tomography (see, e.g., [23], [24]), are difficult to apply, and regularization techniques are needed to restore well-posedness and achieve robust reconstructions. This type of problem is analogous to those typically arising in image processing applications such as object boundary detection and image segmentation [25], so that a whole arsenal of well-established tools is available. In this connection, we have been exploring two popular regularization approaches with *edge-preserving* capabilities, based on different parameterizations of the region under investigation.

A. Pixel-Based Reconstruction: L^p Norm Regularization

One of the simplest reconstruction approaches is based on the parameterization of the test domain to be imaged into a number of adequately small pixels, wherein the object function $\Delta\epsilon_r$ is assumed to be uniform. By discretizing the test domain $\mathcal{D}^{(test)}$ into N pixels centered at \mathbf{r}'_k , $k = 1, \dots, N$, with area Δs , the linear model in (13) can be discretized accordingly as

$$e^s(\mathbf{r}, t) \approx \sum_{k=1}^N \Delta\epsilon_r(\mathbf{r}'_k) \bar{u}(\mathbf{r}, \mathbf{r}'_k, t) \Delta s. \quad (36)$$

Assuming that the scattered field e^s is sampled at N_r receiver locations \mathbf{r}_i^r , $i = 1, \dots, N_r$, and N_t time instants t_{ij} , $j = 1, \dots, N_t$ (Fig. 2), the problem can be cast into matrix form as

$$\underline{y} = \underline{\underline{A}} \cdot \underline{x} + \underline{n}, \quad (37)$$

where the *known term* \underline{y} is a column vector containing the $N_r \cdot N_t$ space-time samples of e^s , $\underline{\underline{A}}$ is a $(N_r \cdot N_t) \times N$ matrix containing the space-time discretization of the TD kernel \bar{u} in (14), \underline{x} is a N -element column vector containing the *unknown* dielectric contrast $\Delta\epsilon_r$ at each pixel, and the *noise* vector \underline{n} accounts for measurement uncertainty and unmodeled effects. Specifically,

$$\underline{y} = [\underline{y}_1^T \quad \cdots \quad \underline{y}_{N_r}^T]^T, \quad \underline{y}_i^T = [e^s(\mathbf{r}_i^r, t_{i1}), \dots, e^s(\mathbf{r}_i^r, t_{iN_t})], \quad (38)$$

$$\underline{\underline{A}} = \begin{bmatrix} \underline{A}_{11} & \cdots & \underline{A}_{1N} \\ \vdots & \ddots & \vdots \\ \underline{A}_{N_r 1} & \cdots & \underline{A}_{N_r N} \end{bmatrix}, \quad \underline{A}_{ik}^T = \Delta s [\bar{u}(\mathbf{r}_i^r, \mathbf{r}_k', t_{i1}), \dots, \bar{u}(\mathbf{r}_i^r, \mathbf{r}_k', t_{iN_t})], \quad (39)$$

$$\underline{x} = [\Delta\epsilon(\mathbf{r}_1'), \dots, \Delta\epsilon(\mathbf{r}_N')]^T, \quad (40)$$

with $k = 1, \dots, N$, $i = 1, \dots, N_r$, and with T denoting the transpose. Typical regularization approaches convert this inverse scattering problem into an optimization problem where a cost functional, containing both data and model information from (37) and possible prior information about the object function, is minimized. Among them, total variation (TV) [26] is well-known for its *edge-preserving* capabilities and has found several applications in EM inverse scattering scenarios (see, e.g., [6], [27], [28]). Here, we use a more general approach, based on L^p norm regularization (LNR). The inversion of the linear model in (37) is posed as the minimization of the cost functional [29]

$$J_{LNR}(\underline{x}) = \|\underline{y} - \underline{\underline{A}} \cdot \underline{x}\|_2^2 + \beta_1 \|\underline{\underline{D}}_\nu \cdot \underline{x}\|_p^p + \beta_2 \|K(\underline{x})\|_2^2, \quad (41)$$

where $\|\cdot\|_2$ is the standard Euclidean norm and $\|\cdot\|_p$ ($0 < p \leq 2$) is an L^p norm [29]. In (41), the first term penalizes lack of data fidelity, while the other terms introduce some loose *a priori* knowledge about the object geometry, with β_1 and β_2 denoting regularization parameters. In particular, the second term highlights the expected piecewise smoothness in the reconstructed object function by penalizing the L^p norm of a spatial gradient operator $\underline{\underline{D}}_\nu$. In our implementation we chose [6], [30]

$$\|\underline{\underline{D}}_\nu \cdot \underline{x}\|_p^p \equiv \sum_{k=1}^N \left\{ \nu [\underline{\underline{D}}_x \cdot \underline{x}]_k^2 + (2 - \nu) [\underline{\underline{D}}_z \cdot \underline{x}]_k^2 \right\}^{p/2}, \quad (42)$$

where $\underline{\underline{D_x}}$ and $\underline{\underline{D_z}}$ denote standard first-order finite-difference operators in the horizontal and vertical direction, respectively, and the parameter ν controls the orientation preference of the smoothness regularization ($\nu = 1$ yields no orientation preference) [30]. As will be shown, a regularization with orientation preference may give better results, in view of the limited-viewing configuration of the problem, which gives rise to different resolutions achievable in the horizontal and vertical directions. The formulation in (42) reduces to Tikhonov-like regularization [31] for $p = 2$, and to TV-like regularization [26] for $p = 1$. As compared with the L^2 norm in the standard Tikhonov regularization [31], L^p norms with $0 < p \leq 1$ on the spatial gradient term in (42) penalize *large* jumps less, thus allowing sharper edges to form in the reconstructed object function and yielding visually better (less blurred) reconstructions.

The last term in (41) is related to possible prior information about the object function sign. Assuming, for instance, a *negative* value for $\Delta\epsilon_r$ in the target region, the L^2 norm of the operator K penalizes *positive* values (or viceversa) of the reconstructed object function

$$[K(\underline{x})]_k = \begin{cases} x_k, & x_k > 0, \\ 0, & x_k \leq 0. \end{cases} \quad (43)$$

In our implementation, the cost functional in (41) is minimized via an iterative procedure based on half-quadratic approximations (see [26] for details). The proper choice of the regularization parameters β_1 and β_2 in (41) is an important issue, and several strategies have been proposed (see, e.g., [32]), but this is not the focus here. Our choice of the regularization parameters was done by trial and error.

B. Object-Based Reconstruction: Curve Evolution

Object-based techniques have been widely investigated in image processing, with many important applications in problems like object boundary detection and image segmentation [25]. The basic idea underlying these approaches is to exploit parametric deformable shape models for the object function which incorporate possible *a priori* information about the target geometry. Prominent among them are *curve evolution* (CE) techniques (see, e.g., [33], [34]), where a *gradient flow* is designed which attracts initial closed curves to the target boundary. Applications of such approaches to EM inverse scattering problems have

been proposed in [35]–[37]. In particular, in [37], a CE approach is applied to UWB GPR underground imaging of shallowly buried low-contrast plastic targets in the presence of a *flat* air-ground interface. This approach is applied here to the more challenging problem of a *rough* air-ground interface. The algorithm in [37] is briefly reviewed. Assuming a scenario with a *single* homogeneous target occupying the region \mathcal{D} (Fig. 2) bounded by a continuous curve \vec{C} , the unknown permittivity contrast function can be expressed as

$$\Delta\epsilon_r(\mathbf{r}) = \overline{\Delta\epsilon_r}\Pi_{\mathcal{D}}(\mathbf{r}), \quad (44)$$

where $\overline{\Delta\epsilon_r} = \epsilon_{r2} - \epsilon_{r1}$, and $\Pi_{\mathcal{D}}$ is the characteristic function of the target region \mathcal{D} ,

$$\Pi_{\mathcal{D}}(\mathbf{r}) = \begin{cases} 1, & \mathbf{r} \in \mathcal{D}, \\ 0, & \mathbf{r} \notin \mathcal{D}. \end{cases} \quad (45)$$

The more general case of multiple connected or unconnected components is addressed in [37]. The parameterization in (44) contains *implicit* regularization, since the reconstructed object must be homogeneous. In this connection, the reconstruction task consists in estimating the target contour \vec{C} and the permittivity contrast $\overline{\Delta\epsilon_r}$. Using (44), the linear model in (13) can be rewritten as

$$e^s(\mathbf{r}, t) \approx \overline{\Delta\epsilon_r} \iint_{\mathcal{D}} \bar{u}(\mathbf{r}, \mathbf{r}', t) d\mathbf{r}'. \quad (46)$$

We assume, as in Sec. V-A, that a set of $N_r \cdot N_t$ samples of the scattered field is collected at N_r receiver locations \mathbf{r}_i^r , $i = 1, \dots, N_r$, at N_t time instants t_{ij} , $j = 1, \dots, N_t$ (Fig. 2). The problem of estimating target boundary and dielectric contrast is posed again as an optimization problem involving minimization of the following quadratic functional [37]

$$J_{CE}(\vec{C}, \overline{\Delta\epsilon_r}) = \frac{1}{2} \sum_{i=1}^{N_r} \sum_{j=1}^{N_t} (e_{ij}^s - \overline{\Delta\epsilon_r} U_{ij})^2 + \beta \int_{\vec{C}} dl. \quad (47)$$

As in (41), the cost functional in (47) contains a data fidelity term, where $e_{ij}^s = e^s(\mathbf{r}_i^r, t_{ij})$ and

$$U_{ij} = \iint_{\mathcal{D}} \bar{u}(\mathbf{r}_i^r, \mathbf{r}', t_{ij}) d\mathbf{r}'. \quad (48)$$

The second term serves as regularization by penalizing the arc-length of the estimated curve, with the choice of the regularization parameter β in (47) affecting its smoothness. Again, in our implementation, β is empirically selected by trial and error, taking into account prior expectations about target geometry (e.g., convexity).

To minimize the cost functional in (47), we use a CE approach to design a gradient flow that attracts an initial closed curve to the boundary of the target region \mathcal{D} . Given a family of smooth curves $\vec{C}(\tau)$ parameterized by τ , we search for the curve in this family and the contrast value $\overline{\Delta\epsilon_r}$ that minimize the cost functional in (47). For a given curve $\vec{C}(\tau)$, the contrast $\overline{\Delta\epsilon_r}$ must satisfy the following equation [37]

$$\frac{\partial J_{CE}(\tau, \overline{\Delta\epsilon_r})}{\partial \overline{\Delta\epsilon_r}} = - \sum_{i=1}^{N_r} \sum_{j=1}^{N_t} [e_{ij}^s - \overline{\Delta\epsilon_r} U_{ij}(\tau)] U_{ij}(\tau) = 0. \quad (49)$$

Moreover, it can be shown (see [37] for details) that the gradient direction of J_{CE} with respect to the curve $\vec{C}(\tau)$ is given by

$$\nabla_{\tau} J_{CE}(\tau, \overline{\Delta\epsilon_r}) = - \sum_{i=1}^{N_r} \sum_{j=1}^{N_t} [e_{ij}^s - \overline{\Delta\epsilon_r} U_{ij}(\tau)] \overline{\Delta\epsilon_r} \bar{u}_{ij}(\mathbf{r}'_c) \hat{\mathbf{n}}_c + \beta \kappa_c \hat{\mathbf{n}}_c, \quad (50)$$

where $\bar{u}_{ij}(\mathbf{r}'_c) = \bar{u}(\mathbf{r}'_i, \mathbf{r}'_c, t_{ij})$, \mathbf{r}'_c denotes points on the curve $\vec{C}(\tau)$, and $\hat{\mathbf{n}}_c$ and κ_c indicate the outward normal and the signed curvature of the curve at \mathbf{r}'_c , respectively [37]. To minimize the cost functional in (47), the curve $\vec{C}(\tau)$ is evolved along the direction of steepest descent, i.e., along the negative gradient of J_{CE} with respect to $\vec{C}(\tau)$,

$$\frac{d\vec{C}(\tau)}{d\tau} = -\nabla_{\tau} J_{CE}(\tau, \overline{\Delta\epsilon_r}) = \sum_{i=1}^{N_r} \sum_{j=1}^{N_t} [e_{ij}^s - \overline{\Delta\epsilon_r} U_{ij}(\tau)] \overline{\Delta\epsilon_r} \bar{u}_{ij}(\mathbf{r}'_c) \hat{\mathbf{n}}_c - \beta \kappa_c \hat{\mathbf{n}}_c, \quad (51)$$

where, for each τ , the optimal value of the contrast $\overline{\Delta\epsilon_r}$ is given by (49). For numerical implementation, the evolution in (51) is discretized in τ and stepped forward, alternatively updating the curve $\vec{C}(\tau)$ and the contrast $\overline{\Delta\epsilon_r}$ (via (49)). In our implementation, we use the level set method [38], [39] which yields a numerically efficient and stable evolution (see [37]).

As compared with pixel-based reconstruction techniques, object-based approaches like curve evolution (CE) offer several computational advantages. First is the natural incorporation of *a priori* information about target constitutive and geometric features, such

as homogeneity and shape smoothness. This allows use of efficient parametric models to describe target shape and permittivity contrast, with consequent *implicit* regularization which mitigates the ill-posedness of the problem and possibly reduces the data size required. Furthermore, it focuses on the key features of the target, which are estimated *directly* rather than extracted via postprocessing of a pixel-based image. Finally, the evolution in (51) depends on *local* properties, and the estimation of the contrast value in (49) is computationally inexpensive, resulting in an overall computational burden that tends to be lower than those of standard pixel-based approaches like LNR.

VI. NUMERICAL RESULTS

A. Reference Solution

The backscattered field observation data used as observations are simulated via a full-wave solution of the forward scattering problem. As in [11], this reference solution is based on the time-harmonic multifilament current method in [40], adapted to moderately rough interfaces. Specifically, the frequency domain (FD) spectrum of the field is evaluated at 100 different frequencies within the pulse bandwidth, without resorting to the plane wave approximation in (1) for the incident field, and with use of the full dispersive permittivity model for the soil. The sampled spectrum is first smoothed through local Padé-approximation [21] and then transformed via standard inverse fast Fourier transform (FFT) routines [21] to obtain the desired TD solution. In our numerical simulation we used a 1024-point FFT; the required TD samples were obtained by linear interpolation of the returned time series.

B. Gaussian Beam Algorithm Parameters

The Born TD kernel \bar{u} in (35), needed for computing the data fidelity term in both LNR and CE cost functionals, is first synthesized in the FD via narrow-waisted GB superposition (cf. (15) and (27)) and subsequently transformed to the TD via a 512-point FFT. In this case, the spectrum is sampled at 50 frequencies within the pulse bandwidth. All other needed frequency samples, as well as the output time waveform, are obtained via linear interpolation. The GB spacings L_1 and L_2 in (15) and (27), respectively, are chosen as half a (ambient) wavelength, i.e., $L_1 = |\lambda_1|/2$, $L_2 = \lambda_0/2$, with $\lambda_0 = 2\pi/k_0$ denoting

the free-space wavelength. This configuration was arrived at using a *pragmatic* stability criterion based on the (in)sensitivity of the outcome to further decreases of the GB spacing (i.e., increases in the number of GBs). To further improve computational efficiency, for a fixed observation point, one can rewrite the CSP GB propagators in (18) and (29) (see, e.g., (11.52)-(11.54)) to isolate frequency-independent parts that need to be computed only once.

C. Simulation Parameters

The numerical simulations that follow are based on the configuration and parameters described in Fig. 5. In this scenario (Fig. 5a), although no specific data model is used for soil, target and roughness, geometric and constitutive parameters were adjusted so as to be consistent with typical moderate roughness (maximum-to-minimum height $\sim 4\text{cm}$, maximum slope $\sim 32^\circ$), slightly lossy soil characteristics [41] ($\epsilon_{r1} = 4$, $\sigma_1 = 0.01\text{ S/m}$), and shallowly-buried plastic mine-like targets ($10\text{cm} \times 6\text{cm}$ ellipse with $\epsilon_{r2} = 3.5$ and $\sigma_2 = 0$, and with center at 10cm below nominal ground). The roughness profile was randomly generated via the quartic spline model in [13, Sec. IV]. The illumination field in (1) was chosen as a vertically incident fourth-order Rayleigh UWB pulse with $T = 1.3\text{ns}$ (Fig. 5c) and cosine tapering (Fig. 5b). This pulse length was found by trial and error to provide a good trade-off between the contrasting high-frequency requirements for good resolution (for both surface estimation and target localization) and low-frequency requirements for adequate soil penetration. Such UWB pulses typically work well for shallow burial depths and slightly lossy soils, as in our example, but may be not suitable for larger burial depths and/or soil losses.

The test domain to be imaged consists of a $20\text{cm} \times 20\text{cm}$ domain surrounding the target, with top side at 5cm below nominal ground. The backscattered field observation data needed for inversion are collected at $N_r = 11$ receivers at $z^r = 0.3\text{m}$ and $x_i^r = -0.5\text{m}, -0.4\text{m}, \dots, 0.5\text{m}$. Note that the relatively low permittivity contrast in this example ($|\Delta\epsilon_r|/\epsilon_{r1} = 0.125$), which justifies the Born approximation in (13), renders the test particularly challenging in view of the weak target scattering as compared with the interface-generated clutter. The burial depth was chosen so as to avoid the relatively clean case where ground-scattered and target-scattered signals are well time-resolved (see Sec.

VI-D).

D. Reconstruction Examples

Before testing the various underground imaging strategies, it is useful to look at the scattered waveforms at the receivers, in order to identify the various contributions and to check the accuracy of the Born-GB forward scattering model. Figure 6a shows the comparison between the total field e (soil+target) and the background field e^b (soil only) waveforms at a fixed receiver location, computed via the full-wave technique described in Sec. VI-A. The two waveforms are practically indistinguishable in the plot scale. Their difference (the target contribution e^s) is displayed in Fig. 6b and is compared with the Born-GB synthesis described in Sec. VI-B. Reasonably good agreement is observed. Looking at the different scales in the two plots, it is noted that background field removal is crucial for achieving reasonable signal-to-noise ratio in the inverse scattering procedure. For rough interfaces and shallowly-buried targets, as in this example, this cannot usually be accomplished by mere time-windowing of the data since ground-scattered and target-scattered signals are *not* well time-resolved. In this connection, the availability of rough interface profile estimations (and hence background field predictions) is of great importance.

Another important issue is the choice of the actual number of unknowns in the object parameterization. This requires trade-off between resolution, robustness and computational burden. Aside from computational considerations, it is understood that beyond a critical level, related to the *essential* dimension of the available information, any attempt at further refinement in the object discretization will increase the sensitivity to noise, *without* actually improving the resolution. The essential information entailed in the forward scattering model can be estimated roughly, for instance, via singular value decomposition (SVD) [21] of the (discretized) scattering kernel in (14), whereas the prior information introduced with the various regularization techniques cannot be easily quantified. We did not attempt to address *optimal* strategies for object discretization and data acquisition. The data and unknown configurations used in the examples, though not necessarily optimal, were found to provide a reasonable trade-off between the above contrasting requirements. In particular, to account for possible redundancy in the observed data, in all examples below we retain a data-to-unknown ratio ~ 3 .

For pixel-based reconstruction with L^p norm regularization (LNR), the test domain to be imaged was discretized into 30×30 square pixels (900 unknowns), and we used $N_t = 300$ time samples of the *late-time* response (causally related to the test domain) at each of the $N_r = 11$ receivers in Fig. 5. The observation vector \underline{y} in (38) was obtained by subtracting the PO-PB-computed background field e^b (see [11]) from the full-wave-computed total field e at each receiver; the (3300×900) data matrix \underline{A} in (39) was computed via the Born-GB algorithm (see Sec. VI-B). In all examples below, the various regularization parameters were selected pragmatically via trial and error, and the results reported correspond to the *best* reconstruction obtained. Reconstruction results, however, were found to be relatively insensitive to variation of these parameters within a calibrated range.

We now move on to presenting various examples of underground imaging results.

D.1 Perfectly Known Roughness Profile and Soil Parameters, and Noiseless Data

We start with the simplest, though somewhat unrealistic, case of perfectly known roughness profile and soil parameters, and noiseless observation data, on which we test the various reconstruction techniques presented so far. A number of representative pixel-based reconstruction examples are shown in Fig. 7. Specifically, the true object function (ground truth) is shown in Fig. 7a, while Figs. 7b-d display a number of LNR reconstructions with $p = 1$, i.e., total-variation-like (TVL), with various combinations of the regularization parameters, namely, *i*) gradient penalty without directional preference, i.e., $\beta_2 = 0$ in (41) and $\nu = 1$ in (42) (Fig. 7b); *ii*) gradient penalty with directional preference, i.e., $\beta_2 = 0$ and $\nu = 0.1$ (Fig. 7c); and *iii*) gradient penalty with directional preference and “positivity penalty”, i.e., $\beta_2 > 0$ and $\nu = 0.1$ (Fig. 7d). It is observed that all TVL images, though not yielding highly accurate point-wise reconstructions, provide reasonably accurate target localization. Note that the limited-viewing geometry (vertical illumination) renders the problem *more ill-posed* in the horizontal direction, resulting in higher vertical than horizontal resolution in all cases. In this connection, gradient penalty with *directional* preference turns out to be more effective, since the data fidelity term is less sensitive to horizontal than vertical blurring. If the gradient penalty is equally imposed on both directions (as in Fig. 7b), it is possible that the vertical direction is under-regularized while the horizontal direction is over-regularized. Therefore, a gradient penalty with $\nu < 1$ (smaller

in the horizontal direction, as in Fig. 7c) may help in compensating for this imbalance, and yield visually better reconstructions. Looking at Fig. 7d, it is realized that prior knowledge of the object function sign, and consequent positivity (or negativity) penalty in the cost functional, may further improve the reconstruction.

For more *quantitative* accuracy assessments, we introduced two simple r.m.s. error metrics in both target and background regions,

$$\Delta e_t \equiv \frac{\sum_{\mathbf{r}'_k \in \mathcal{D}} [\epsilon_{r2} - \epsilon_{r2}^{(est)}(\mathbf{r}'_k)]^2}{N_2 \epsilon_{r2}^2}, \quad \Delta e_b \equiv \frac{\sum_{\mathbf{r}'_k \notin \mathcal{D}} [\epsilon_{r1} - \epsilon_{r1}^{(est)}(\mathbf{r}'_k)]^2}{N_1 \epsilon_{r1}^2}, \quad (52)$$

where $\epsilon_{r1}^{(est)}$ and $\epsilon_{r2}^{(est)}$ denote the estimated relative permittivity in the background and target regions, respectively, while N_1 and N_2 indicate the corresponding number of pixels. These errors are specified (in dB) in the figure captions. It is observed from Fig. 7 that gradient penalty with directional preference coupled with positivity penalty (Fig. 7d) yields the smallest errors ($\Delta e_t = -22\text{dB}$, $\Delta e_b = -37\text{dB}$).

Reconstruction examples for $p = 0.5, 1.5$ and 2 (with directional preference and positivity penalty) are shown in Fig. 8. By comparing these reconstructions, together with that for the corresponding TVL (i.e., $p = 1$) in Fig. 7d, one notes progressive blurring toward the Tikhonov-like ($p = 2$, Fig. 8c) reconstruction. However, LNR with $p < 1$ (e.g., $p = 0.5$ in Fig. 8a) was found not to yield significant improvement as compared with TVL, and the corresponding results are omitted henceforth.

Figure 9 shows a reconstruction obtained via CE, which provides good estimation of both target boundary and permittivity (only 0.6% error in ϵ_{r2}). The reconstruction looks visually better than the corresponding TVL one in Fig. 7d, as also witnessed by the smaller r.m.s. errors, $\Delta e_t = -25\text{dB}$, $\Delta e_b = -39\text{dB}$, computed from (52) with the same pixelization used for LNR. Again, due to the limited-viewing configuration, localization is more accurate vertically than horizontally. For this reconstruction, a reduced data set was used, with only 6 receivers (#1, 3, 5, 7, 9, and 11, in Fig. 5) and 100 time samples each, taking advantage of the more compact object parameterization entailed in CE as compared with pixel-based reconstruction techniques.

D.2 Imperfectly Known Roughness Profile and Soil Parameters, and Noisy Data

We now explore more realistic configurations by removing the assumption of perfect knowledge of the air-soil rough interface as well as soil parameters and observation data. First, to highlight the effect of the air-ground interface roughness, so far assumed to be perfectly known, we show in Fig. 10 a TVL reconstruction obtained by *erroneously* assuming a flat air-ground interface at $z = 0$ in the forward scattering model. Image deterioration, mainly due to poor background field removal, is dramatic and renders the reconstruction practically meaningless. For this configuration, CE likewise gave meaningless results, exhibiting strong sensitivity to initial conditions and regularization parameters. Results can be improved via statistical processing, as shown in [6], but will not become comparable to those in Figs. 7–9.

Now, we use the algorithm in [13] (see also Sec. III) to generate estimations of the rough interface to be used in the subsequent underground imaging problem. We also investigate the effect of measurement uncertainty in the observation data as well as soil parameters. Figure 11 shows the rough interface reconstructions obtained with two different data configurations: *i*) corruption of the observation data with an additive uniform noise ($\pm 10\%$ in amplitude) that accounts for measurement uncertainty and unmodeled effects (e.g., finer-scale roughness scattering), and *ii*) assumption of a -5% error in ϵ_{r1} and σ_1 in the forward scattering model. For these reconstructions, as discussed in [13], we used $N_t = 50$ time samples of the *early-time* response at each of the 11 receivers in Fig. 5; the observation time-windows were chosen so as to roughly gate out scattering contributions from possible regions deeper than $\sim 8\text{cm}$ below nominal ground ($z = 0$), thus avoiding any possible bias due to target scattering. Interface estimates are reasonably accurate, apart from the weakly-illuminated edge regions, as also observed in [13]. We repeat that, for the underground imaging problem, these configurations are particularly challenging. The -5% relative permittivity mismatch in the second example is on the same order as the actual dielectric contrast $\Delta\epsilon_r = -0.5$ to be estimated. The corresponding TVL and CE target reconstructions are shown in Fig. 12, where a slightly narrower illumination ($d = 0.8\text{m}$) is used in order to de-emphasize the effect of the poorly estimated side regions in Fig. 11. Specifically, Figs. 12a and 12b show TLV and CE reconstructions, respec-

tively, for the case of noisy data (with perfectly known soil parameters), whereas Figs. 12c and 12d show the corresponding results for the case of imperfectly known soil parameters (with noiseless data). In these examples, the corresponding estimated air-soil interface profiles in Fig. 11 are used in the forward scattering model. As seen from Figs. 12a and 12b, reconstructions turn out to be robust with respect to noise in the observation data. In particular, CE still yields good target localization and permittivity estimation (1.7% error in ϵ_{r2} ; $\Delta e_t = -23\text{dB}$, $\Delta e_b = -39\text{dB}$). For the TVL reconstruction, r.m.s. errors ($\Delta e_t = -21\text{dB}$, $\Delta e_b = -31\text{dB}$) are still comparable to those obtained in the *ideal* case (Fig. 7d). For the more challenging soil-parameter-mismatch test in Figs. 12c and 12d, although the target shape is not as accurately captured as before, localization and permittivity estimation (1.7% error in ϵ_{r2} , and $\Delta e_t = -26\text{dB}$, $\Delta e_b = -25\text{dB}$ with CE; $\Delta e_t = -15\text{dB}$, $\Delta e_b = -22\text{dB}$ with TVL) are still acceptable for classification purposes. We stress that this latter test is particularly meaningful since the assumed mismatch in the soil parameters reduces the *effective* visibility of the (already low-contrast) target, thus also emphasizing the effect of imperfect background field removal. However, it is understood that for such low-contrast target scattering scenarios, *stronger* uncertainty in the soil parameters may render the target practically invisible.

E. Computational Aspects

The overall computational features of the proposed approach are detailed in Table I for each of the separate tasks required, with specific reference to a 700 MHz PC benchmark implementation. The computational burdens of the first four tasks are strongly tied to those of the narrow-waisted GB forward solvers in [11] and Secs. IV-B and VI-B. In this connection, the reported (average) computing times were found to be about four times and two times smaller than those achievable with standard PO implementations the quasi-real ray tracing in [10], respectively. Meaningful comparisons with full-wave forward solvers are not easy since computational features may be strongly implementation-dependent. To give a rough idea, the (estimated) computing times for estimating the air-soil interface and for building up the forward model (data matrix filling up) when using our full-wave reference solution in the forward scattering models would have been on the order of eight and four hours, respectively. These rough estimates are based on forward scattering benchmark

simulations.

As seen from Table I, the only memory-demanding tasks are those involving the manipulation of the LNR data matrix $\underline{\underline{A}}$ in (41) and a similar (but smaller) data matrix used in the CE (51) via the level set method.

Overall, CE tends to be computationally cheaper than LNR. Although the numerical codes are not fully optimized, overall computing times are on the order of a few minutes, thus leaving room for optimism that extensions of the algorithmic approach to more realistic three-dimensional (3D) scenarios will remain computationally affordable for real-world applications.

VII. CONCLUSIONS

An adaptive approach for underground imaging via ultrawide-band ground penetrating radar in the presence of a moderately rough air-soil interface has been presented. The proposed approach is based on short-pulse Gabor-based narrow-waisted Gaussian beam algorithms as fast forward scattering models, through which the coarse-scale deterministic features of the air-ground interface are first estimated and subsequently exploited to compensate for background clutter and distortion in the interrogating signal. Preliminary 2D results, restricted to slightly lossy soils and shallowly-buried low-contrast dielectric targets, indicate that quite accurate and robust target reconstructions can be obtained with reasonable computing times and resources, even with sparse and corrupted data and imperfect knowledge of soil parameters. The proposed approach looks promising from both the computational and accuracy viewpoints, with potential application to antipersonnel land mine clearance. In this connection, the necessary extension to more realistic 3D configurations, currently under investigation, is fairly straightforward theoretically, but its computational features remain to be explored. Also under investigation are more robust strategies for background field removal.

REFERENCES

- [1] A.C. Dubey, J.F. Harvey, J.T. Broach, E.R. Dugan, Eds., *Detection and Remediation Technologies for Mines and Minelike Targets V*, *Proc. SPIE*, vol. 4038, pp. 1485–1496, Aug. 2000.
- [2] T. Dogaru and L. Carin, “Time-domain sensing of targets buried under a rough air-ground interface,” *IEEE Trans. Antennas Propagat.*, vol. 46, No. 3, pp. 360–372, Mar. 1998.
- [3] R.A. Weisenseel, W.C. Karl, D.A. Castañón, E.L. Miller, C.M. Rappaport, and C.A. DiMarzio, “Statistical fusion of GPR and EMI data,” in *Detection and Remediation Technologies for Mines and Minelike Targets IV*, A.C. Dubey, J.F. Harvey, J. Broach, and R.E. Dugan, Eds., *Proc. SPIE*, vol. 3710, pp. 1179–1187, Aug. 1999.
- [4] T. Dogaru, L. Collins, and L. Carin, “Optimal time-domain detection of a deterministic target buried under a randomly rough interface,” *IEEE Trans. Antennas Propagat.*, vol. 49, No. 3, pp. 313–326, Mar. 2001.
- [5] H. Zhan, C.M. Rappaport, M. El-Shenawee, and E.L. Miller, “Mine detection under rough ground surfaces using 2-D FDTD modeling and hypothesis testing,” *Proc. 2001 IEEE Antennas Propagat. Int. Symposium*, Boston, MA, USA, July 8–13, 2001, vol. 3, p. 756.
- [6] H. Feng, D. A. Castañón, W. C. Karl, and E. L. Miller, “GPR imaging approaches for buried plastic landmine detection,” in *Detection and Remediation Technologies for Mines and Minelike Targets V*, A.C. Dubey, J.F. Harvey, J.T. Broach, E.R. Dugan, Eds., *Proc. SPIE*, vol. 4038, pp. 1485–1496, Aug. 2000.
- [7] J.J. Maciel and L.B. Felsen, “Systematic study of fields due to extended apertures by Gaussian beam discretization,” *IEEE Trans. Antennas Propagat.*, vol. 37, No. 7, pp. 884–892, July 1989.
- [8] J.J. Maciel and L.B. Felsen, “Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part I - plane layer,” *IEEE Trans. Antennas Propagat.*, vol. 38, No. 10, pp. 1607–1617, Oct. 1990.
- [9] J.J. Maciel and L.B. Felsen, “Gaussian beam analysis of propagation from an extended aperture distribution through dielectric layers, Part II - circular cylindrical layer,” *IEEE Trans. Antennas Propagat.*, vol. 38, No. 10, pp. 1618–1624, Oct. 1990.
- [10] V. Galdi, L.B. Felsen, and D.A. Castañón, “Quasi-ray Gaussian beam algorithm for time-harmonic two-dimensional scattering by moderately rough interfaces,” *IEEE Trans. Antennas Propagat.*, vol. 49, No. 9, pp. 1305–1314, Sept. 2001.
- [11] V. Galdi, L.B. Felsen, and D.A. Castañón, “Quasi-ray Gaussian beam algorithm for short-pulse two-dimensional scattering by moderately rough dielectric interfaces,” submitted to *IEEE Trans. Antennas Propagat.*, Apr. 2001.
- [12] V. Galdi, D.A. Castañón, and L.B. Felsen, “Multifrequency reconstruction of moderately rough interfaces via quasi-ray Gaussian beams,” to be published in *IEEE Trans. Geosci. Remote Sensing*, 2002.
- [13] V. Galdi, J. Pavlovich, W.C. Karl, D.A. Castañón, and L. B. Felsen, “Moderately rough dielectric interface reconstruction via short-pulse quasi-ray Gaussian beams,” submitted to *IEEE Trans. Antennas Propagat.*, Oct. 2001.
- [14] V. Galdi, H. Feng, D.A. Castañón, W.C. Karl, and L.B. Felsen, “Multifrequency subsurface sensing in the presence of a moderately rough air-soil interface via quasi-ray Gaussian beams,” invited paper in *Radio Science*, Special Issue on 2001 URSI EMT Symposium, submitted Nov. 2001.
- [15] W.C. Chew, *Waves and Fields in Inhomogeneous Media*. Oxford (UK): Oxford Press, Oxford, 1996.
- [16] J.B. Keller, “Accuracy and validity of the Born and Rytov approximations,” *J. Opt. Soc. Am.*, vol. 59, pp. 1003–1004, 1969.

- [17] T.M. Habashy, R.W. Groom, and B. Spies, "Beyond the Born and Rytov approximation: A nonlinear approach to electromagnetic scattering," *J. Geophys. Res.*, vol. 98, pp. 1759–1775, 1993.
- [18] N. Dasgupta, N. Geng, T. Dogaru, and L. Carin, "On the extended-Born technique for scattering from buried dielectric targets," *IEEE Trans. Antennas Propagat.*, vol. 47, pp. 1739–1743, 1999.
- [19] Y.M. Wang and W.C. Chew, "An iterative solution of the two-dimensional inverse scattering problem," *Int. J. Imaging Syst. Technol.*, vol. 1, pp. 100–108, 1989.
- [20] W.C. Chew and Y.M. Wang, "Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method," *IEEE Trans. Medical Imaging*, vol. 9, pp. 218–225, 1990.
- [21] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge (UK): Cambridge Univ. Press, 1992.
- [22] M. Bertero, "Linear inverse and ill-posed problems," in *Advances in Electronics and Electron Physics*, vol. 75, pp. 1–20. San Diego (CA): Academic Press, 1989.
- [23] A.J. Devaney, "A filtered backpropagation algorithm for diffraction tomography," *Ultrason. Imag.*, vol. 4, pp. 336–349, 1982.
- [24] T. Melamed, Y. Ehrlich, and E. Heyman, "Short-pulse inversion of inhomogeneous media: a time-domain diffraction tomography," *Inverse Problems*, vol. 12, pp. 977–993, 1996.
- [25] R.C. Gonzales and R.E. Woods, *Digital Image Processing*. New York (NY): Addison Wesley, 1993.
- [26] C.R. Vogel and M.E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Trans. Image Processing*, vol. 7, No. 6, pp. 813–824, June 1998.
- [27] D.C. Dobson and F. Santosa, "An image-enhanced technique for electrical impedance tomography," *Inverse Problems*, vol. 10, No. 2, pp. 317–334, 1994.
- [28] P. van den Berg and R.E. Kleinman, "A total variation enhanced modified gradient algorithm for profile reconstruction," *Inverse Problems*, vol. 13, No. 6, pp. 1607–1620, 1997.
- [29] W.C. Karl, "Regularization in image restoration and reconstruction", in *Handbook of Image and Video Processing*, A. Bovik, Ed., pp 141–160. New York (NY): Academic Press, 2000.
- [30] J. Kaufhold, R.C. Chan, W.C. Karl, and D. A. Castañón, "Ultrasound tissue analysis and characterization," in *Battlefield Biomedical Technologies*, H.H. Pien, Ed., *Proc. SPIE*, Vol. 3712, pp. 73–83, July 1999.
- [31] A.N. Tikhonov and V.Y. Arsenin, *Solution of Ill-Posed Problems*. Washington, D.C.: V.H. Winston, 1977.
- [32] P.C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Review*, vol. 34, No. 4, Dec. 1992.
- [33] A. Yezzi, S. Kichenassamy, K. Kumar, P. Olver, and A. Tennenbaum, "A geometric snake model for segmentation of medical imagery," *IEEE Trans. Medical Imaging*, vol. 16, No. 2, pp. 199–209, Apr. 1997.
- [34] J. Shah, "Riemannian drums, anisotropic curve evolution, and segmentation," *J. Visual Communication and Image Representation*, vol. 11, No. 2, pp. 142–153, Sept. 1999.
- [35] A. Litman and D. Lesselier, "Reconstruction of a two-dimensional binary obstacle by controlled evolution of a level set," *Inverse Problems*, vol. 14, pp. 685–706, 1998.
- [36] O. Dorn, E.L. Miller, and C.M. Rappaport, "A shape reconstruction method for electromagnetic tomography using adjoint fields and level sets," *Inverse Problems*, vol. 16, No. 5, pp. 1119–1156, Oct. 2000.
- [37] H. Feng, W.C. Karl, and D.A. Castañón, "A curve evolution approach to object-based tomographic reconstruction," submitted to *IEEE Trans. Image Processing*, June 2001.
- [38] S. Osher and J. Sethian, "Fronts propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations," *J. Comp. Phys.*, vol. 79, pp. 12–49, 1988.

- [39] F. Santosa, "A level-set approach for inverse problems involving obstacles," *ESAIM: Control, Optimisation, and Calculus of Variations*, Vol. 1, pp. 17–33, 1996.
- [40] Y. Leviatan and A. Boag, "Analysis of electromagnetic scattering from dielectric cylinders using a multifilament current model," *IEEE Trans. Antennas Propagat.*, vol. 35, No. 10, pp. 1119-1127, Oct. 1987.
- [41] J.E. Hipp, "Soil electromagnetic parameters as functions of frequency, soil density, and soil moisture," *Proc. IEEE*, vol. 62, pp. 98-103, Jan. 1974.

TABLE I

SUMMARY OF COMPUTATIONAL FEATURES. REPORTED COMPUTING TIMES ARE INTENDED AS AVERAGE VALUES FOR A BENCHMARK IMPLEMENTATION ON A 700 MHz PC.

Task	Implementation	Computing time [secs.]		Memory [Mbytes]	
		LNR	CE	LNR	CE
Interface estimation	C++	55		-	-
Background field removal	C++	1	0.4	-	-
Data matrix filling up	C++	120	62	-	-
Target imaging	Matlab	185	127	22.6	4.1

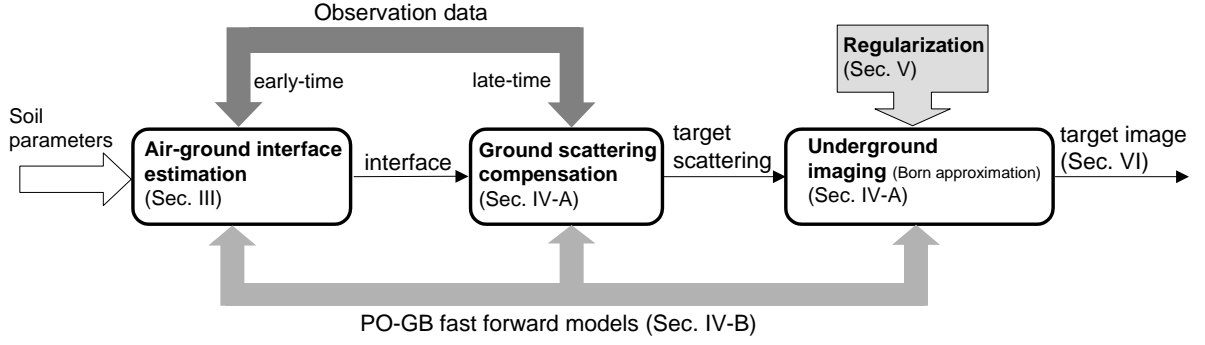


Fig. 1. Schematic flow-chart of the proposed adaptive approach, coordinated with the paper layout.

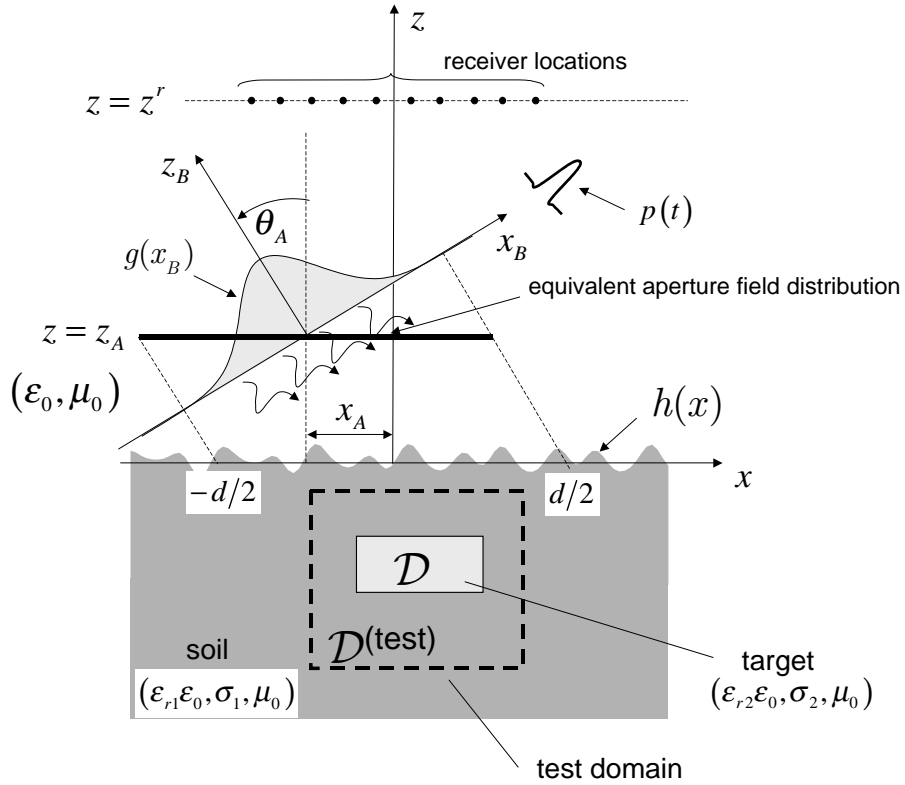


Fig. 2. Problem geometry. An aperture-generated, quasi-plane-wave, TM-polarized pulsed field impinges from free space onto a dielectric halfspace with known relative permittivity ϵ_{r1} and conductivity σ_1 , bounded by a moderately rough interface profile $z = h(x)$, wherein a target with dielectric permittivity ϵ_{r2} and conductivity $\sigma_2 \approx 0$ occupies the region \mathcal{D} . In order to image the test domain $\mathcal{D}^{(test)}$, the reflected field is sampled at N_t time instants at N_r fixed receiver locations $x_1^r, \dots, x_{N_r}^r$ on the observation plane $z = z^r$.

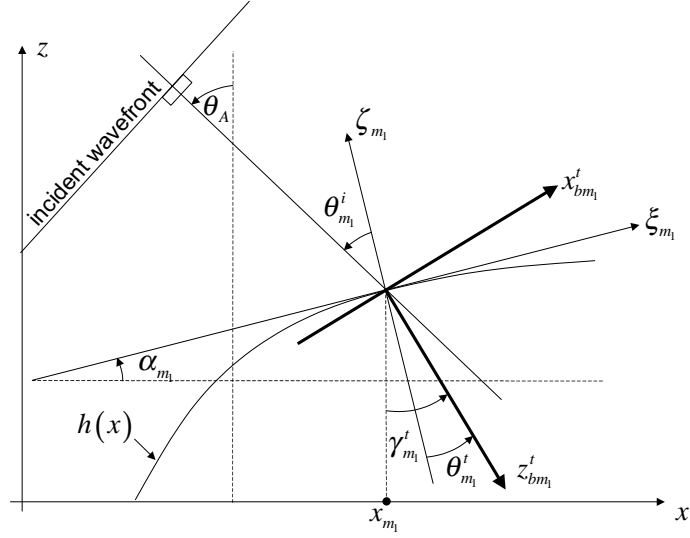


Fig. 3. Global and local coordinates for beams transmitted from free space into the soil. $(x_{bm_1}^t, z_{bm_1}^t)$ is the m_1 -th transmitted-beam-centered local coordinate system; (ξ_{m_1}, ζ_{m_1}) is the local tangent-normal coordinate system, centered at lattice points $(x_{m_1}, h(x_{m_1}))$; $\alpha_{m_1} = \tan^{-1}[dh(x_{m_1})/dx]$ is the local slope; $\theta_{m_1}^t$ is the local refraction angle measured from the surface normal; $\gamma_{m_1}^t = \theta_{m_1}^t + \alpha_{m_1}$ is the local refraction angle measured from the z -axis.

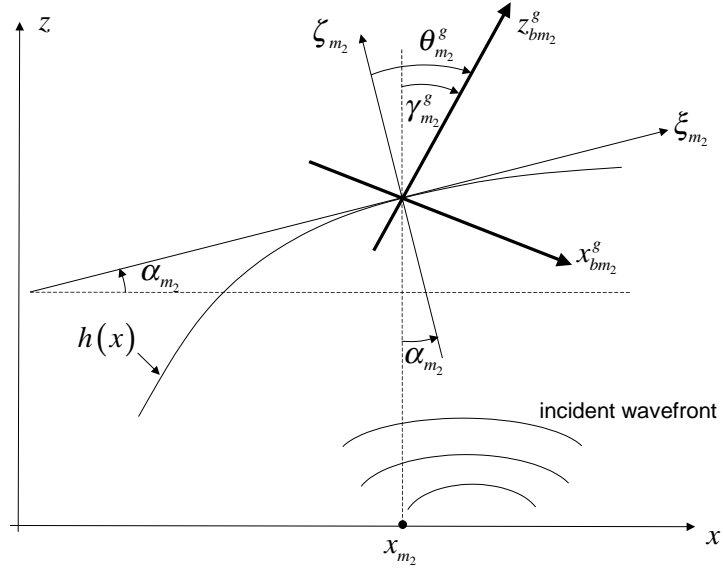


Fig. 4. As in Fig. 3, but for beams transmitted from soil into free space, with planar approximation of the incident cylindrical wavefront in (22). $(x_{bm_2}^g, z_{bm_2}^g)$ is the m_2 -th transmitted-beam-centered local coordinate system; $\theta_{m_2}^g$ is the local refraction angle measured from the surface normal; $\gamma_{m_2}^g = \theta_{m_2}^g - \alpha_{m_2}$ is the local refraction angle measured from the z -axis.

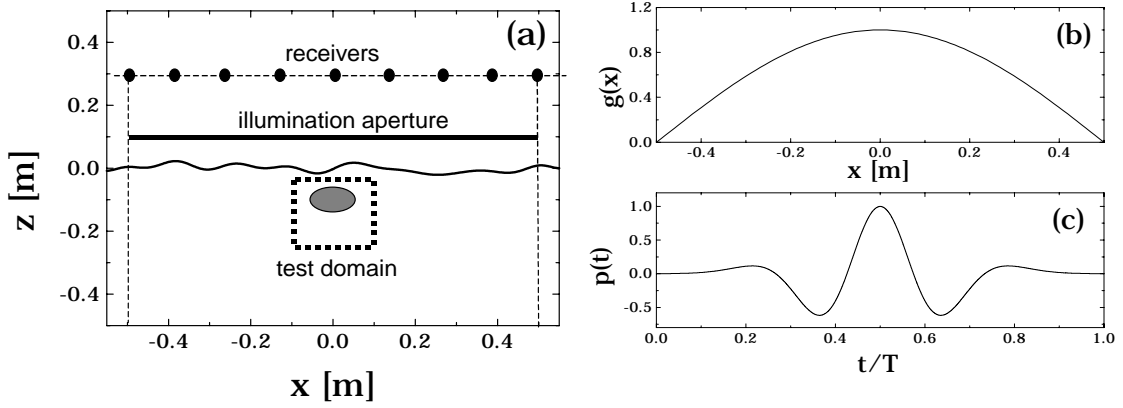


Fig. 5. Simulation geometry and parameters (cf. Fig. 2). (a): An elliptic ($10\text{cm} \times 6\text{cm}$) dielectric mine-like target with $\epsilon_{r2} = 3.5$ and $\sigma_2 = 0$ is buried with center at 10cm below the nominal ground ($z = 0$). The rough surface profile realization was randomly generated so as to simulate typical moderate roughness (maximum-to-minimum height $\sim 4\text{cm}$, maximum slope $\sim 32^\circ$) for a class of realistic soils ($\epsilon_{r1} = 4$, $\sigma_1 = 0.01\text{ S/m}$). A $20\text{cm} \times 20\text{cm}$ test domain surrounding the target (with top side at 5cm below nominal ground) is to be imaged. The reflected field is sampled at $N_r = 11$ receivers located at $z^r = 0.3\text{m}$ and $x_i^r = -0.5\text{m}, -0.4\text{m}, \dots, 0.5\text{m}$. (b): Incident field tapering $g(x) = \cos(\pi x/d)$. Parameters: $d = 1\text{m}$, $\theta_A = 0$, $x_A = 0$, $z_A = 0.1\text{m}$. (c): Fourth-order Rayleigh pulsed excitation $p(t)$ ($cT = 0.4d$, i.e., $T \sim 1.3\text{ns}$).

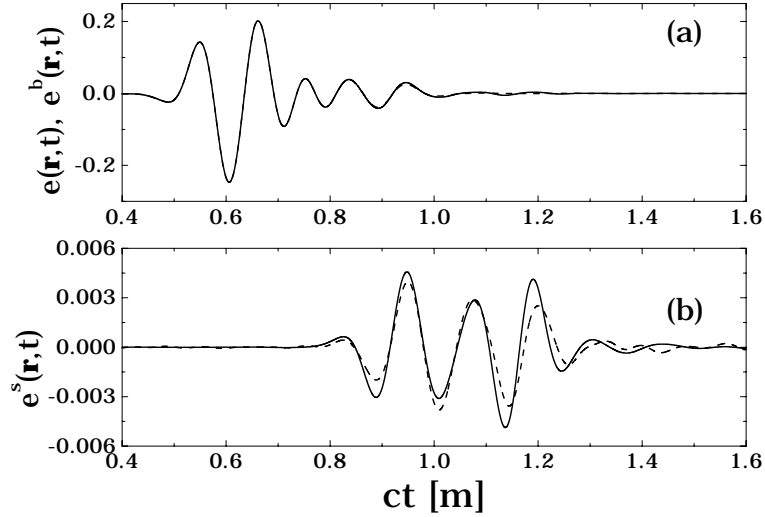


Fig. 6. Parameters as in Fig. 5. Field observed at receiver #8 ($x_8^r = 0.3\text{m}$, $z^r = 0.3\text{m}$) (a): — Total field e (reference solution); - - - Background field e^b (reference solution). Both are coincident on this drawing. (b): Target-scattered field $e^s = e - e_b$. — Reference solution; - - - Beam-Born approximation in (13) (parameters as in Sec. VI-B).

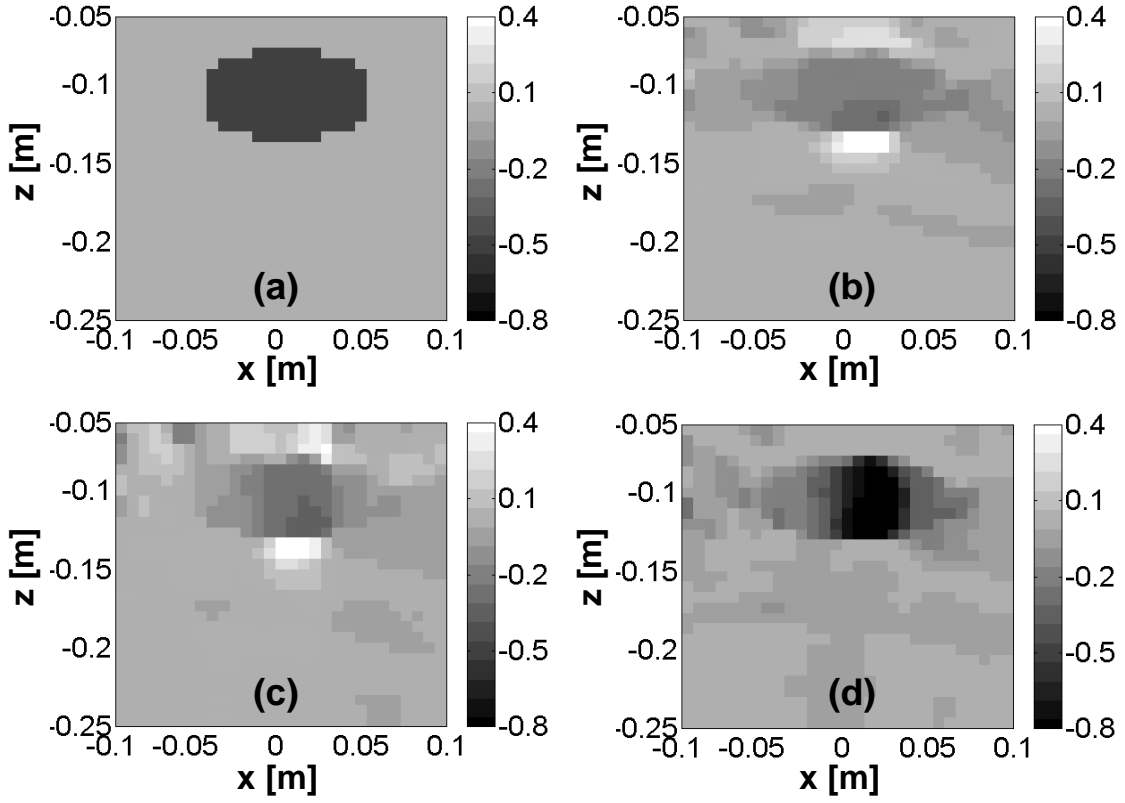


Fig. 7. Parameters as in Fig. 5. (a): Ground truth ($\Delta\epsilon_r$ reference configuration). (b): TVL ($p = 1$) reconstruction ($\beta_1 = 1.6 \cdot 10^{-6}$, $\nu = 1$, $\beta_2 = 0$; $\Delta e_t = -19\text{dB}$, $\Delta e_b = -35\text{dB}$). (c): TVL reconstruction with directional preference ($\beta_1 = 4.6 \cdot 10^{-6}$, $\nu = 0.1$, $\beta_2 = 0$; $\Delta e_t = -19\text{dB}$, $\Delta e_b = -35\text{dB}$). (d): TVL reconstruction with directional preference and positivity penalty ($\beta_1 = 3.4 \cdot 10^{-6}$, $\nu = 0.1$, $\beta_2 = 2 \cdot 10^{-3}$; $\Delta e_t = -22\text{dB}$, $\Delta e_b = -37\text{dB}$). The exact air-soil interface profile is used in the forward scattering model.

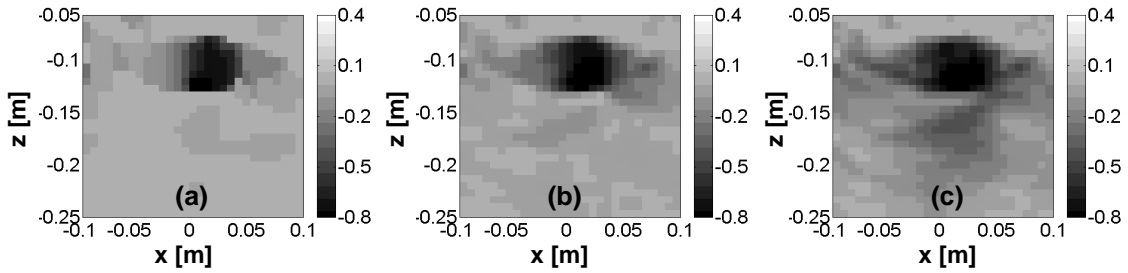


Fig. 8. As in Fig. 7, but using LNR with $p \neq 1$, directional preference and positivity penalty ($\beta_1 = 3.4 \cdot 10^{-6}$, $\nu = 0.1$, $\beta_2 = 2 \cdot 10^{-3}$). (a): $p = 0.5$ ($\Delta e_t = -22\text{dB}$, $\Delta e_b = -40\text{dB}$). (b): $p = 1.5$ ($\Delta e_t = -23\text{dB}$, $\Delta e_b = -33\text{dB}$). (c): $p = 2$ (Tikhonov, $\Delta e_t = -24\text{dB}$, $\Delta e_b = -27\text{dB}$). The case $p = 1$ (TVL) is displayed in Fig. 7d.

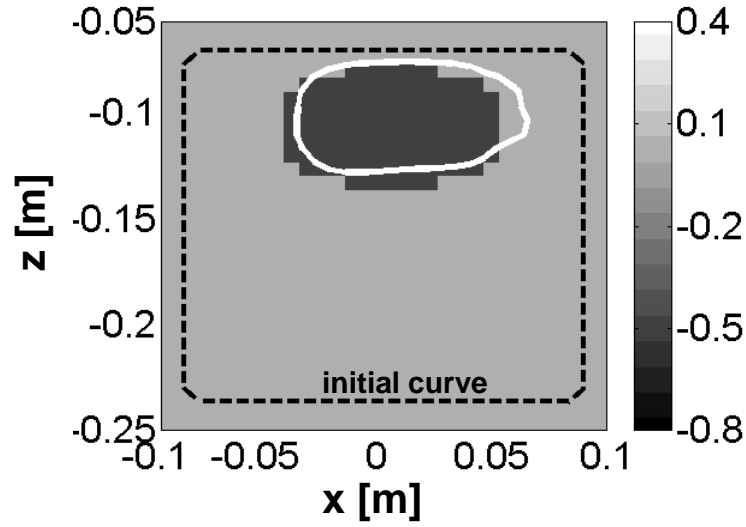


Fig. 9. Parameters as in Fig. 5. Curve evolution reconstruction (white curve) is superposed on ground truth. Curve initial conditions are displayed by dashed lines. Evolution parameters: 150 steps with $\beta = 3.5 \cdot 10^{-5}$ plus 620 steps with $\beta = 0.3 \cdot 10^{-5}$. The estimated target relative permittivity is $\epsilon_{r2} = 3.52$ (0.6% error). The r.m.s. errors in target and background regions are $\Delta e_t = -25\text{dB}$ and $\Delta e_b = -37\text{dB}$, respectively. The exact air-soil interface profile is used in the forward scattering model.

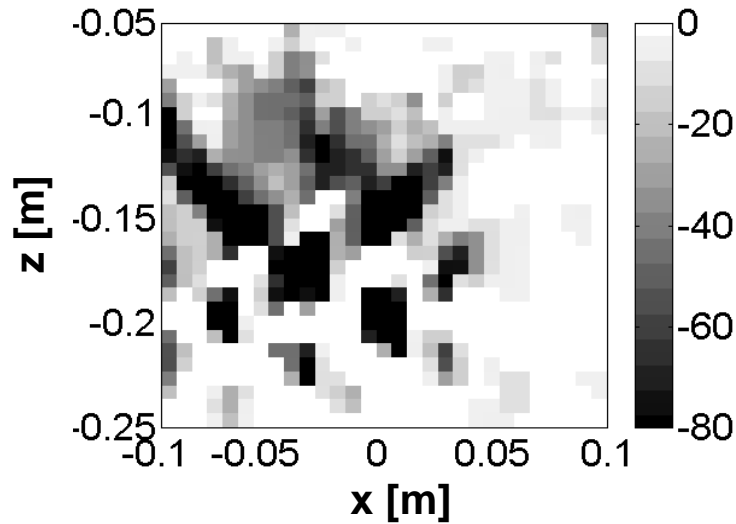


Fig. 10. As in Fig. 6d ($\beta_1 = 3.4 \cdot 10^{-6}$, $\nu = 0.1$, $\beta_2 = 2 \cdot 10^{-3}$), but assuming a flat interface at $z = 0$ in the forward scattering model ($\Delta e_t = 27\text{dB}$, $\Delta e_b = 28\text{dB}$).

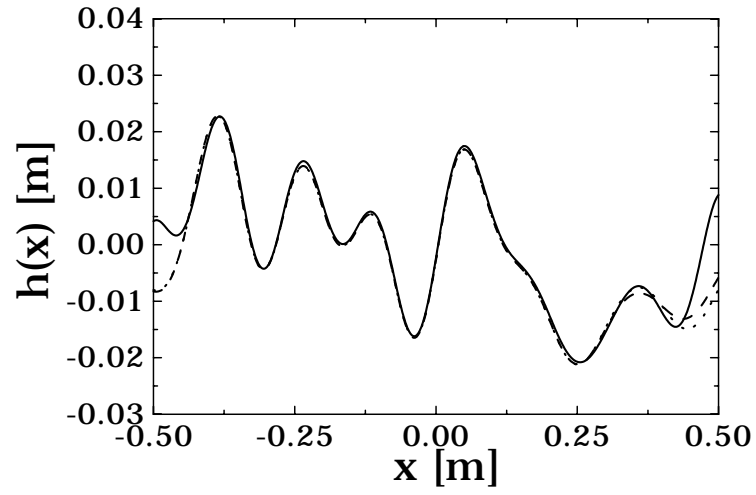


Fig. 11. Parameters as in Fig. 5. Rough surface profile reconstruction via early-time response processing (cf. Sec. III and [13]). The early-time field at the receivers is sampled at $N_t = 50$ time instants with the observation windows chosen so as to roughly gate out causal scattering contributions from regions deeper than $\sim 8\text{cm}$ below nominal ground. — Actual profile; - - - Reconstruction with observation data corrupted by a $\pm 10\%$ additive uniform noise; ····· Reconstruction with -5% error in ϵ_{r1} and σ_1 .

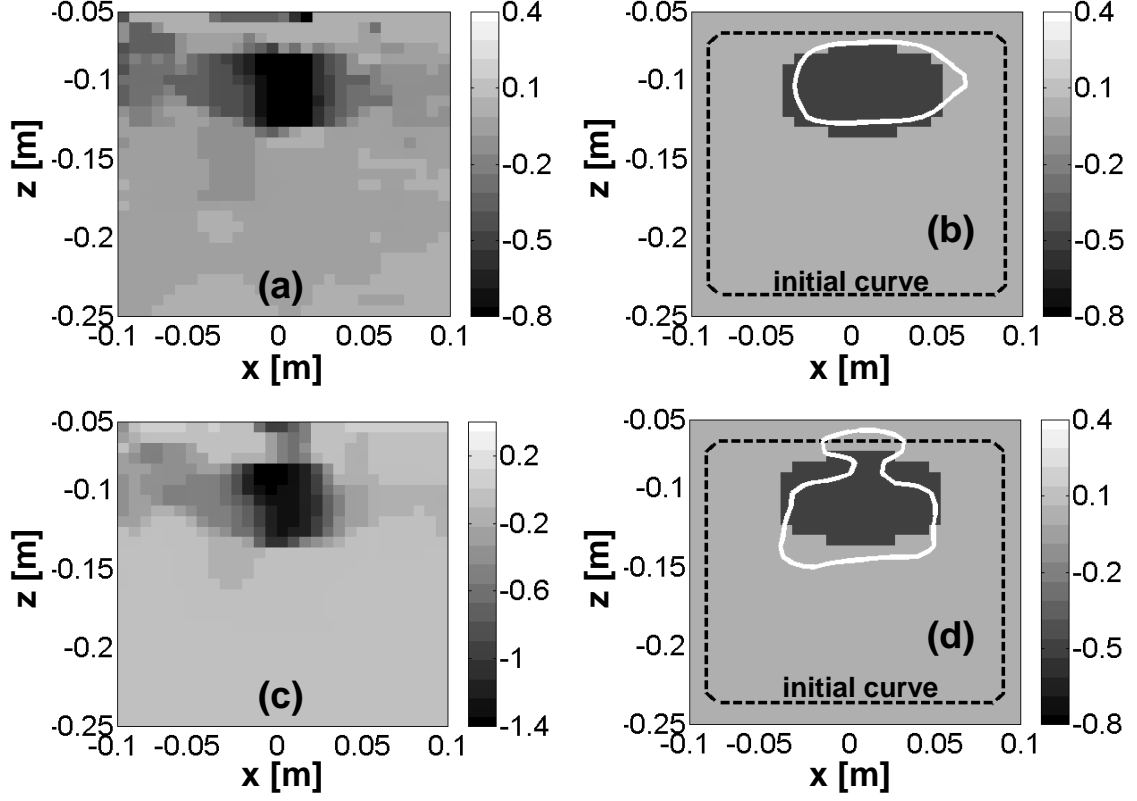


Fig. 12. Parameters as in Fig. 5, but with $d = 0.8$. Reconstruction examples with corrupted data via TVL regularization and CE. (a): TVL reconstruction with field data observations corrupted by $\pm 10\%$ additive uniform noise ($\beta_1 = 3.4 \cdot 10^{-6}$, $\nu = 0.1$, $\beta_2 = 2 \cdot 10^{-3}$, $\Delta e_t = -21\text{dB}$, $\Delta e_b = -31\text{dB}$). (b): As in Fig. 12a, but CE reconstruction (white curve) superposed on ground truth (150 steps with $\beta = 3.5 \cdot 10^{-5}$ plus 627 steps with $\beta = 0.3 \cdot 10^{-5}$; estimated target permittivity: $\epsilon_{r2} = 3.56$, i.e., 1.7% error; $\Delta e_t = -23\text{dB}$, $\Delta e_b = -39\text{dB}$). (c): TVL reconstruction with -5% error in ϵ_{r1} and σ_1 ($\beta_1 = 3.4 \cdot 10^{-6}$, $\nu = 0.1$, $\beta_2 = 2 \cdot 10^{-3}$, $\Delta e_t = -15\text{dB}$, $\Delta e_b = -22\text{dB}$). (d): As in Fig. 12c, but CE reconstruction (white curve) superposed on ground truth (150 steps with $\beta = 3.5 \cdot 10^{-5}$ plus 640 steps with $\beta = 0.21 \cdot 10^{-5}$; estimated target permittivity: $\epsilon_{r2} = 3.56$, i.e., 1.7% error; $\Delta e_t = -26\text{dB}$, $\Delta e_b = -25\text{dB}$). The *estimated* air-soil interface profiles in Fig. 11 are used in the forward scattering model.

Approaches to underground imaging for object localization

Vincenzo Galdi^{a,b}, William C. Karl^a,
David A. Castañón^a, and Leopold B. Felsen^c

^aMDSP Lab., Dept. of Electrical & Computer Eng., Boston University, Boston, MA, USA

^bWaves Group, University of Sannio, Benevento, Italy

^cDept. of Aerospace & Mechanical Eng. and Dept. of Electrical and Computer Eng.
Boston University, Boston, MA, USA

Also, University Professor Emeritus, Polytechnic University, Brooklyn, NY, USA

ABSTRACT

In this paper, we work towards a robust approach for imaging weak-contrast objects buried under a rough soil/air interface using data from an electromagnetic GPR (ground penetrating radar) array. A major source of variability in observed GPR signals is due to reflection from a rough and random ground. Our approach to imaging is based on use of physical and statistical modeling techniques to estimate and compensate for this rough soil/air interface. An approximate physical model based on Gaussian beams is used to model the interaction of the illumination with the ground and estimate the surface profile. This estimated surface profile is then used to correct the raw data for the effects of the rough surface. The corrected data may subsequently be used to reconstruct the subsurface region and localize anomalies. In this stage, statistical models can be used to account for both noise and residual unmodeled effects.

Keywords: Ground penetrating radars, rough surfaces, Gaussian beams

1. INTRODUCTION

In applications such as humanitarian demining, one is interested in detecting shallowly buried small objects with conductivity properties which differ slightly from the surrounding medium. When using ground penetrating radar (GPR) signals, a major source of variability in the received signals is due to reflections from the variable, unknown surface profile. Since the objects of interest are shallowly buried, the backscattered signals from the subsurface objects are difficult to separate from the stronger surface reflections.

In this paper, we work towards a robust approach for imaging weak-contrast objects buried under a rough soil/air interface using data from an electromagnetic GPR array. In particular, we are interested in the use of physical and statistical modeling techniques to estimate and compensate for the effects of unknown air/soil interfaces. We propose to estimate the nature of this interface, and use the estimated characteristics to isolate the return obtained by subsurface scattering. We use an approximate physical model based on Gaussian beams to model the reflected and transmitted fields at the air/soil interface parametrically, in terms of an unknown ground shape. The Gaussian beam model is fast, which allows for real-time generation of parameter estimates. Based on the received signals at the GPR array and the field models, we obtain an estimated surface profile, which is then used to correct the received data to compensate for the effects of the rough surface.

The corrected data may subsequently be used to reconstruct the subsurface region and localize anomalies. In this stage, statistical models can be used to account for noise, estimation error and residual unmodeled effects. In the present paper we present the Gaussian beam air/soil interaction model, and the resulting estimation algorithm for surface profile. We outline the subsequent steps on how the subsurface imaging process would proceed. We plan to evaluate this algorithm on experimental data collected at Northeastern University's subsurface test facility using both a forward-looking and a downward-looking GPR array.

The rest of this paper is organized as follows: In the next section, we overview the formulation and solution of the rough surface reconstruction problem using our proposed Gaussian beam models. To simplify the notation and exposition, we restrict our discussions to a 2-D geometry. Extensions of the results to 3-D geometries are included in our references. In Section 3, we present the approach for constructing subsurface images using diffraction tomography ideas integrated with the Gaussian Beam models. Section 4 presents a summary and future directions for investigation.

2. ROUGH SURFACE RECONSTRUCTION

2.1. Statement of the Problem

Most available algorithms for rough surface estimation have concentrated on conducting surfaces and time-harmonic excitation, and usually require densely sampled measurements. For instance, Wombell and DeSanto¹ used Kirchhoff approximations and Fourier transforms to estimate surface profiles based on measurements of the scattered field in all spectral directions. Ying and Noguchi² used nonlinear optimization techniques for direct estimation of surfaces illuminated by monochromatic Gaussian beams, based on the far-field scattering amplitude for all spectral directions. In a different approach, Schatzberg and Devaney³ used the Rytov approximation and backpropagation to estimate surface profiles based on full measurements (amplitude and phase) of the scattered wave. In contrast with the above contributions, our investigation here is concerned with moderately rough interfaces between air and a homogeneous dielectric half space (soil), and with the estimation of the interface profile from a limited number of spatially/temporally sampled reflected field measurements, using limited aperture illumination, as is typical for realistic ground penetrating radar (GPR) configurations. In a recent investigation⁴ we addressed this problem for narrow-band stepped-frequency GPR configurations by exploiting the time-harmonic Gabor-based Gaussian beam forward scattering model in Ref. 5. Here, we extend the approach to wide-band pulsed illumination, typical of many realistic GPR systems.

The proposed extension is based on a recently developed Gabor-based narrow-waisted pulsed-beam (PB) representation of short-pulse scattering by moderately rough dielectric interfaces.⁶ The problem geometry is sketched in Fig. 1: in an (x, z) two-dimensional (2D) coordinate space, a pulsed field with transverse magnetic (TM) polarization is assumed to impinge vertically from free-space onto a lossy, homogeneous dielectric half-space (soil) of known relative permittivity ϵ_r and conductivity σ , bounded by a moderately rough interface described by the continuous function $h(x)$. We assume that the transmitting array generates a y -directed tapered well-collimated pulsed field with transverse aperture width d , which can be approximated by a pulsed truncated tapered plane wave propagating in the negative z -direction (see Fig. 1),

$$e^i(x, z, t) \sim g(x)p(t - c^{-1}z), \quad (1)$$

where $p(t)$ is a pulse of length $T \ll d/c$, with c representing the speed of light, and $g(x)$ is the taper profile.

The 2D y -directed reflected field e^r is sampled at N_t time instants at N_r receiver locations (at $z = z^r$). To begin with, we ignore the presence of buried objects and the *noisy* (incoherent) contribution of finer-scale roughness, and we seek to estimate the coarse shape of the surface, acknowledging the implicit limits of retrievable information through inverse scattering. As in Ref. 4, due to the inherent *ill-posedness* of this inverse scattering problem, we exploit a *robust* inversion strategy based on an appropriate compact low-dimensional representation of the surface, which provides an “implicit” regularization. The surface estimation problem is thus posed as a nonlinear optimization problem, whose solution requires repeated evaluations of the reflected field at the receiver locations. The efficiency of such an approach strongly relies on the availability of a fast forward scattering model. An efficient forward solver is necessary also for computing the field transmitted into the soil, as required in Born-approximation-based techniques for underground imaging later on (see Sec. 3.1).

2.2. Forward Scattering Model

The forward scattering model is detailed in Ref. 6. It is based on the Kirchhoff Physical Optics (PO) approximation in conjunction with the Gabor-based narrow-waisted PB discretization of 1D aperture field distributions in Ref. 7. In this section, we briefly review the basic theory underlying the algorithm.

In the asymptotic high-frequency range, and for smooth roughness over all wavelength scales in the pulse spectrum, the reflected field spectrum is approximated by the PO integral

$$E^r(x, z, \omega) \sim - \int_{C_{PO}} J_{PO}(x', \omega) \frac{\partial}{\partial \zeta} G(x, z; x', h(x'); k_0) d\ell', \quad (2)$$

where C_{PO} extends over the illuminated portion of the 1D interface $z = h(x)$, $d\ell'$ is the incremental arc-length measured along the surface tangent, and $\partial/\partial\zeta$ denotes the normal derivative. The incident field tapering is chosen so that the illuminated portion of the interface, C_{PO} , is essentially confined to the interval $|x| \leq d/2$. Furthermore, G is the frequency-domain (FD) line-source free-space Green’s function,

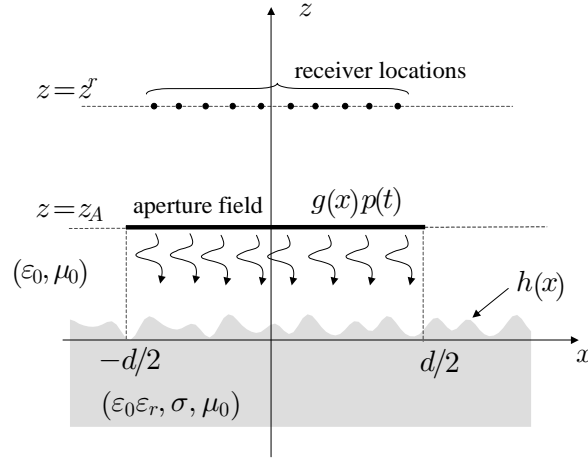


Figure 1. Setup geometry for the inverse scattering problem. A TM polarized pulsed EM field illuminates a homogeneous dielectric half-space with known relative permittivity ϵ_r and conductivity σ and with a moderately rough interface whose coarse scale profile is described by the continuous function $h(x)$. The reflected field is sampled at N_r receiver locations and N_t time instants, at $z = z^r$.

$$G(x, z; x', z'; k) = \frac{i}{4} H_0^{(1)}(kR), \quad R = [(x - x')^2 + (z - z')^2]^{1/2}, \quad (3)$$

with $H_0^{(1)}(\cdot)$ denoting the zeroth order Hankel function of the first kind. The induced PO “surface current” density J_{PO}^r is given by twice the tangential reflected field at the interface, obtained from the canonical solution of infinite plane-wave scattering by a plane dielectric boundary locally tangent to the rough surface profile

$$J_{PO}^r(x, \omega) = 2\mathcal{R}(x, \omega) E^i(x, h(x), \omega). \quad (4)$$

In (4), E^i is the frequency spectrum of (1), and \mathcal{R} denotes the local TM plane-wave Fresnel reflection coefficient,

$$\mathcal{R}(x, \omega) = \frac{\cos \theta^i - (\epsilon_r^e - \sin^2 \theta^i)^{1/2}}{\cos \theta^i + (\epsilon_r^e - \sin^2 \theta^i)^{1/2}}, \quad (5)$$

where $\theta^i(x) = -\tan^{-1}[h'(x)]$, $' \equiv d/dx$, is the local incident angle relative to the normal, and $\epsilon_r^e(\omega) = \epsilon_r + i\sigma/(\omega\epsilon_0)$ is the complex, frequency dependent, *effective* permittivity of the soil. In the following, we restrict consideration to slightly lossy soils, i.e.,

$$\frac{\sigma}{\omega\epsilon_0\epsilon_r} \ll 1, \quad \omega < \Omega, \quad (6)$$

with Ω denoting the effective bandwidth of the pulse $p(t)$. This condition is typically fulfilled in many GPR applications^{8,9} and significantly simplifies the time domain (TD) inversion.⁶

The PO radiation integral in (2) is not very different from the Kirchhoff aperture integrals analyzed in Ref. 7, except that the line integration in (2) is performed along a rough surface profile instead of on a projected equivalent planar aperture. In Ref. 7, the 1D space-time aperture field distribution is first parameterized in the FD in terms of x -domain discretized m -indexed Gabor basis functions with narrow width L , centered on the Gabor lattice points $x_m = mL$; these initial conditions generate *narrow-waisted*, quasi-ray, complex-source-point Gaussian beams. For Rayleigh pulses,

$$p(t) = \Xi \frac{d^j}{dt^j} \exp \left[- \left(\frac{t - T/2}{\zeta T} \right)^2 \right], \quad (7)$$

the resulting TD analytic Fourier inversion integral can be approximated by rapidly computable closed form expressions. The same approach can be applied to discretizing PO integrals as in (2). One obtains the following approximate pulsed beam (PB) expansion for the reflected field (see Ref. 6 for details)

$$e^r(x, z, t) \sim \sum_{|m| \leq (d/2L)} c_m^r b_m^r(x, z, t - t_m^r), \quad (8)$$

where the Gabor expansion coefficients c_m^r and the time delay t_m^r are approximated by sampling the TD-PO surface current at the lattice points $x_m = mL$,

$$c_m^r = (L/\sqrt{2})^{1/2} g(x_m) \mathcal{R}_0(x_m), \quad t_m = c^{-1} (z_A - z_m - x_m \sin \theta_m^i \cos \theta_m^i), \quad (9)$$

where $z_m = h(x_m)$, $\theta_m^i \equiv \theta^i(x_m)$ is the local incidence angle, and $\mathcal{R}_0(x_m)$ is the local Fresnel reflection coefficient evaluated at the center angular frequency Ω_0 of the pulse. The PB propagator in (8) is given by⁶

$$b_m^r(x, z, t) = \text{Re} \left\{ (-1)^j \beta_m^r \left[\mathcal{T}_m^r \Gamma\left(\frac{3+2j}{4}\right) M_1^{(j)}\left(\frac{t - \tau_m^r}{\mathcal{T}_m^r}\right) - 2i(t - \tau_m^r) \Gamma\left(\frac{5+2j}{4}\right) M_2^{(j)}\left(\frac{t - \tau_m^r}{\mathcal{T}_m^r}\right) \right] \right\}, \quad (10)$$

where

$$\beta_m^r = 2^{j+1/2} \pi^{-1/2} (\mathcal{T}_m^r)^{-j-5/2} \Lambda_m^r \Xi_\zeta T, \quad \Lambda_m^r = -i 2^{5/4} \exp(i\pi/4) c^{-1} \sqrt{\frac{L}{8\pi}} \frac{\zeta_m}{(R_m^r)^{3/2}}, \quad (11)$$

$$\zeta_m = (x - x_m) \sin \theta_m^i + (z - z_m) \cos \theta_m^i, \quad (12)$$

$$\tau_m^r = c^{-1} (R_m^r + x_m \sin \theta_m^i \cos \theta_m^i + T/2), \quad \mathcal{T}_m^r = \frac{L \cos^2 \theta_m^i}{2c\sqrt{2\pi}} \sqrt{1 - \frac{z_{bm}^r}{R_m^r}}, \quad T_m^r = \sqrt{(\mathcal{T}_m^r)^2 + \varsigma^2 T^2}, \quad (13)$$

and (x_{bm}^r, z_{bm}^r) are beam-centered coordinates,

$$\begin{bmatrix} x_{bm}^r \\ z_{bm}^r \end{bmatrix} = \begin{bmatrix} \cos(2\theta_m^i) & -\sin(2\theta_m^i) \\ \sin(2\theta_m^i) & \cos(2\theta_m^i) \end{bmatrix} \begin{bmatrix} x - x_m \\ z - z_m \end{bmatrix}, \quad R_m = [(x_{bm}^r)^2 + (z_{bm}^r)^2]^{1/2}. \quad (14)$$

Moreover, in (10), $\Gamma(\cdot)$ denotes the gamma function,¹⁰ and

$$M_1^{(j)}(t) = {}_1F_1\left(\frac{3+2j}{4}, \frac{1}{2}, -t^2\right), \quad M_2^{(j)}(t) = {}_1F_1\left(\frac{5+2j}{4}, \frac{3}{2}, -t^2\right), \quad (15)$$

with ${}_1F_1(u, v, t)$ denoting the Kummer confluent hypergeometric function.¹⁰ The functions $M_{1,2}^{(j)}$ can be computed efficiently using a rapidly converging power series expansion derived in Ref. 6. The propagator in (10) is a pulsed Gaussian beam, whose collimation is controlled by the discretization period L . The expansion in (8) has been thoroughly calibrated against a rigorous reference solution,⁶ and has been found to provide accurate and robust syntheses for $L \ll d$, i.e., for *narrow-waisted*, poorly collimated PB propagators. The corresponding PB expansion for the transmitted field can be obtained similarly (see Sec. 3.1), as can results for oblique incidence.⁶ Extension to 2D surfaces, and 3D field scattering can be derived based on 2D aperture PB discretization.¹¹

2.3. Surface Estimation

As noted earlier, the proposed *robust* inversion strategy is based on a low-dimensional compact geometrical parameterization of the unknown interface profile, whose unknown parameters are estimated by *fitting* the model-based forward scattering prediction to the available (measured/simulated) data, i.e., minimizing a suitable *cost functional*. The choice of the interface profile parameterization requires tradeoff between compactness and versatility, keeping in mind that the number of unknown parameters N to be estimated must be *smaller* than the collected scattered field data size, i.e., $N \sim N_r \times N_t$. As in Ref. 4, we parameterize the rough surface coarse shape in terms of quartic splines,¹²

$$h(x) \approx \sum_{n=1}^N c_n s_n^{(4)}(x), \quad (16)$$

where c_n are unknown coefficients, and $s_n^{(4)}$ represents a quartic B-spline basis function.¹²

Let $\hat{e}_{pq}^r = \hat{e}^r(x_p^r, z^r, t_q)$ denote the y -directed reflected field *measured* at time t_q at sensor position x_p^r . Given a vector of spline coefficients $\underline{c} = \{c_1, \dots, c_N\}$ and the incident field in (1), we use the PB algorithm in Sec. 2.2 for the surface profile in (16) to generate predictions of the reflected field waveforms at each receiver. Let $e_{pq}^r = e^r(x_p^r, z^r, t_q; \underline{c})$ denote the *model-based prediction* of the reflected field at time t_q at sensor position x_p^r , for surface profile coefficients \underline{c} . Accordingly, we define the cost functional as follows

$$J(\underline{c}) = \|\underline{e}^r(\underline{c}) - \hat{\underline{e}}^r\|^2 = \sum_{p=1}^{N_r} \sum_{q=1}^{N_t} \gamma_p (e_{pq}^r - \hat{e}_{pq}^r)^2, \quad \gamma_p = \left[\max_q \{\hat{e}_{pq}^r\} \right]^{-2}. \quad (17)$$

As in Ref. 4, the *regularized* inverse scattering problem is formalized as finding the coefficient vector \underline{c} in (16) which minimizes the cost functional $J(\underline{c})$ in (17), i.e. finding $\hat{\underline{c}}$ such that

$$\hat{\underline{c}} = \arg \min_{\underline{c}} J(\underline{c}). \quad (18)$$

The cost functional in (18) is generally *non-convex* with respect to \underline{c} , and therefore its optimization requires particular care due to the possible presence of local minima. In the narrow-band frequency range investigated in Ref. 4, it was found that the smoothness of the cost functional was strongly dependent on the frequency content of the excitation field, and the global optimization was achieved via a *continuation method* based on the selected use of the data at the various available frequencies. For the pulsed excitation of interest here, the pulse length cT plays the key role. In particular, to enhance resolution in the reconstruction, the pulse should be as short as possible. However, an exceedingly wide-band excitation would most likely result in a highly non-convex cost functional with many local minima, whose global minimization would be very expensive from the computational viewpoint. Furthermore, for underground imaging, it is also important to operate at sufficiently low frequencies for adequate soil penetration. In our numerical experiments, we found that a suitable trade-off between these contrasting requirements can be actually achieved by using typical wide-band pulses in the GPR frequency range.

2.4. Results

In order to test our surface profile estimation strategy, we generated synthetic field measurement data using a rigorous reference solution of the forward scattering problem (see Ref. 6 for details), with the (fourth-order Rayleigh) wide-band pulse excitation in Fig. 2 (2.45 GHz center frequency) and a cosine-tapered transverse field distribution,

$$g(x) = \begin{cases} \cos(\pi x/d), & |x| \leq d/2, \\ 0, & |x| > d/2, \end{cases} \quad (19)$$

on an aperture of width $d = 1\text{m}$. The chosen excitation, typical of realistic GPR applications, was found to yield relatively smooth cost functionals, whose minimization was performed via the Polak-Ribiere version of the conjugate gradient (CG) algorithm,¹³ assuming as initial guess a flat interface at $z = 0$. A typical reconstruction example is shown in Fig. 3, for the simulation parameters listed in the figure caption. The soil parameters were chosen so as to simulate a class of realistic soils (Puerto Rican clay loam⁹). The surface realization, parameterized in this experiment with 20 B-spline basis functions, was generated randomly so as to mimic typical ($\sim 2 - 3$ cm peak-to-peak) natural roughness. In this experiment, $N_r = 11$ equispaced receivers were used, with $N_t = 100$ time samples, and convergence of the CG minimization was achieved in 20 iterations. As one can see, the reconstruction is reasonably accurate throughout most of the interval, except near the edges of the illumination region. This kind of accuracy was observed in many numerical experiments, with typical overall computing times of 1 min. on a laptop PC.

3. SUBSURFACE IMAGING

Having established the possibility of estimating the coarse shape of the rough air-ground interface, we now turn to the problem of *subsurface imaging* in the presence of a *known* moderately rough interface.

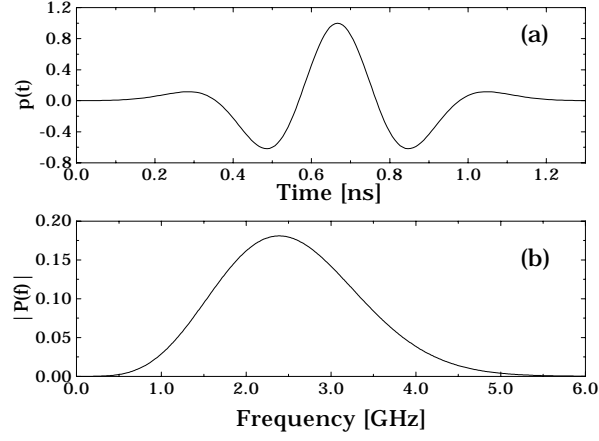


Figure 2. Fourth-order Rayleigh pulse in (7) ($j = 4$, $T = 1.3$ ns, $\Xi = T/30000$, $\varsigma = 1/\sqrt{50}$). (a): Temporal profile; (b): Spectrum (magnitude).

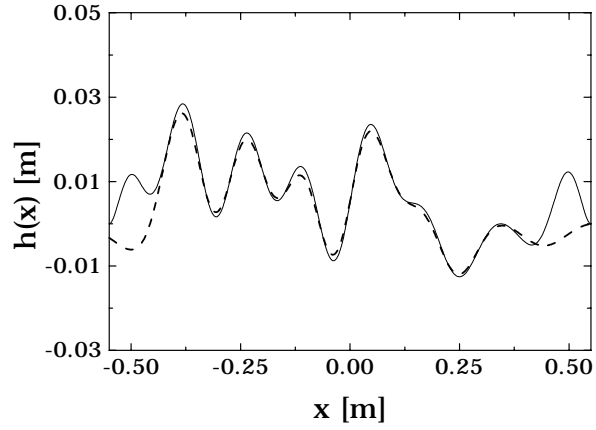


Figure 3. Rough-surface reconstruction example. Soil parameters: $\epsilon_r = 4.5$, $\sigma = 0.012$ S/m. Incident field in (1), with $g(x)$ in (19), and $p(t)$ as in Fig. 2, and with $z_A = 0.2$ m and $d = 1$ m. Collected data: $N_r = 11$ receivers at $z_{obs} = 0.3$ m, equispaced within $[-0.5\text{m}, 0.5\text{m}]$; $N_t = 100$ time samples at each receiver. — True; - - - Reconstruction.

3.1. Forward Scattering Model

We refer to the problem geometry in Fig. 1, but now include a buried target with dielectric permittivity $\epsilon_r^{(tar)}$ and conductivity $\sigma^{(tar)}$ occupying the region $D^{(tar)}$ in the half-space $z < h(x)$. In the frequency domain, the TM-polarized *total* field observed at the receiver location $\mathbf{r}^r = (x^r, z^r)$ can be written as

$$E(\mathbf{r}^r, \omega) = E^b(\mathbf{r}^r, \omega) + E^s(\mathbf{r}^r, \omega), \quad (20)$$

where E^b represents the *background field* (i.e., the field in the absence of the target), and

$$E^s(\mathbf{r}^r, \omega) = \frac{\omega^2}{c^2} \int \int_{D^{(tar)}} G_b(\mathbf{r}^r, \mathbf{r}', \omega) E(\mathbf{r}', \omega) O(\mathbf{r}', \omega) d\mathbf{r}' \quad (21)$$

is the field scattered by the target. In (21), G_b represents the FD Green's function of the rough-interface dielectric half-space, E is the *total* field in the target region, and

$$O(\mathbf{r}', \omega) = [\epsilon_r^{(tar)}(\mathbf{r}') - \epsilon_r] + i \frac{[\sigma^{(tar)}(\mathbf{r}') - \sigma]}{\omega \epsilon_0} = \Delta \epsilon_r(\mathbf{r}') + i \frac{\Delta \sigma(\mathbf{r}')}{\omega \epsilon_0} \quad (22)$$

is referred to as the *object function*. The integration in (21) is limited over the region $D^{(tar)}$ in which the object function is nonzero. For pulsed excitation, with time dependence $p(t)$, by Fourier inversion of (20) and (21) one obtains

$$e(\mathbf{r}^r, t) = e^b(\mathbf{r}^r, t) + e^s(\mathbf{r}^r, t), \quad (23)$$

$$e^s(\mathbf{r}^r, t) = - \int \int_{D^{(tar)}} \left\{ \Delta \epsilon_r(\mathbf{r}') \frac{\partial^2}{\partial t^2} [g_b(\mathbf{r}^r, \mathbf{r}', t) \otimes e(\mathbf{r}', t)] + \frac{\Delta \sigma(\mathbf{r}')}{\epsilon_0} \frac{\partial}{\partial t} [g_b(\mathbf{r}^r, \mathbf{r}', t) \otimes e(\mathbf{r}', t)] \right\} d\mathbf{r}', \quad (24)$$

where g_b represents the space-time Green's function of the rough-interface dielectric half-space (i.e., the Fourier inverse transform of G_b in (21)), e is the *total* field in the target region for pulsed excitation $p(t)$, and \otimes indicates time convolution. Assuming that the rough interface profile is known, one can employ the pulsed beam synthesis in (8) to generate prediction of the *background field* contribution e^b in (23) for a given pulsed excitation, and therefore *isolate* the contribution e^s from the buried target. The relation in (24) between the scattered field at the receivers and the dielectric/conductivity inhomogeneities $\Delta \epsilon$ and $\Delta \sigma$ is *nonlinear*, due to the presence of the total field e which is itself dependent on $\Delta \epsilon$ and $\Delta \sigma$. However, for plastic anti-personnel land mines, of particular interest in our investigation, the dielectric properties of the buried target are typically close to those of the background soil, i.e., $\epsilon_r^{(tar)} \approx \epsilon_r$, $\sigma^{(tar)} \approx 0$; therefore one can resort to approximate linearized approaches valid in the *weak scattering* limit. The simplest such approach is the Born approximation,¹⁴ where one replaces the *total* field e inside the target with the *transmitted* field e^t in the target region, in the absence of the inhomogeneity. For nonconducting targets ($\sigma^{(tar)} \approx 0$, i.e., $\Delta \sigma(\mathbf{r}') \approx 0$), this yields a *linear* model relating the scattered field at the receivers to the dielectric inhomogeneity $\Delta \epsilon_r$,

$$e^s(\mathbf{r}^r, t) \approx - \int \int_{D^{(tar)}} \left\{ \Delta \epsilon_r(\mathbf{r}') \frac{\partial^2}{\partial t^2} [g_b(\mathbf{r}^r, \mathbf{r}', t) \otimes e^t(\mathbf{r}', t)] + \frac{\sigma}{\epsilon_0} \frac{\partial}{\partial t} [g_b(\mathbf{r}^r, \mathbf{r}', t) \otimes e^t(\mathbf{r}', t)] \right\} d\mathbf{r}'. \quad (25)$$

The limitations of this approximation, which neglects multibounce interactions inside the target, have been thoroughly investigated and are well documented in the technical literature (see, e.g., Ref. 15). Moreover, more sophisticated and accurate *nonlinear* versions have been proposed (see, e.g., Refs. 16,17).

To implement the linear model in (25) for a given pulsed excitation, one needs to compute the transmitted field e^t in the target region (in the absence of the target), and the space-time Green's function g_b of the soil. As shown in Sec. 2.2, the computation of the field transmitted through moderately rough dielectric interfaces can be carried out efficiently through the PB algorithm in Ref. 6. The transmitted field PB synthesis is formally analogous to that in (8)–(14) for the reflected field. One obtains (see Ref. 6 for details)

$$e^t(x, z, t) \sim \sum_{|m| \leq (d/2L)} c_m^t b_m^t(x, z, t - t_m^t), \quad (26)$$

with

$$b_m^t(x, z, t) = \text{Re} \left\{ (-1)^j \beta_m^t \left[\mathcal{T}_m^t \Gamma \left(\frac{3+2j}{4} \right) M_1^{(j)} \left(\frac{t - \tau_m^t}{\mathcal{T}_m^t} \right) - 2i(t - \tau_m^t) \Gamma \left(\frac{5+2j}{4} \right) M_2^{(j)} \left(\frac{t - \tau_m^t}{\mathcal{T}_m^t} \right) \right] \right\}, \quad (27)$$

where

$$\beta_m^t = 2^{j+1/2} \pi^{-1/2} (\mathcal{T}_m^t)^{-j-5/2} \Lambda_m^t \Xi_\zeta T, \quad \tau_m^t = c^{-1} [\sqrt{\epsilon_r} (R_{m0}^t + x_m \sin \theta_{m0}^R \cos \alpha_m) + T/2], \quad (28)$$

$$\Lambda_m^t = i(\epsilon_r)^{1/4} 2^{5/4} \exp(-\kappa R_{m0}^t + i\pi/4) c^{-1} \sqrt{\frac{L}{8\pi}} \frac{\zeta_m}{(R_{m0}^t)^{3/2}}, \quad \kappa = \frac{\sigma}{2c\epsilon_0 \sqrt{\epsilon_r}}, \quad (29)$$

$$T_m^t = \sqrt{\epsilon_r} \frac{L \cos \alpha_m \cos \theta_{m0}^R}{2c\sqrt{2\pi}} \sqrt{1 - \frac{z_{bm0}^t}{R_{m0}^t}}, \quad \mathcal{T}_m^t = \sqrt{(T_m^t)^2 + \zeta^2 T^2}, \quad (30)$$

$$R_{m0}^t = \sqrt{(x_{m0}^t)^2 + (z_{m0}^t)^2}, \quad \begin{bmatrix} x_{bm0}^t \\ z_{bm0}^t \end{bmatrix} = \begin{bmatrix} \cos \gamma_{m0}^t & \sin \gamma_{m0}^t \\ \sin \gamma_{m0}^t & -\cos \gamma_{m0}^t \end{bmatrix} \begin{bmatrix} x - x_m \\ z - h(x_m) \end{bmatrix}, \quad (31)$$

$$\theta_{m0}^R \equiv \sin^{-1} (\sin \theta_m^i / \sqrt{\epsilon_r}), \quad \gamma_{m0}^t = \theta_{m0}^R + \alpha_m. \quad (32)$$

Figure 4 shows a typical transmitted field waveform, for the excitation and specified rough surface geometry in Fig. 3, computed via the PB synthesis in (26)–(32) (with 50 beams) and via the rigorous reference solution detailed in Ref. 6. As for the reflected field, the PB synthesis provides a robust, fast and reasonably accurate approximation for the transmitted field, which can be fruitfully exploited for efficient implementation of (25).

The computation of the space-time Green's function g_b for the rough-interface dielectric half-space geometry could be accomplished by generalizing the transmitted field PB synthesis in (26)–(32) to the case of cylindrical incident wavefronts and impulsive time excitation. However, for the present, we are interested in a more efficient approach to the PB expansion of the *whole* convolution $g_b \otimes e^t$ and its time derivatives in (25). This approach, now under investigation, involves more cumbersome algebra, but would substantially increase the computational efficiency, avoiding the time-consuming numerical convolution.

3.2. Subsurface Reconstruction

We now discuss how to use the forward scattering model to invert GPR data and generate estimates of the subsurface property distribution – i.e. how to do imaging. Such an inversion scheme can be constructed by discretizing the test domain $D^{(test)}$ to be imaged into a number of adequately small pixels and relating these values to the observations of the scattered field waveforms at the receivers through the linear forward scattering model in (25). Assuming that N_t time samples of the scattered field e^s are collected at N_r receivers, and that the test domain is discretized into N_p pixels, the forward model in (25) can be discretized accordingly and cast into matrix form as:

$$\underline{y} = \underline{A} \cdot \underline{x} + \underline{n}, \quad (33)$$

where \underline{y} is a column vector containing the $N_r \cdot N_t$ time samples of the *known term* in (25),

$$e^s(\mathbf{r}^r, t) + \frac{\sigma}{\epsilon_0} \int \int_{D^{(test)}} \frac{\partial}{\partial t} [g_b(\mathbf{r}^r, \mathbf{r}', t) \otimes e^t(\mathbf{r}', t)] d\mathbf{r}', \quad (34)$$

\underline{x} is a N_p -element column vector containing the *unknown* dielectric contrast $\Delta\epsilon_r$ at each pixel, \underline{A} is a $(N_r \cdot N_t) \times N_p$ matrix containing the space-time discretization of the integral

$$- \int \int_{D^{(test)}} \frac{\partial^2}{\partial t^2} [g_b(\mathbf{r}^r, \mathbf{r}', t) \otimes e^t(\mathbf{r}', t)] d\mathbf{r}', \quad (35)$$

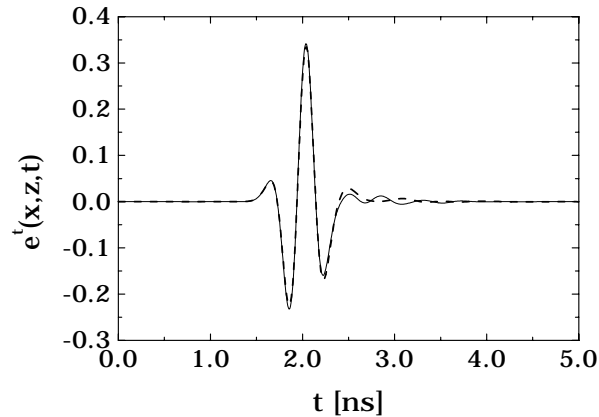


Figure 4. Parameters as in Fig. 3. Transmitted field through the true rough interface profile observed at $x = 0$, $z = -0.1\text{m}$. — Reference solution; - - - PB synthesis in (26)–(32) (50 beams).

and the *noise* vector \underline{n} accounts for measurement uncertainty and unmodeled small scale *clutter*. Due to the inherent ill-posedness of this inverse scattering problem, and the various approximations in our model, the reconstruction *cannot* be simply accomplished via the (pseudo-)inverse of the matrix \underline{A} . In order to obtain reasonably accurate reconstructions, it is essential to incorporate stabilization steps, known as regularization. In a number of recent investigations restricted to flat air-ground interfaces,^{18,19} we have explored various *object-based* regularization approaches, such as *total variation*¹⁸ and *curve evolution*,¹⁹ which have been found to provide quite accurate and robust reconstructions. We are currently working on extending these approaches to the more challenging case of rough interfaces.

4. CONCLUSIONS

We have presented an approach for imaging objects buried under a rough air/soil interface from ground penetrating radar data. Our approach is based on estimation of the rough surface followed by subsurface imaging. We have a Gaussian-beam approximate physical model for the scattering problem that is efficient and well matched to the estimation/inversion problem. This model was demonstrated for the problem of surface profile estimation. We presented an approach to the subsequent subsurface estimation problem based on an approximate Greens function for the subsurface scattering.

We are currently pursuing several extensions to the present work. We are considering extensions to include sensor fusion, particularly using sensors which provide surface information such as optical range sensors. This information would improve the surface estimates, and allow for more accurate isolation of subsurface scattering signals. We are also extending our models to represent two different physical GPR systems for which we have experimental data: a downward-looking GPR array and a forward-looking GPR array. Once these models are complete, we will evaluate their performance on the collected GPR data.

ACKNOWLEDGMENTS

This work was supported by ODDR&E under MURI Grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from Grant No. 9900448 by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University, Brooklyn, NY USA.

REFERENCES

1. R. J. Wombell and J. A. DeSanto, “The reconstruction of shallow rough-surface profiles from scattered field data,” *Inverse Probl.* **7**, pp. L7–L12, 1991.

2. C. Ying and A. Noguchi, "Rough surface inverse scattering problem with Gaussian beam illumination," *IEICE Trans. Electron.* **E77-C**, pp. 1781–1785, 1994.
3. A. Schatzberg and A. J. Devaney, "Rough surface inverse scattering with the Rytov approximation," *J. Opt. Soc. Am. A* **10**, pp. 942–950, 1993.
4. V. Galdi, D. A. Castañón, and L. B. Felsen, "Reconstruction of moderately rough interfaces via quasi-ray Gaussian beams," submitted to *IEEE Trans. Geosci. Remote Sensing*, Sept. 2000.
5. V. Galdi, L. B. Felsen, and D. A. Castañón, "Quasi-ray Gaussian beam algorithm for time-harmonic two-dimensional scattering by moderately rough interfaces," *IEEE Trans. Antennas Propagat.* **49**(9), 2001 (in print).
6. V. Galdi, L. B. Felsen, and D. A. Castañón, "Quasi-ray Gaussian beam algorithm for short-pulse two-dimensional scattering by moderately rough dielectric interfaces," submitted to *IEEE Trans. Antennas Propagat.*, Mar. 2001.
7. V. Galdi, L. B. Felsen, and D. A. Castañón, "Narrow-waisted Gaussian beam discretization for short-pulse radiation from one-dimensional large apertures," *IEEE Trans. Antennas Propagat.* **49**(10), 2001 (in print).
8. T. B. Hansen and P. M. Johansen, "Inversion scheme for ground penetrating radar that takes into account the planar air-soil interface," *IEEE Trans. Geosci. Remote Sensing* **38**, pp. 496–506, 2000.
9. J. E. Hipp, "Soil electromagnetic parameters as functions of frequency, soil density, and soil moisture," *Proc. IEEE* **62**, pp. 98–103, 1974.
10. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, NY, 1964.
11. V. Galdi, L. B. Felsen, and D. A. Castañón, "Time-domain radiation from large two-dimensional apertures via narrow-waisted Gaussian beams," submitted to *IEEE Trans. Antennas Propagat.*, Dec. 2000.
12. L. L. Shumaker, *Spline Functions: Basic Theory*, Wiley, New York, NY, 1981.
13. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing* (2nd ed.), Cambridge Univ. Press, Cambridge, UK, 1992.
14. W. C. Chew, *Waves and Fields in Inhomogeneous Media*, Oxford Press, Oxford, UK, 1996.
15. J.B. Keller, "Accuracy and validity of the Born and Rytov approximations," *J. Opt. Soc. Am.* **59**, pp. 1003–1004, 1969.
16. T. M. Habashy, R. W. Groom, and B. Spies, "Beyond the Born and Rytov approximation: A nonlinear approach to electromagnetic scattering," *J. Geophys. Res.* **98**, pp. 1759–1775, 1993.
17. N. Dasgupta, N. Geng, T. Dogaru, and L. Carin, "On the Extended-Born technique for scattering from buried dielectric targets," *IEEE Trans. Antennas Propagat.* **47**, pp. 1739–1743, 1999.
18. H. Feng, D. A. Castañón, and W. C. Karl, and E. L. Miller, "GPR imaging approaches for buried plastic landmine detection," in *Detection and Remediation Technologies for Mines and Minelike Targets V*, A. C. Dubey, J. F. Harvey, J. T. Broach, and R. E. Dugan, Eds., *Proc. SPIE* **4038**, p. 1485–1496, 2000.
19. H. Feng, D. A. Castañón, and W. C. Karl, "A shape-based approach to buried landmine imaging," *Proc. 5th Joint Conf. on Information Sciences (JCIS 2000)*, Atlantic City, NJ, Feb. 27-Mar. 3, 2000.

TWO-DIMENSIONAL SCATTERING BY MODERATELY ROUGH DIELECTRIC INTERFACES VIA NARROW-WAISTED GAUSSIAN BEAMS*

Vincenzo Galdi^(1,2), Leopold B. Felsen^(1,3), and David A. Castañon⁽¹⁾

⁽¹⁾ Dept. of Electrical & Computer Eng., Boston University, 8 Saint Mary's Street, Boston, MA 02215, USA

⁽²⁾ Waves Group, University of Sannio, I-82100, Benevento, Italy

⁽³⁾ Dept. of Aerospace & Mechanical Eng., Boston University, 110 Cummington St., Boston, MA 02215, USA
Also University Professor Emeritus, Polytechnic University, Brooklyn, NY 11201 USA

Abstract: For one-dimensional aperture excitations, a two-dimensional Gabor-based Gaussian beam (GB) algorithm is presented which describes field scattering by, and transmission through, a moderately rough interface between two dielectric media. The proposed algorithm extends the results in Maciel and Felsen[1], and is shown to produce accurate and computationally efficient solutions for this complex propagation environment, over a range of calibrated combinations of the problem parameters.

INTRODUCTION AND PROBLEM FORMULATION

Gabor-based Gaussian beam (GB) algorithms, in conjunction with the complex source point (CSP) method for generating beam-like wave objects, have found application in a variety of high frequency wave propagation and diffraction scenarios. Of special interest for efficient numerical implementation is the non-collimated narrow-waisted species of GB which reduces the computationally intensive complex ray tracing for collimated GB propagation and scattering to quasi-real ray tracing, without the failure of strictly real ray field algorithms in caustic and other transition regions. The Gabor-based narrow-waisted CSP-GB method has been applied previously (Maciel and Felsen[1]) to two-dimensional (2D) propagation from extended nonfocused and focused aperture distributions through arbitrarily curved 2D layered environments. In the present 2D study, the method is applied to 1D-aperture-excited field scattering from, and transmission through, a moderately rough interface between two dielectric media, as depicted in Fig. 1. In the 2D coordinate space (x, z) , a y -directed electric field with time-harmonic dependence $\exp(-i\omega t)$ and spatial distribution $f(x)$ is assumed to occupy the aperture region $|x| \leq d/2$ at $z = z_A$ in free space, i.e., $E^{inc}(x, z_A) = f(x), |x| \leq d/2$; $E^{inc}(x, z_A) = 0, |x| > d/2$. The aperture-radiated field illuminates a homogeneous dielectric half-space, with relative permittivity ϵ_r , bounded by a moderately rough (both in height and slope) interface described by a continuous function $z = h(x)$. Our ultimate goal is to develop an efficient forward scattering solver for inverse problems concerned with profile and object reconstruction (Galdi et al.[2]).

GABOR-BASED GAUSSIAN BEAM DISCRETIZATION

Aperture-Radiated Field. The Gabor-based CSP-GBs are generated by Gaussian initial fields in the x -domain ($z = z_A$) aperture plane. The initial fields are centered on a discretized Gabor lattice in the (x, k_x) phase space indexed by (x_m, k_{xn}) , with $(m, n) = 0, \pm 1, \pm 2, \dots$, where k_x is the x -domain spectral wavenumber; L_x and β_x denote the spatial and spectral lattice periods, respectively, subject to the constraint $L_x \beta_x = 2\pi$. Thus, m tags the spatial displacement $x_m = mL_x$, while n tags the spectral displacement (tilt) $k_{xn} = n\beta_x$ of the Gabor-GBs. As

* Work supported by ODDR&E under MURI grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from Grant No. 9900448 by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University, Brooklyn, NY 11201, USA.

shown in Maciel and Felsen[1], for narrow-waisted (NW) GBs with $L_x \leq \lambda_0 \ll d$, λ_0 being the free-space wavelength, the $n \neq 0$ tilted GBs are *evanescent* (i.e, non-contributing) sufficiently far from the $z = z_A$ plane. Coupled with high frequency asymptotics, this property allows the conventional computationally intensive determination of the exact excitation amplitudes due to a Gabor-discretized large aperture distribution to be avoided and replaced with adequate accuracy by sampling the aperture profile at the x_m lattice point locations (Maciel and Felsen[1]). Field synthesis at the observer is accomplished by adding contributions from the ray-like $n = 0$ nontilted NW-CSP-GBs. The incident field is thus written as (Galdi et al.[3])

$$E^{inc}(x, z) \approx \sum_{|m| \leq L_x/2d} A_m B_m(x, z), \quad A_m = \left(L_x/\sqrt{2}\right)^{1/2} f(x_m), \quad (1)$$

where the CSP beam propagator B_m is given by

$$B_m(x, z) \sim ik_0 2^{5/4} \left(\frac{L_x}{8\pi k_0 \tilde{R}_m} \right)^{1/2} \exp \left\{ i \left[k_0 (\tilde{R}_m + ib) + \pi/4 \right] \right\} \frac{(z + ib)}{\tilde{R}_m}, \quad (2)$$

with $k_0 = 2\pi/\lambda_0$ being the free-space wavenumber, $\tilde{R}_m = \left[(x - x_m)^2 + (z + ib)^2 \right]^{1/2}$ (with $\Re(\tilde{R}_m) \geq 0$) denoting the *complex distance* between the observation point (x, z) and the CSP $(x_m, -ib)$, and $b = L_x^2/\lambda_0$ representing the Fresnel length of the beam. Here and henceforth, the tilde \sim identifies CSP complex quantities.

Reflected and Transmitted Fields. In the presence of the rough dielectric interface environment, the NW-GBs launched from the aperture plane can be tracked like *quasi-real* rays, as shown in Maciel and Felsen[1]. We begin with the canonical problem of CSP Gaussian beam reflection from a curved segment of a dielectric boundary, as depicted in Fig. 2, which furnishes the building-blocks for the rough surface scattering algorithm (whose complete description can be found in Galdi et al.[3]). Referring to Fig. 2, an incident Gaussian beam is generated by a CSP at $\tilde{P}_s \equiv (x_s + ib \cos \alpha_s, z_s + ib \sin \alpha_s)$, α_s being the real departure angle of the beam axis relative to the x -axis. For electrically large and smooth scatterers, and when the observation point $P \equiv (x, z)$ lies in the paraxial region of a reflected beam, the reflected field can be approximated in terms of the on-axis field of that beam (i.e., a *real* ray) and a complex phase correction (Maciel and Felsen[1]). Denoting the on-axis parameters by the subscript zero, one finds (Galdi et al.[3])

$$E^{refl}(P) \approx E^{refl}(P_0) \Big|_{\tilde{P}_i = P_{i0}} \exp(ik_0 \tilde{\delta}_p) = \Gamma E^{inc}(P_{i0}) \left(\frac{\tilde{f}_{r0}}{\tilde{R}_0} \right)^{1/2} \exp \left[ik_0 (L_{r0} + \tilde{\delta}_p) \right]. \quad (3)$$

Here (see Fig. 2): $\tilde{R}_0 = \overline{F_{r0} P_0}$; $\tilde{R} = \overline{F_{r0} P}$; $\tilde{L}_{i0} = \overline{\tilde{P}_s P_{i0}}$; $L_{r0} = \overline{P_{i0} P_0}$; $\tilde{\delta}_p = \tilde{R} - \tilde{R}_0$ is the complex phase correction; $\tilde{F}_{r0} = (x_{i0} - \tilde{f}_{r0} \cos \alpha_r, z_{i0} - \tilde{f}_{r0} \sin \alpha_r)$ is the *complex* (virtual) focus obtained by analytic continuation of the standard ray-optical formulas, with $\tilde{f}_{r0} = \tilde{L}_{i0} r_c \cos \theta_i / (2\tilde{L}_{i0} + r_c \cos \theta_i)$; the complex incidence point \tilde{P}_i is approximated by the real beam-axis incidence point $P_{i0} \equiv (x_{i0}, z_{i0})$; α_r is the real departure angle of the reflected beam axis relative to the x -axis; r_c is the curvature radius at P_{i0} ; and Γ is the relevant TM plane-wave Fresnel reflection coefficient. As shown in (Maciel and Felsen[1]), this corresponds to tracing a ray along a *complex* trajectory from the CSP to the intersection of the *real* beam axis with the real surface; from there the path to the observer proceeds entirely in real configuration space, along the beam axis. Transmitted fields and possible multiple interactions can be modeled in a similar fashion (see Galdi et al.[3] for details).

RESULTS AND CONCLUSIONS

From an extensive set of reflected/transmitted field examples for different choices of problem parameters, calibrating each through comparison with an independent numerical reference solution (Galdi et al.[3], Leviatan and Boag[4]), we have chosen the lossy interface profile $h(x)$ in Fig. 3, excited by a nonphased cosine aperture field distribution ($f(x) = \cos(\pi x/d)$) of width $d = 80\lambda_0$, located at $z_A = 6.4\lambda_0$ on the height scale of Fig. 3, i.e., very close to the maximum profile height of $6\lambda_0$. The Fresnel zone reflected field observed at $z_{or} = 20\lambda_0$ is shown in Fig. 4. Here, $L_x = d/80 = \lambda_0$. On the plot, the beam-generated solution (dashed) agrees very well with the reference solution (solid) in this typical example. In our numerical experiments, we have found accurate results for moderate roughness with maximum slopes $\leq 40^\circ$ and (average) curvature radii larger than a wavelength, and for incidence directions far from grazing, with typical computing times of about 9 ms per point on a 500 MHz laptop. These results suggest that our algorithm is promising as a fast forward solver in complex wave scattering scenarios.

REFERENCES

- [1] Maciel, J.J., and Felsen, L.B., *IEEE Trans. Antennas Propagat.*, **vol. 37**, no. 7, July 1989, pp. 884-892; *ibid.*, **vol. 38**, no. 10, Oct. 1990, pp. 1607-1617; *ibid.*, **vol. 38**, no. 10, Oct. 1990, pp. 1618-1624.
- [2] Galdi, V., Castañón, D.A., and Felsen, L.B., "Narrow-waisted Gaussian beams for reconstruction of moderately rough one-dimensional dielectric interfaces," this Conference. Also submitted to *IEEE Trans. Geosci. Remote Sensing*, Sept. 2000.
- [3] Galdi, V., Felsen, L.B., and Castañón, D.A., "Quasi-ray Gaussian beam algorithm for time-harmonic two-dimensional scattering by moderately rough interfaces," *IEEE Trans. Antennas Propagat.*, **vol. 49**, no. 9, Sept. 2001 (in print).
- [4] Leviatan, Y., and Boag, A., *IEEE Trans. Antennas Propagat.*, **vol. 35**, no. 10, Oct. 1987, pp. 1119-1127.

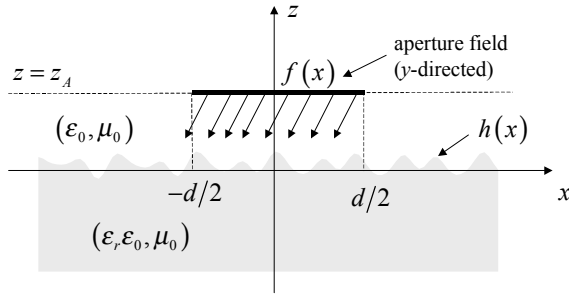


Figure 1 – Problem geometry.

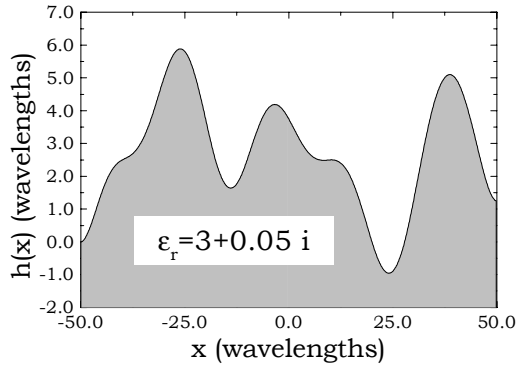


Figure 3 – Rough surface geometry and parameters ϵ_r is the relative (to ϵ_0) permittivity.

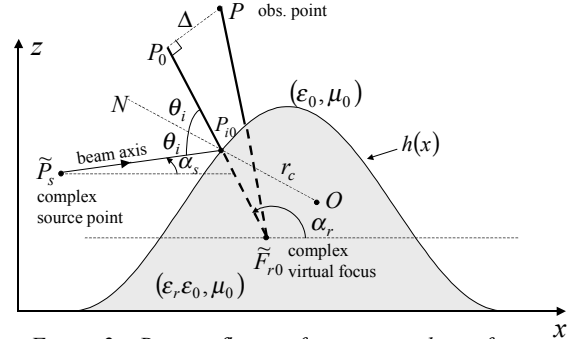


Figure 2 – Beam reflection from a curved interface. ϵ and μ are the permittivity and permeability, respectively.

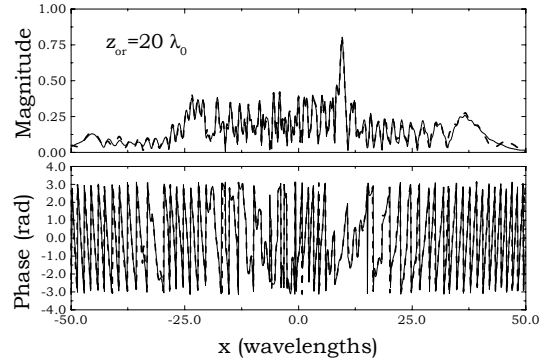


Figure 4 – Beam-computed (dashed) and reference (solid) reflected fields.

TIME-DEPENDENT TWO-DIMENSIONAL RADIATION FROM LARGE ONE-DIMENSIONAL APERTURES VIA NARROW-WAISTED GABOR-BASED GAUSSIAN BEAM WAVE PACKETS*

Vincenzo Galdi^(1,2), Leopold B. Felsen^(1,3), and David A. Castañón⁽¹⁾

⁽¹⁾ Dept. of Electrical & Computer Eng., Boston University, 8 Saint Mary's Street, Boston, MA 02215, USA

⁽²⁾ Waves Group, University of Sannio, I-82100, Benevento, Italy

⁽³⁾ Dept. of Aerospace & Mechanical Eng., Boston University, 110 Cummington St., Boston, MA 02215, USA
Also University Professor Emeritus, Polytechnic University, Brooklyn, NY 11201 USA

Abstract: In this paper, we present a computationally efficient and physically appealing discretized Gabor-based algorithm constructed with *narrow-waisted* ray-like pulsed Gaussian beams, for the two-dimensional (2D) electromagnetic fields radiated from short-pulsed large 1D truncated aperture distributions. This analysis extends the time-harmonic results in Maciel and Felsen[1] to the time domain.

INTRODUCTION AND PROBLEM FORMULATION

Narrow-waisted, ray-like, Gabor-based Gaussian beam (GB) algorithms have been shown capable of providing robust and efficient syntheses for radiation from finite one-dimensional (1D) aperture distributions (Maciel and Felsen[1]) and the interaction of these 2D fields with complex 2D environments (Maciel and Felsen[2], Galdi et al.[3]). The investigations have been restricted to time-harmonic excitation, whereby the complex source point (CSP) method can be exploited effectively for tracking each basis beam. Here, we present the extension of the frequency domain (FD) results for aperture radiation in Maciel and Felsen[1] to short-pulse time-dependent (TD) excitation. Referring to the problem geometry in Fig. 1, we consider a 1D (y -directed) space-time aperture field distribution of width d at $z = 0$, with separable space-time dependence and linear time delay

$$e_y(x, z=0, t) = f(x, t), |x| \leq d/2, \quad f(x, t) = g(x) p(t - c^{-1}x \sin \theta_A), \quad (1)$$

where c is the speed of the light and $p(t)$ is a pulse with characteristic width $T_p \ll d/c$. This distribution gives rise to a space-time pulse propagating in the θ_A direction (see Fig. 1). The FD algorithm in Maciel and Felsen[1] evolves from the rigorous Kirchhoff integration over the aperture distribution, which is then parameterized via the discrete Gabor basis and evaluated asymptotically for high frequencies to furnish the GB propagators to the observer. The TD results here are obtained by Fourier inversion from the FD and yield pulsed beams (PB).

NARROW-WAISTED GAUSSIAN BEAM DISCRETIZATION

Frequency Domain. In the 2D domain (x, z) , the high-frequency narrow-waisted (NW) basis beams are excited by Gaussian initial field profiles which, in the 1D aperture plane $z = 0$, are centered on the Gabor lattice points $x_m = mL_x$, $k_{xm} = n\beta_x$, with $(m, n) = 0, \pm 1, \pm 2, \dots$, in the (x, k_x) phase space (k_x is the x -domain spectral wavenumber). The spatial and spectral periods L_x and β_x , respectively, are constrained by $L_x\beta_x = 2\pi$. For NW beams with $L_x \ll d$, the spectral displacement (tilt) $\beta_x = 2\pi/L_x \gg 1$, thereby making the $n \neq 0$ basis beams evanescent and therefore negligible “sufficiently far” from the aperture plane (Maciel and Felsen[1]). By high frequency asymptotics, the NW-GB propagators can be modeled via the complex-source-point (CSP)

* Work supported by ODDR&E under MURI grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from Grant No. 9900448 by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University, Brooklyn, NY 11201, USA.

method, with their excitation amplitudes determined approximately from sampling of the aperture profile at the spatial lattice points x_m . Accordingly, for time-harmonic $\exp(-i\omega t)$ excitation, the radiated field can be represented by beam superposition as (Galdi et al.[4]),

$$E_y(x, z, \omega) \approx \sum_{|m| \leq d/2L_x} C_m B_m(x, z, \omega), \quad C_m = \left(L_x / \sqrt{2}\right)^{1/2} g(x_m), \quad (2)$$

$$B_m(x, z, \omega) \sim -ik 2^{5/4} \left(\frac{L_x}{8\pi}\right)^{1/2} \exp\left\{i\left[k\left(\tilde{R}_m + x_m \sin\theta_A + ib\right) + \pi/4\right]\right\} \frac{(z - ib \cos\theta_A)}{\tilde{R}_m^{3/2}}. \quad (3)$$

Here, $k = \omega/c = 2\pi/\lambda$ is the ambient wavenumber with λ as the wavelength, $\tilde{P}_m \equiv (x_m + ib \sin\theta_A, ib \cos\theta_A)$ is the CSP, the complex displacement magnitude $b = (L_x \cos\theta_A)^2 / \lambda$ is the Fresnel length of the beam, and $\tilde{R}_m = \left[(x - x_m - ib \sin\theta_A)^2 + (z - ib \cos\theta_A)^2\right]^{1/2}$ (with $\Re(\tilde{R}_m) \geq 0$), is a *complex distance*; through $L_x = 2\pi/\beta_x$, C_m is λ -dependent.

Time Domain. When Fourier-inverting to the short-pulse TD to generate pulsed beams, we choose L_x to be λ -independent because this renders the aperture-sampled TD-Gabor PB-excitation coefficients λ -independent. Specifically, via the analytic signal formulation (one-sided Fourier transform), one has (Galdi et al.[4]),

$$e_y(x, z, t) \approx \sum_{|m| \leq d/2L_x} c_m b_m(x, z, t), \quad c_m = \left(L_x / \sqrt{2}\right)^{1/2} g(x_m), \quad (4)$$

$$b_m(x, z, t) = \Re e \left[\frac{1}{\pi} \int_0^\infty B_m(x, z, \omega) P(\omega) \exp(-i\omega t) d\omega \right], \quad (5)$$

where $P(\omega)$ represents the spectrum of the pulse $p(t)$. We then find that, subject to certain constraints which are validated subsequently by numerical experimentations, the approximate inversion integrals in (5) for Gaussian-envelope-modulated pulses can be evaluated in closed form in terms of efficiently computable hypergeometric functions (Galdi et al.[4]). The constraints are formalized through nondimensional critical estimators which contain all relevant problem parameters, and show how changes in one parameter can be compensated by corresponding changes in the other parameters so as to remain in the legitimized range. The results can be generalized to arbitrarily phased (e.g., focused) aperture distributions (Galdi et al.[4]).

RESULTS AND CONCLUSIONS

For illustration, we consider a non-phased (i.e., $\theta_A = 0$) cosine-tapered distribution with a fourth-order Rayleigh pulse excitation,

$$g(x) = \cos(\pi x/d), \quad p(t) = \left(\frac{T_p^4}{30000}\right) \frac{d^4}{dt^4} \exp\left[-\frac{50(t - T_p/2)}{T_p^2}\right]. \quad (6)$$

Figure 2 shows typical temporal and spatial profiles, computed via the PB algorithm (dashed curves) and a reference solution obtained by rigorous space-time Kirchhoff aperture numerical integration (solid curves). The two results coincide on the scale of the drawings. The same type of agreement is observed for more complex (phased, focused) aperture field distributions (Galdi et al.[4]). These results confirm that the previously established utility of the Gabor-based narrow-waisted FD-GB algorithm for radiation from distributed apertures in Maciel and Felsen[1] remains intact in the TD. In the FD, it has been established that the GB basis beams can be tracked efficiently from the aperture through interactions with complex scattering environments (Maciel and

Felsen[2]), and are therefore useful forward solvers in inversion scenarios (Galdi et al.[5]). Interaction of the PB propagators with complex environments is presently under consideration (Galdi et al.[6]).

REFERENCES

- [1] Maciel, J.J., and Felsen, L.B., *IEEE Trans. Antennas Propagat.*, vol. 37, no. 7, July 1989, pp. 884-892.
- [2] Maciel, J.J., and Felsen, L.B., *IEEE Trans. Antennas Propagat.*, vol. 38, no. 10, Oct. 1990, pp. 1607-1617; *ibid.*, vol. 38, no. 10, Oct. 1990, pp. 1618-1624.
- [3] Galdi, V., Felsen, L.B., and Castañón, D.A., "Scattering by moderately rough dielectric interfaces via narrow waisted Gaussian beams," this Conference. Also, *IEEE Trans. Antennas Propagat.*, vol. 49, no. 9, Sept. 2001 (in print).
- [4] Galdi, V., Felsen, L.B., and Castañón, D.A., "Narrow-waisted Gaussian beam discretization for short-pulse radiation from one-dimensional large apertures," *IEEE Trans. Antennas and Propagat.*, vol. 49, no. 10, Oct. 2001 (in print).
- [5] Galdi, V., Castañón, D.A., and Felsen, L.B., "Narrow-waisted Gaussian beams for reconstruction of moderately rough one-dimensional dielectric interfaces," this Conference. Also submitted to *IEEE Trans. Geosci. Remote Sensing*, Sept. 2000.
- [6] Galdi, V., Felsen, L.B., and Castañón, D.A., "Time-domain 2D scattering by moderately rough dielectric interfaces via narrow-waisted Gaussian beams," to be presented at 2001 IEEE Antennas and Propagat. Int. Symposium, Boston, MA, USA, July 8-13, 2001.

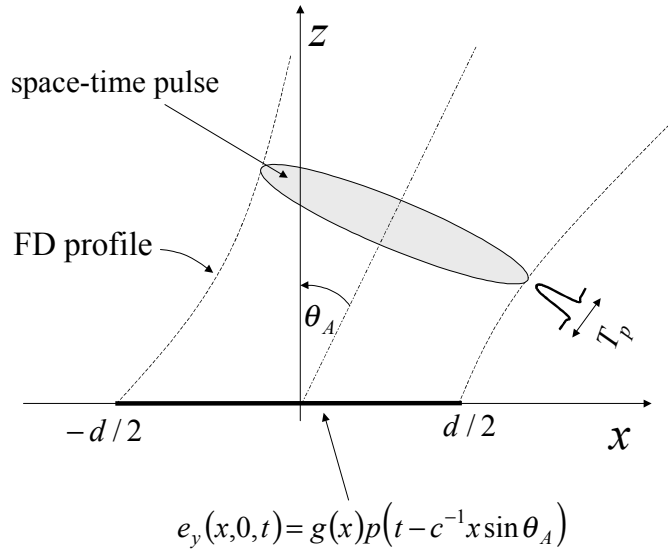


Figure 1

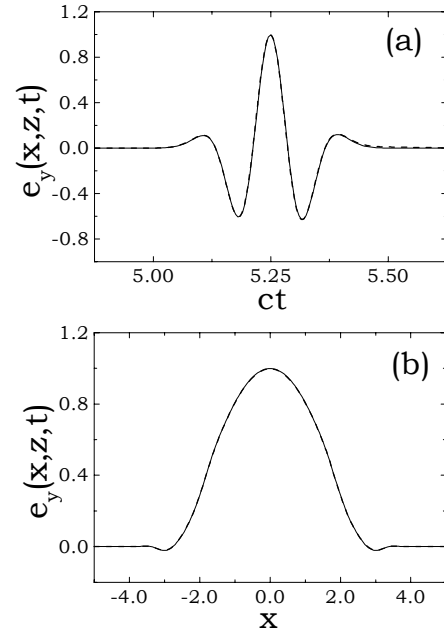


Figure 2

Figure 1 – Problem geometry and coordinate system.

Figure 2 – Beam-computed (dashed) and reference (solid) radiated field due to cosine-tapered focused aperture distribution in with $ct_p = 0.5$, $d = 5$ (arbitrary units). (a): Temporal profiles at $x = 0$ and $z = 5$ (30 beams);

(b): Corresponding spatial profile at $ct = 5.25$. The adequacy of the number of beams retained in the algorithm is determined by the insensitivity of the outcome to variations (scramblings) of the beam/lattice parameters (cf. Maciel and Felsen[1], [2]).

TWO-DIMENSIONAL NARROW-WAISTED GAUSSIAN BEAMS FOR RECONSTRUCTION OF MODERATELY ROUGH ONE-DIMENSIONAL DIELECTRIC INTERFACES *

Vincenzo Galdi^(1,2), David A. Castañón⁽¹⁾, and Leopold B. Felsen^(1,3)

⁽¹⁾ Dept. of Electrical & Computer Eng., Boston University, 8 Saint Mary's Street, Boston, MA 02215, USA

⁽²⁾ Waves Group, University of Sannio, I-82100, Benevento, Italy

⁽³⁾ Dept. of Aerospace & Mechanical Eng., Boston University, 110 Cummington St., Boston, MA 02215, USA
Also University Professor Emeritus, Polytechnic University, Brooklyn, NY 11201 USA

Abstract: In this paper, we present a new technique for determining the surface profile of a one-dimensional moderately rough interface between air and a homogeneous dielectric half space. Based on sparsely sampled step-frequency ground penetrating radar measurements, the proposed inversion scheme uses a quasi-ray Gaussian beam fast forward model, coupled with a low-order parameterization of the surface profile in terms of B-splines. The profile estimation problem is posed as a parameter optimization problem, which is solved using a multiresolution continuation method. Numerical experiments establish that the algorithm is efficient, and computes very accurate profiles throughout most of the illuminated region even in noisy environments, losing accuracy only in regions with very weak illumination.

INTRODUCTION AND PROBLEM FORMULATION

Estimation of rough surfaces from inverse scattering has received considerable attention in the past decade. However, most of the available algorithms have focused on conducting surfaces. Wombell and DeSanto[1] used Kirchhoff approximations and Fourier transforms to estimate surface profiles based on measurements of the scattered field in all spectral directions. Ying and Noguchi[2] used nonlinear optimization techniques for direct estimation of surfaces illuminated by monochromatic Gaussian beams, based on the far-field scattering amplitude for all spectral directions. In a different approach, Schatzberg and Devaney[3] used Rytov approximation and backpropagation to estimate surface profiles based on full measurements of the scattered wave. In contrast with the above contributions, our work in this paper is focused on estimating surface profiles based on reflection from a moderately rough interface between air and a homogeneous dielectric half space (soil). The problem geometry is illustrated in Fig. 1: in a y -independent (x, z) coordinate system, a y -directed field generated by an aperture field distribution at $z = z_A$, $E^{inc}(x, z_A) = f(x)$, $|x| \leq d/2$, is assumed to illuminate a homogeneous dielectric half-space, with known relative permittivity ϵ_r , bounded by an irregular interface described by the continuous function $z = h(x)$. We assume that the reflected field is measured at N_r receivers with discrete spatial locations (x_1, \dots, x_{N_r}) at $z = z_{obs}$, using a stepped-frequency ground penetrating radar (GPR) with N_λ operating wavelengths (frequencies).

ALGORITHMIC ASPECTS

Surface Parameterization. The proposed inversion strategy is based on a low-dimensional compact geometrical parameterization of the unknown interface profile, whose unknown parameters are estimated by fitting the model-based forward scattering prediction to the available (measured/simulated) data, i.e., minimizing a suitable cost functional. The choice of the interface profile parameterization is a key issue, and requires tradeoff between

* Work supported by ODDR&E under MURI grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from Grant No. 9900448 by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University, Brooklyn, NY 11201, USA.

compactness and versatility. We assume that the surface shape can be approximated by a finite set of quartic B-spline basis functions with unknown coefficients,

$$h(x) \approx \sum_{n=1}^N c_n s_n^{(4)}(x), \quad (1)$$

where $s_n^{(4)}$ indicates a quartic B-spline basis function.

Forward Scattering Model. We exploit a computationally efficient forward model relating the reflected field at the receivers to the surface profile. This model, detailed in Galdi et al.[4], utilizes Gabor-based Gaussian beam algorithms in conjunction with the complex source point (CSP) method for generating beam-like wave objects (Maciel and Felsen[5]). The algorithm has previously been validated and calibrated for the range of parameters involved in this inverse scattering scenario.

Surface Estimation. The surface estimation problem is posed as a nonlinear optimization problem, similar in spirit to the work of Ying and Noguchi[2]. Let $\mathcal{E}_{p,q}^r$ denote the y -directed reflected field measured at wavelength λ_q at sensor position x_p . Given a vector of spline coefficients $\underline{c} = \{c_1, \dots, c_N\}$ and the aperture-excited incident field, we can use the above beam algorithm for the surface profile in (1) to generate predictions of the reflected field waveforms at each receiver, i.e., $E_{p,q}^r \equiv E^r(x_p, z_{obs}; \lambda_q; \underline{c})$. Accordingly, we define the error functional as follows

$$\Phi(\underline{c}) \equiv \left\| \underline{E}^r(\underline{c}) - \underline{\mathcal{E}}^r \right\|^2, \quad (2)$$

where $\underline{E}^r(\underline{c}) = \{E_{p,q}^r\}$, $\underline{\mathcal{E}}^r = \{\mathcal{E}_{p,q}^r\}$, $p = 1, \dots, N_r$, $q = 1, \dots, N_\lambda$. The measured data are simulated via the multifilament current method (Galdi et al.[4], Leviatan and Boag[6]). The regularized inverse scattering is formalized as finding the coefficient vector $\underline{c} = \{c_1, \dots, c_N\}$ in (1) which minimizes the cost functional in (2), i.e. finding $\hat{\underline{c}}$ such that

$$\hat{\underline{c}} = \arg \min_{\underline{c}} \Phi(\underline{c}). \quad (3)$$

We show that this optimization problem has local minima, and develop a multiresolution continuation strategy based on selective use of temporal frequency information to approach convergence to globally optimal estimates of surface profiles (see Galdi et al.[4] for details).

RESULTS AND CONCLUSIONS

From an extensive set of numerical experiments for a broad range of problem parameters, we have chosen the typical reconstruction example in Fig. 2. The surface and soil parameters are specified in the figure caption. The incident field is generated by a cosine-tapered aperture field distribution (i.e., $f(x) = \cos(\pi x/d)$) of width $d = 0.8$ located at $z_A = 0.6$. The complex reflected field is measured at 10 equispaced observation points at $z_{obs} = 0.6$ spanning the illuminated region, and at four operating wavelengths ($\lambda = 0.2, 0.1, 0.067, 0.05$). The length scales are normalized with respect to the smallest radius of curvature of the local interface profile maxima. The multiresolution optimization strategy detailed in Galdi et al.[4] is exploited to minimize the obtained cost functional.

It can be observed that, with the exception of the poorly illuminated edge regions, the reconstruction is reasonably accurate. This kind of accuracy was confirmed in other numerical experiments, even in noisy environments, with computing time of 1-2 minutes on a 400 MHz laptop (Galdi et al.[4]). We are currently working on extending the algorithm to short-pulse excitation (Pavlovich et al.[7]).

REFERENCES

- [1] Wombell, R.J., and DeSanto, J.A., *Inverse Probl.*, **vol. 7**, 1991, pp. L7-L12; *ibid.*, *J. Opt. Soc. Am. A*, **vol. 8**, no. 12, Dec. 1991, pp. 1892-1897.
- [2] Ying, C., and Noguchi, A., *IEICE Trans. Electron.*, **vol. E77-C**, no. 11, Nov. 1994, pp. 1781-1785.
- [3] Schatzberg, A., and Devaney, A.J., *J. Opt. Soc. Am. A*, **vol. 10**, no. 5, May 1993, pp. 942-950.
- [4] Galdi, V., Felsen, L.B., and Castañon, D.A., "Two-dimensional scattering by moderately rough dielectric interfaces via narrow waisted Gaussian beams," this Conference. Also, *IEEE Trans. Antennas Propagat.*, **vol. 49**, no. 9, Sept. 2001 (in print).
- [5] Maciel, J.J., and Felsen, L.B., *IEEE Trans. Antennas Propagat.*, **vol. 37**, no. 7, July 1989, pp. 884-892; *ibid.*, **vol. 38**, no. 10, Oct. 1990, pp. 1607-1617; *ibid.*, **vol. 38**, no. 10, Oct. 1990, pp. 1618-1624.
- [6] Leviatan, Y., and Boag, A., *IEEE Trans. Antennas Propagat.*, **vol. 35**, no. 10, Oct. 1987, pp. 1119-1127.
- [7] Pavlovich, J., Galdi, V., Karl, W.C., Castañon, D.A., and Felsen, L.B., "Time-domain reconstruction of moderately rough dielectric interfaces via quasi-ray Gaussian beams," to be presented at 2001 IEEE Antennas and Propagat. Int. Symposium, Boston, MA, USA, July 8-13, 2001.

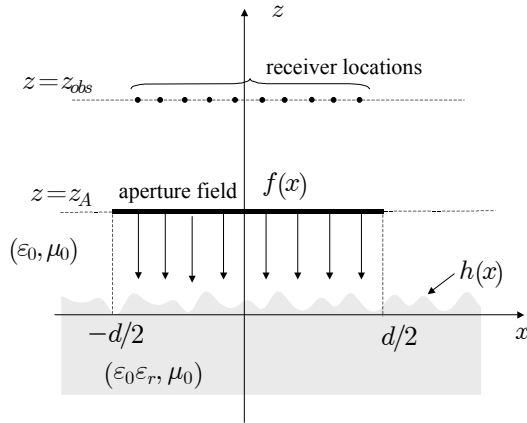


Figure 1 - Setup geometry for the inverse scattering problem. An aperture-excited, time-harmonic, TM-polarized EM field illuminates a homogeneous dielectric half-space with known relative permittivity ϵ_r and with a moderately rough interface whose coarse scale profile is described by the continuous function $h(x)$. The scattered field is sampled at N_r receiver locations at $z = z_{obs}$.

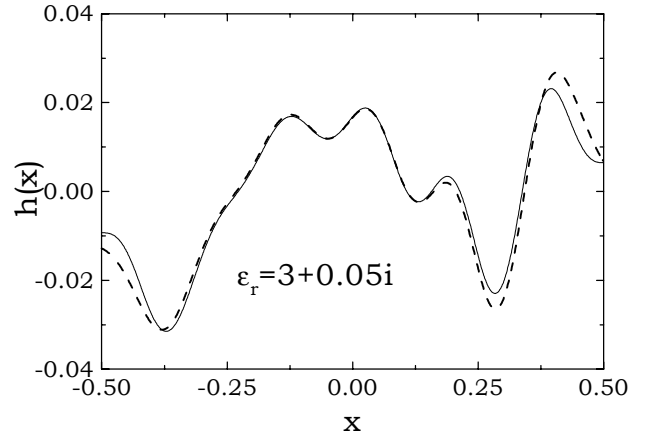


Figure 2 – Reconstruction example.

Profile parameters (arbitrary units): $\epsilon_r = 3 + i0.05$, maximum slope 34° ; aperture parameters (nonphased cosine distribution): $d = 0.8$, $z_A = 0.6$; 10 equispaced observation points at $z_{obs} = 0.6$; operating wavelengths: $\lambda = 0.2, 0.1, 0.067, 0.05$.

NARROW-WAISTED GAUSSIAN BEAM DISCRETIZATION FOR PULSED RADIATION FROM 2D LARGE APERTURES *

Vincenzo Galdi^(1,3), Leopold B. Felsen^(1,2), and David A. Castañon⁽¹⁾

⁽¹⁾ Department of Electrical and Computer Engineering
Boston University, Boston, MA 02215, USA

E-mail: vgaldi@bu.edu, lfelsen@bu.edu, dac@bu.edu

⁽²⁾ Department of Aerospace and Mechanical Engineering
Boston University, Boston, MA 02215, USA

Also, University Professor Emeritus,
Polytechnic University, Brooklyn, NY 11201, USA

⁽³⁾ Waves Group, University of Sannio, Benevento, Italy

I. INTRODUCTION AND PROBLEM FORMULATION

In this paper, we present a computationally efficient and physically appealing discretized Gabor-based parameterization, in terms of *narrow-waisted* ray-like pulsed Gaussian beams, for the three-dimensional (3D) vector electromagnetic fields radiated from short-pulsed arbitrarily polarized large 2D truncated aperture distributions. This analysis extends the time-harmonic results in [1] to the time domain (TD), following the guidelines of our previous investigation of pulsed radiation of 2D fields from 1D large truncated apertures [2].

Referring to the problem geometry in Fig. 1, a 2D space-time aperture field distribution

$$\mathbf{e}_0(x, y, z = 0, t) = \mathbf{f}(x, y, t), \quad (x, y) \in \Gamma_A, \quad \mathbf{u}_z \cdot \mathbf{f} = 0, \quad (1)$$

is assumed to occupy the aperture region Γ_A (of typical dimension d_A) in the $z = 0$ plane. Here and henceforth, bold face symbols denote vector quantities, \mathbf{u}_α denotes a unit vector in the α -direction, and the subscript “0” identifies quantities in the $z = 0$ plane. The resulting space-time field radiated into the half-space $z > 0$ can be rigorously synthesized by Kirchhoff integration [3, p. 175],

$$\mathbf{e}(\mathbf{r}, t) = -\frac{1}{2\pi} \iint_{-\infty}^{\infty} (\mathbf{r} - \mathbf{r}') \times \left\{ \mathbf{u}_z \times \left[\frac{\mathbf{f}(x, y, t - R/c)}{R^3} + \frac{1}{cR^2} \frac{\partial}{\partial t} \mathbf{f}(x, y, t - R/c) \right] \right\} dx' dy', \quad (2)$$

where $\mathbf{r} = x\mathbf{u}_x + y\mathbf{u}_y + z\mathbf{u}_z$, $\mathbf{r}'_0 = x'\mathbf{u}_x + y'\mathbf{u}_y$, $R = |\mathbf{r} - \mathbf{r}'_0|$, and c is the ambient propagation speed. In what follows, we shall be dealing with the simple case of space-time separable aperture fields having linear time delay,

*Work supported by ODDR&E under MURI grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from a Grant by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University.

$$\mathbf{f}(x, y, t) = \mathbf{g}(x, y)p[t - c^{-1} \sin \theta_A(x \cos \phi_A + y \sin \phi_A)], \quad (3)$$

with spectrum (capital letters identify frequency domain quantities),

$$\mathbf{F}(x, y, \omega) = P(\omega)\mathbf{g}(x, y) \exp[ik \sin \theta_A(x \cos \phi_A + y \sin \phi_A)], \quad (4)$$

where $p(t)$ is a short pulse of length $T \ll d_A/c$ with spectrum $P(\omega)$. Due to space limitations, we omit the technical details and concentrate on presenting the main results. The interested reader is referred to [4] for a comprehensive treatment.

II. FREQUENCY DOMAIN BEAM DISCRETIZATION

We first recall relevant results of our previous frequency domain (FD) investigation in [1], before proceeding to their inversion (via analytic signal transform) into the time domain. For an implied $\exp(-i\omega t)$ dependence, the FD counterpart of (2) for the radiated 3D field, using the linearly-phased aperture distribution in (4), is given by [3, p. 107]

$$\mathbf{E}(\mathbf{r}) = 2 \int_{-\infty}^{\infty} \nabla G_{3D}(\mathbf{r}, \mathbf{r}'_0) \times [\mathbf{u}_z \times \mathbf{F}(x', y')] dx' dy', \quad (5)$$

where $G_{3D}(\mathbf{r}, \mathbf{r}'_0) = \exp(ikR)/(4\pi R)$ is the 3D-FD Green's function with $k = \omega/c = 2\pi/\lambda$ as the ambient wavenumber, and λ as the wavelength. As shown in [1], [4], the radiated field in (5) can be parameterized effectively in terms of Gabor-based *narrow-waisted complex source point* (CSP) Gaussian beams

$$\tilde{\mathbf{E}}(\mathbf{r}) \sim \sum_{m,p} (\mathbf{C}_{mp} \times \mathbf{u}_z) \times \tilde{\mathcal{B}}_{mp}(\mathbf{r}), \quad \mathbf{C}_{mp} = (L/\sqrt{2})\mathbf{g}(mL, pL), \quad (6)$$

$$\tilde{\mathcal{B}}_{mp}(\mathbf{r}) = ikL \frac{\exp[ik(\tilde{R}_{mp} + ib + \mathcal{L}_{mp})]}{\sqrt{2\pi}\tilde{R}_{mp}^2} (\mathbf{r} - \tilde{\mathbf{r}}'_{mp}), \quad \tilde{R}_{mp} = |\mathbf{r} - \tilde{\mathbf{r}}'_{mp}|, \quad (7)$$

$$\tilde{\mathbf{r}}'_{mp} = (mL + ib \sin \theta_A \cos \phi_A)\mathbf{u}_x + (pL + ib \sin \theta_A \sin \phi_A)\mathbf{u}_y + ib \cos \theta_A \mathbf{u}_z, \quad (8)$$

$$\mathcal{L}_{mp} = L \sin \theta_A (m \cos \phi_A + p \sin \phi_A), \quad b = (L \cos \theta_A)^2/\lambda, \quad (9)$$

where $\tilde{\mathbf{r}}'_{mp}$ is the CSP at the (m, p) -indexed (x, y) plane lattice point, \mathbf{C}_{mp} are Gabor expansion coefficients (obtained by aperture profile sampling), and $\tilde{\mathcal{B}}_{mp}$ is the *ray-like* GB propagator in the (θ_A, ϕ_A) direction, obtained from the 3D-FD Green's function by analytic continuation of the spatial coordinates into complex space (see (8)). The expansion in (6) is a high frequency asymptotic ray-like approximation of the rigorous, Gabor-based, Gaussian beam (GB) discretization presented in [1], which is tied to a discretized lattice in a four-dimensional $(x, y; k_x, k_y)$ space-spectrum phase space, where k_x, k_y are the spectral wavenumbers. Equation (6) has been found to yield accurate syntheses for sufficiently small values of the spatial lattice period $L \lesssim \lambda \ll d_A$, i.e., for narrow-waisted beams. In (6)–(8), the tilde \sim identifies dependence on analytically continued CSP coordinates as well as the field produced thereby. Due to the finite extent of the aperture, i.e., $\mathbf{g}(x, y) = 0$ for $(x, y) \notin \Gamma_A$, the summation in (6) is *self-truncating* up to a number of beams given roughly by $(d_A/L)^2$.

III. TIME DOMAIN BEAM DISCRETIZATION

The corresponding pulsed beam (PB) expansion for the field radiated by the space-time aperture distribution in (3) can be formally obtained by Fourier inversion of (6) filtered by the pulse spectrum $P(\omega)$. Although this cannot be accomplished analytically for the most general pulse shape, we have found rapidly computable closed form approximations for the important class of Rayleigh pulses

$$p(t) = \Xi p_g^{(j)}(t - T/2, T), \quad P(\omega) = \Xi(-i\omega)^j \exp(i\omega T/2) P_g(\omega, T), \quad (10)$$

where the superscript (j) indicates j -th order differentiation, Ξ is an amplitude coefficient, and

$$p_g(t, T) = \exp\left[-\left(\frac{t}{\sigma T}\right)^2\right], \quad P_g(\omega, T) = \sqrt{\pi}\sigma T \exp\left(-\frac{\sigma^2 T^2 \omega^2}{4}\right). \quad (11)$$

Following a procedure analogous to that presented in [2] (see [4] for details), one obtains a PB expansion formally similar to (6),

$$\mathbf{e}(\mathbf{r}, t) \sim \sum_{m,p} (\mathbf{c}_{mp} \times \mathbf{u}_z) \times \mathbf{b}_{mp}(\mathbf{r}, t), \quad \mathbf{c}_{mp} = \mathbf{C}_{mp} = (L/\sqrt{2})\mathbf{g}(mL, pL). \quad (12)$$

The PB propagator \mathbf{b}_{mp} is given in the beam-centered coordinates \mathbf{r}_{bmp} by [4]

$$\mathbf{b}_{mp}(\mathbf{r}, t) = \frac{\Xi \Lambda_{mp} \sigma T}{\mathcal{T}_{mp}} p_g^{(j+1)}[(t - T/2 - \tau_{mp})\sigma, \mathcal{T}_{mp}] (\mathbf{r} - \mathbf{r}'_{mp}), \quad (13)$$

$$\begin{bmatrix} x_{bmp} \\ y_{bmp} \\ z_{bmp} \end{bmatrix} = \begin{bmatrix} \cos \theta_A \cos \phi_A & \cos \theta_A \sin \phi_A & -\sin \theta_A \\ -\sin \phi_A & \cos \phi_A & 0 \\ \sin \theta_A \cos \phi_A & \sin \theta_A \sin \phi_A & \cos \theta_A \end{bmatrix} \begin{bmatrix} x - mL \\ y - pL \\ z \end{bmatrix}, \quad (14)$$

$$\Lambda_{mp} = -\frac{L}{\sqrt{2}\pi c R_{mp}^2}, \quad \tau_{mp} = c^{-1} \left(z_{bmp} + \frac{\rho_{bmp}^2}{2z_{bmp}} + \mathcal{L}_{mp} \right), \quad (15)$$

$$T_{mp} = \frac{\rho_{bmp} L \cos \theta_A}{\sqrt{\pi} c z_{bmp}}, \quad \mathcal{T}_{mp} = \sqrt{T_{mp}^2 + \sigma^2 T^2}, \quad \rho_{bmp} = (x_{bmp}^2 + y_{bmp}^2)^{1/2}, \quad (16)$$

with $\mathbf{r}'_{mp} = mL\mathbf{u}_x + pL\mathbf{u}_y$, $R_{mp} = |\mathbf{r} - \mathbf{r}'_{mp}|$. The PB expansion in (12) involves rapidly computable analytic functions, and therefore can be evaluated very efficiently.

IV. RESULTS AND CONCLUSIONS

The PB synthesis in (12) has been validated and calibrated against a reference solution generated by Gaussian quadrature numerical integration of (2). Convergence issues have been addressed analytically in [4] and formalized in terms of nondimensional estimators. Typical results from a broad range of numerical experiments are shown in Fig. 2, for an x -polarized, square, cosine-tapered aperture field distribution

$$\mathbf{g}(x, y) = \begin{cases} \cos(\pi x/d) \cos(\pi y/d) \mathbf{u}_x, & |x|, |y| \leq d/2, \\ 0, & |x|, |y| > d/2, \end{cases} \quad (17)$$

and a fourth-order Rayleigh pulse ($j = 4$, $\Xi = T^4/30000$, $\sigma = 1/\sqrt{50}$). Figure 2 compares the reference solution and the PB synthesis (20×20 beams) for the x -component of radiated field at a fixed observation point in the near zone of the aperture. Excellent agreement is observed with the indicated number of beams, whose use is far more efficient than the Kirchhoff numerical integration. Similar results are obtained for the z -component, and also for more complicated phasing (as in rectangular focused aperture distributions) [4]. This renders the proposed pulsed beam synthesis attractive as a numerically efficient, accurate, robust algorithm for future scattering scenarios.

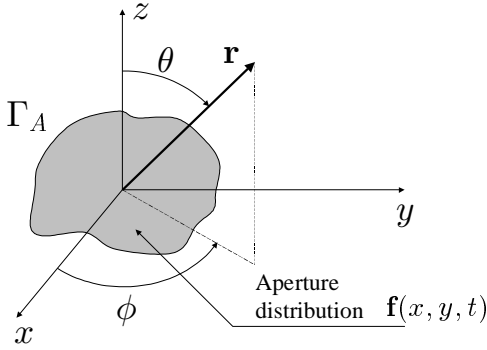


Fig. 1

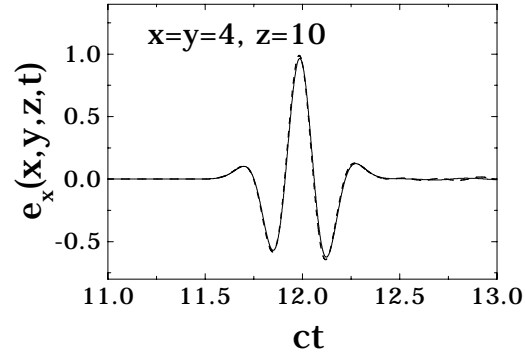


Fig. 2

Fig. 1 - Two-dimensional aperture field distribution and coordinate systems.

Fig. 2 - x -polarized, pulsed, linear-delay, square cosine aperture distribution in (3),(17), with $cT = 1$, $d = 10 = 10cT$ (arbitrary units), $\theta_A = 30^\circ$, $\phi_A = 45^\circ$. Temporal behavior of the x -component of radiated field observed at $(x = y = 4, z = 10)$. — Reference solution; - - - Pulsed beam synthesis ($L = d/20$, i.e., 400 beams).

REFERENCES

- [1] J.J. Maciel and L.B. Felsen, "Discretized Gabor-based beam algorithm for time-harmonic radiation from two-dimensional truncated planar aperture distributions - I: Formulation and solution"; *ibid*, "II: Asymptotics and numerical tests," submitted to *IEEE Trans. Antennas Propagat.*, Dec. 2000.
- [2] V. Galdi, L.B. Felsen, and D.A. Castañon, "Narrow-waisted Gaussian beam discretization for two-dimensional time-dependent radiation from large one-dimensional plane apertures," submitted to *IEEE Trans. Antennas Propagat.*, Sept. 2000.
- [3] T.B. Hansen and A.D. Yaghjian, *Plane-Wave Theory of Time-Domain Fields: Near-Field Scanning Applications*, Piscataway (NJ): IEEE Press, 1999.
- [4] V. Galdi, L.B. Felsen, and D.A. Castañon, "Time-domain radiation from large two-dimensional apertures via narrow-waisted Gaussian beams," submitted to *IEEE Trans. Antennas Propagat.*, Dec. 2000.

TIME-DOMAIN TWO-DIMENSIONAL SCATTERING BY MODERATELY ROUGH DIELECTRIC INTERFACES VIA NARROW-WAISTED GAUSSIAN BEAMS *

Vincenzo Galdi^(1,3), Leopold B. Felsen^(1,2), and David A. Castañon⁽¹⁾

⁽¹⁾ Department of Electrical and Computer Engineering

Boston University, Boston, MA 02215, USA

E-mail: vgaldi@bu.edu, lfelsen@bu.edu, dac@bu.edu

⁽²⁾ Department of Aerospace and Mechanical Engineering

Boston University, Boston, MA 02215, USA

Also, University Professor Emeritus,

Polytechnic University, Brooklyn, NY 11201, USA

⁽³⁾ Waves Group, University of Sannio, Benevento, Italy

I. INTRODUCTION AND PROBLEM FORMULATION

This paper treats the interaction of short-pulse electromagnetic radiation with one-dimensional (1D) moderately rough dielectric interfaces, which arise in many realistic problems, such as buried object imaging. Our ultimate goal is to incorporate physics-based, computationally efficient, forward models in such inverse scattering scenarios, in order to improve the overall computational performance. In this connection, we recently analyzed time-harmonic forward and inverse scattering from moderately rough surfaces, using a Gabor-based Gaussian beam algorithm [1], [2]. Here, we extend this forward scattering algorithm to pulsed excitation, via a Physical-Optics(PO)-based approach in conjunction with the pulsed beam discretization in [3]. The problem geometry is shown in Fig. 1: a TM-polarized incident field generated by a pulsed aperture field distribution of width d at $z = z_A$ is assumed to impinge from free-space onto a lossless, nondispersive, homogeneous dielectric half-space of relative permittivity ϵ_r , bounded by a moderately rough interface described by the continuous function $h(x)$. We assume a tapered, separable, nonphased space-time field distribution

$$\mathbf{e}_y^{inc}(x, z = z_A, t) = \begin{cases} f(x)p(t - c^{-1}z_A), & |x| \leq d/2, \\ 0, & |x| > d/2, \end{cases} \quad (1)$$

where $p(t)$ is a pulse of length $T \ll d/c$, with c representing the speed of light, and $f(x)$ is the amplitude taper profile. The interface is assumed to be located in the collimation zone of the aperture so that the incident field can be approximated by

*Work supported by ODDR&E under MURI grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellowship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from a Grant by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University.

a pulsed truncated tapered plane wave propagating in the negative z -direction (see Fig. 1). Due to space limitations, we present only the main results for the scattered field, omitting technical details which can be found in [4].

II. FREQUENCY DOMAIN FORMULATION

In the proposed approach, the time domain (TD) is accessed by Fourier inversion from the frequency domain (FD) via the analytic transform (with $\omega \geq 0$). In the asymptotic high-frequency range, and for smooth roughness over all wavelength scales in the pulse spectrum, the scattered field can be approximated by integration of the PO “current” over the interface profile

$$E_y^s(x, z, \omega) \sim \int_{C_{PO}} P(\omega) J_{PO}(x') \mathbf{u}_n \cdot \nabla G_{2D}(x, z; x', h(x')) dl', \quad (2)$$

where C_{PO} extends over the illuminated portion of the surface, $P(\omega)$ is the pulse spectrum, \mathbf{u}_n is the outward normal unit vector, and dl' is the incremental arc-length. Furthermore, G_{2D} is the FD line-source Green’s function,

$$G_{2D}(x, z; x', z') = \frac{i}{4} H_0^{(1)}(\omega R/c), \quad R = [(x - x')^2 + (z - z')^2]^{1/2}, \quad (3)$$

and J_{PO} is the PO surface current density

$$J_{PO}(x) = 2\mathcal{R}(x) E_y^{inc}(x, h(x)), \quad (4)$$

with \mathcal{R} denoting the TM plane-wave Fresnel reflection coefficient, and E_y^{inc} representing the incident field.

III. TIME DOMAIN BEAM DISCRETIZATION

The PO radiation integral in (2) is not very different from the Kirchhoff aperture integrals analyzed in [3], except that the line integration in (2) is performed along a rough surface profile instead of on a projected equivalent planar aperture. In [3], the 1D space-time aperture field distribution is first parameterized in the FD in terms of x -domain discretized m -indexed Gabor basis functions with narrow width L , centered on the Gabor lattice points $x_m = mL$; these initial conditions generate *narrow-waisted*, quasi-ray, complex-source-point Gaussian beams. For Rayleigh pulses, the resulting TD analytic Fourier inversion integral can be approximated by rapidly computable closed form expressions. The same approach can be applied to discretizing PO integrals as in (2). Thus, for fourth-order Rayleigh pulses

$$p(t) = \Xi \frac{d^4}{dt^4} \exp \left[- \left(\frac{t - T/2}{\sigma T} \right)^2 \right], \quad (5)$$

one obtains the following pulsed beam (PB) expansion for the scattered field (see [3] for details)

$$e_y^s(x, z, t) \sim \sum_{|m| \leq (D/2L)} c_m b_m(x, z, t - t_m), \quad (6)$$

where the interval $[-D/2, D/2]$ spans the x -domain projection of the illuminated portion of the surface. The Gabor expansion coefficients c_m and the time delay

t_m are approximated by sampling the TD-PO surface current at the lattice points $x_m = mL$,

$$c_m = \left(\frac{L}{\sqrt{2}}\right)^{1/2} \mathcal{R}(\theta_{im}) f(x_m), \quad t_m = c^{-1}(z_A - z_m - x_m \sin \theta_{im}), \quad (7)$$

where $z_m = h(x_m)$. Moreover, $\theta_{im} = -\sin^{-1}[h'(z_m)/\sqrt{1+(h'(z_m))^2}]$ is the local incidence angle, with $h' \equiv dh/dx$, and

$$\mathcal{R}(\theta_{im}) = \frac{\cos \theta_{im} - (\epsilon_r - \sin^2 \theta_{im})^{1/2}}{\cos \theta_{im} + (\epsilon_r - \sin^2 \theta_{im})^{1/2}} \quad (8)$$

is the local TM plane-wave Fresnel reflection coefficient. The PB propagator in (6) is given by

$$b_m(x, z, t) = \text{Re} \left\{ \gamma_m \left[\mathcal{T}_m \Gamma\left(\frac{11}{4}\right) M_1\left(\frac{t - \tau_m}{\mathcal{T}_m}\right) - 2i(t - \tau_m) \Gamma\left(\frac{13}{4}\right) M_2\left(\frac{t - \tau_m}{\mathcal{T}_m}\right) \right] \right\}, \quad (9)$$

where

$$\gamma_m = -i \frac{16 \cdot 2^{1/4} \exp(i\pi/4) L^{1/2} \sigma \Xi T [\sin \theta_{im}(x - x_m) - \cos \theta_{im}(z - z_m)]}{\pi c R_m^{3/2} \mathcal{T}_m^{13/2}},$$

$$\tau_m = \frac{x_m \sin \theta_{im} + R_m + T/2}{c}, \quad T_m = \frac{L \cos^2 \theta_{im}}{c(8\pi)^{1/2}} \left(1 - \frac{z_{bm}}{R_m}\right)^{1/2}, \quad \mathcal{T}_m = (T_m^2 + \sigma^2 T^2)^{1/2},$$

and (x_{bm}, z_{bm}) are beam-centered coordinates,

$$\begin{bmatrix} x_{bm} \\ z_{bm} \end{bmatrix} = \begin{bmatrix} \cos(2\theta_{im}) & -\sin(2\theta_{im}) \\ \sin(2\theta_{im}) & \cos(2\theta_{im}) \end{bmatrix} \begin{bmatrix} x - x_m \\ z - z_m \end{bmatrix}, \quad R_m = (x_{bm}^2 + z_{bm}^2)^{1/2}.$$

Moreover, in (9), $\Gamma(\cdot)$ denotes the gamma function [5], and

$$M_1(t) = {}_1F_1\left(\frac{11}{4}, \frac{1}{2}, -t^2\right), \quad M_2(t) = {}_1F_1\left(\frac{13}{4}, \frac{3}{2}, -t^2\right), \quad (10)$$

where ${}_1F_1(u, v, t)$ is the Kummer confluent hypergeometric function [5]. The functions M_1 and M_2 can be efficiently computed using a rapidly converging power series expansions derived in [3]. The propagator in (9) is a pulsed Gaussian beam, whose collimation is controlled by the discretization period L . The expansion in (6) yields accurate syntheses for $L \ll D$, i.e., for *narrow-waisted*, poorly collimated PB propagators.

The corresponding expansion for the transmitted field can be obtained similarly, as can results for oblique incidence and slight Ohmic losses [4].

IV. RESULTS AND CONCLUSIONS

In order to validate and calibrate the above algorithm, we developed an independent reference solution, based on the time-harmonic multifilament current method exploited in [1]. The spectrum of the scattered field was obtained by solving the scattering problem at 100 different frequencies within the pulse bandwidth. The

resulting frequency samples were smoothed through local Padé-approximation and filtered by the pulse spectrum $P(\omega)$. The TD solution was then obtained by standard inverse FFT algorithms. Typical results are shown in Fig. 2. Specifically, for the temporal behavior of the scattered field at a fixed observation point in the presence of the the interface profile in Fig. 2a, Fig. 2b shows the comparison between the reference solution and the PB synthesis (100 beams) in (6), which involves a modest computational effort when compared with conventional Kirchhoff-PO integration. Good agreement is observed even in the finer details. Convergence and accuracy issues are discussed in [4]. Similar results have been obtained for the transmitted field. Applications to inverse-scattering scenarios are presently under consideration.

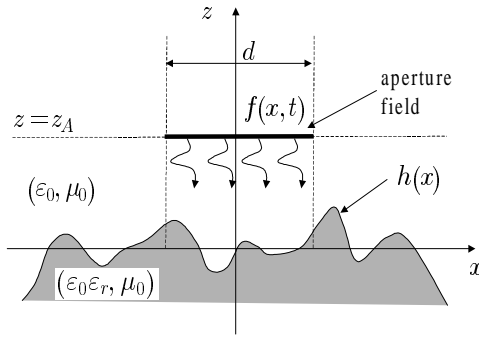


Fig. 1

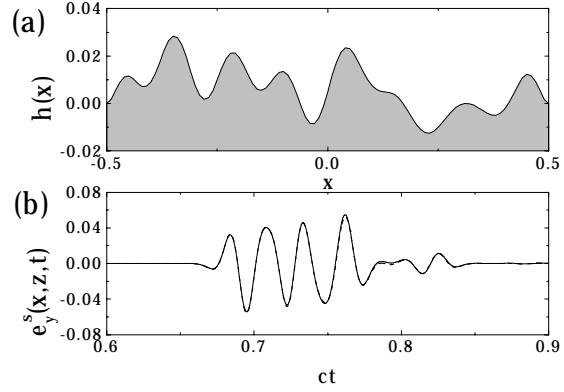


Fig. 2

Fig. 1 - Problem geometry and coordinate system.

Fig. 2 - (a): Rough surface profile (arbitrary units). (b): Temporal behavior of scattered field at $x = 0, z = 0.5$. Aperture tapering: $f(x) = \exp(-18x^2/(\pi d^2))$, $d = 0.8$. Simulation parameters: $\epsilon_r = 3$, $D = 1$, $z_A = 0.2$, $cT = 0.08$, $\Xi = T^4/30000$, $\sigma = 1/\sqrt{50}$.
— Reference solution; - - - Pulsed beam synthesis ($L = D/100$, i.e., 100 beams).

REFERENCES

- [1] V. Galdi, L.B. Felsen, and D.A. Castañón, “Quasi-ray Gaussian beam algorithm for time-harmonic scattering by moderately rough interfaces,” *IEEE Trans. Antennas Propagat.*, 2001 (in print).
- [2] V. Galdi, D.A. Castañón, and L.B. Felsen, “Reconstruction of moderately rough interfaces via quasi-ray Gaussian beams,” submitted to *IEEE Trans. Geosci. Remote Sensing*, Sept. 2000.
- [3] V. Galdi, L.B. Felsen, and D.A. Castañón, “Narrow-waisted Gaussian beam discretization for two-dimensional time-dependent radiation from large apertures,” submitted to *IEEE Trans. Antennas Propagat.*, Sept. 2000.
- [4] V. Galdi, L.B. Felsen, and D.A. Castañón, “Time-domain two-dimensional scattering by moderately rough dielectric interfaces via narrow-waisted Gaussian beams,” submitted to *IEEE Trans. Antennas Propagat.*, Jan. 2001.
- [5] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*. New York (NY): Dover, 1964.

TIME-DOMAIN RECONSTRUCTION OF MODERATELY ROUGH DIELECTRIC INTERFACES VIA QUASI-RAY GAUSSIAN BEAMS *

Julia Pavlovich⁽¹⁾, Vincenzo Galdi^(1,2), W. Clem Karl⁽¹⁾,
David A. Castañon⁽¹⁾, and Leopold B. Felsen⁽³⁾

⁽¹⁾ Multidimensional Signal Processing Laboratory
Department of Electrical and Computer Engineering
Boston University, Boston, MA 02215, USA

⁽²⁾ Waves Group, University of Sannio, Benevento, Italy

⁽³⁾ Department of Aerospace and Mechanical Engineering
and Department of Electrical and Computer Engineering
Boston University, Boston, MA 02215, USA

Also, University Professor Emeritus,

Polytechnic University, Brooklyn, NY 11201, USA

E-mail: oska@bu.edu, vgaldi@bu.edu, wckarl@bu.edu, dac@bu.edu, lfelsen@bu.edu

I. PROBLEM FORMULATION

The problem of determining the properties of rough surfaces from electromagnetic scattered field data arises in many important subsurface sensing problems such as underground imaging using data from a Ground Penetrating Radar (GPR). In GPR systems, arrays of above-ground transmitters and receivers illuminate areas of interest and receive backscattered signals from shallowly buried objects and from surface reflections. The shape of the air-ground interface is unknown, and it is a principal corruptor of the backscattered signal from subsurface targets of interest. In order to enhance subsequent detection, classification and inverse scattering processing, it is important to compensate for the distortion introduced by the air-ground interface. As a first step toward this goal, in a recent investigation [1] we addressed the problem of estimating the profile of one-dimensional (1D) moderately rough dielectric interfaces from a limited number of multifrequency measurements under time-harmonic excitation. In this paper, we extend the approach to the case of time-dependent (short pulse) excitation. The proposed extension is based on a recently developed Gabor-based narrow-waisted pulsed-beam (PB) representation of short-pulse scattering by moderately rough surfaces [2].

The problem geometry is depicted in Fig. 1: in an (x, z) 2D coordinate space, a TM-polarized incident field generated by a pulsed aperture field distribution of width

*Work supported by ODDR&E under MURI grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The work of V. Galdi was also supported by a European Union postdoctoral fellow ship through the University of Sannio, Benevento, Italy. L.B. Felsen also acknowledges partial support from a Grant by the US-Israel Binational Science Foundation, Jerusalem, Israel, and from Polytechnic University.

d at $z = z_A$ is assumed to impinge from free-space onto a lossless, nondispersive, homogeneous dielectric half-space of known relative permittivity ϵ_r , bounded by a moderately rough interface described by the continuous function $h(x)$. We assume a tapered, separable, nonphased space-time aperture field distribution

$$e_y^{inc}(x, z = z_A, t) = \begin{cases} f(x)p(t), & |x| \leq d/2, \\ 0, & |x| > d/2, \end{cases} \quad (1)$$

where $p(t)$ is a pulse of length $T \ll d/c$, with c representing the speed of light, and $f(x)$ is the amplitude taper profile. The interface is located in the collimation zone of the aperture so that the incident field can be approximated by a pulsed truncated tapered plane wave propagating in the negative z -direction (see Fig. 1). The y -directed scattered field e_y^s is sampled at N_t time instants at N_r receiver locations (at $z = z_{obs}$). This setup is typical of GPR systems for detecting shallowly buried anti-personnel mines, of interest for humanitarian demining. However, in this investigation we ignore the presence of any buried objects. Furthermore, we are not taking into account the *noisy* (incoherent) contribution of finer-scale roughness, and we aim to estimate the coarse shape of the surface, acknowledging the implicit limits of retrievable information through inverse scattering. Due to the inherent *ill-posedness* of this kind of inverse scattering problems, it is essential to incorporate stabilization steps, known as regularization methods. As in [1], we introduced an “implicit” regularization by exploiting an appropriate discrete representation of the surface. Accordingly, the surface estimation problem is posed as a nonlinear optimization problem, whose solution requires repeated evaluations of the scattered field at the receiver locations. The efficiency of such an approach thus strongly relies on the availability of a fast forward scattering model.

II. FORWARD MODEL

Our forward scattering model is detailed in [2]. Technically, it is based on the Kirchhoff Physical Optics (PO) approximation and on the Gabor-based narrow-waisted PB discretization of 1D aperture field distributions in [3]. The time domain is accessed by analytic Fourier inversion, starting from the frequency domain PO model discretized in terms of *complex source point* narrow-waisted *quasi-ray* Gaussian beams. For the class of Rayleigh (differentiated Gaussian) time pulses we obtained closed form solutions in terms of rapidly computable analytic functions [2]. The algorithm has been validated and calibrated against an independently generated reference solution based on an extension [2] of the multifilament current method [4]. This reference solution is also used to simulate the measured data needed in the inversion procedure (see Sec. III). The proposed forward scattering model has been shown capable of furnishing accurate and numerically efficient predictions in a calibrated range of parameters, so that it appears to be suitable for the inverse scattering scenario described in Sec. I.

III. TIME DOMAIN INVERSION APPROACH

As stated previously, the proposed *robust* inversion strategy is based on a low-dimensional compact geometrical parameterization of the unknown interface profile, whose unknown parameters are estimated by *fitting* the model-based forward

scattering prediction to the available (measured/simulated) data, i.e., minimizing a suitable *error functional*. The choice of the interface profile parameterization requires tradeoff between compactness and versatility, keeping in mind that the number of unknown parameters N to be estimated must be *smaller* than the collected scattered field data size, i.e., $N \sim N_r \times N_t$. We assume that the surface shape can be approximated by a finite set of quartic B-spline basis functions with unknown coefficients,

$$h(x) \approx \sum_{n=1}^N c_n s_4(x). \quad (2)$$

Let $E_{pq}^s = E_y^s(x_p, t_q)$ denote the y -directed scattered field *measured* at time t_q at sensor position x_p . Given a vector of spline coefficients $\underline{c} = \{c_1, \dots, c_N\}$ and the outgoing field from the aperture distribution in (1), we can use the PB algorithm in [2] for the surface profile in (2) to generate predictions of the scattered field waveforms at each receiver. Let $e_{pq}^s = e_y^s(x_p, t_q; \underline{c})$ denote the *model-based prediction* of the scattered field at time t_q at sensor position x_p , for surface profile coefficients \underline{c} . With this notation, we define the error functional as follows

$$J(\underline{c}) = \|\underline{\mathbf{e}}^s(\underline{c}) - \underline{\mathbf{E}}^s\|^2, \quad (3)$$

where $\underline{\mathbf{e}}^s = \{e_{pq}^s\}$, $\underline{\mathbf{E}}^s = \{E_{pq}^s\}$, $p = 1, \dots, N_r$, $q = 1, \dots, N_t$. As anticipated, the measured data are simulated via the reference solution described in [2]. The regularized inverse scattering problem can now be formalized as finding the coefficient vector \underline{c} in (2) which minimizes the error functional $J(\underline{c})$ in (3), i.e. finding $\hat{\underline{c}}$ such that

$$\hat{\underline{c}} = \arg \min_{\underline{c}} J(\underline{c}). \quad (4)$$

This minimization is nontrivial since the cost functional in (4) is likely non-convex with respect to \underline{c} . Therefore, standard descent-based optimization algorithms are prone to get stuck in local minima, and global optimization techniques are needed. In the frequency domain approach, i.e., for frequency-stepped time-harmonic excitation [1], we used a *continuation method* based on the selected use of the data at the various available frequencies. We are presently working on the extension of this *multiresolution* optimization strategy to the case of short-pulse excitation.

IV. RESULTS AND CONCLUSIONS

To illustrate the results achievable with our technique we show in Fig. 2 a reconstruction example for stepped frequency GPR measurements, obtained using the algorithm in [1]. The soil was modeled as lossless, nondispersive, with relative permittivity $\epsilon_r = 3$. The maximum slope of the rough surface profile is 34° . The interrogating field is generated by a cosine-tapered aperture field distribution of width $d = 0.8$ located at $z_A = 0.6$. The complex scattered field is measured at 10 equispaced observation points at $z_{obs} = 0.6$ spanning the illuminated region, and at four operating wavelengths ($\lambda = 0.2, 0.1, 0.067, 0.05$). The length scales are normalized with respect to the smallest radius of curvature R_c of the local interface profile maxima; it has been shown in [5] that for $\lambda/R_c \lesssim 0.2$, the high-frequency narrow-waisted beam algorithm actually works well over the range of problem parameters in our example here, although the lower limit $\lambda/R_c \sim 0.2$ is in

the “low frequency” region. The resulting error functional to be minimized is similar to (3), and the multiresolution minimization strategy detailed in [1] is exploited. As one can see, the reconstruction is accurate throughout most of the interval, except near the edges of the illumination region. This kind of accuracy was observed in many numerical experiments, even in noisy environments [1].

We are currently conducting numerical experiments with short pulse GPR excitations. These results will be reported at the conference.

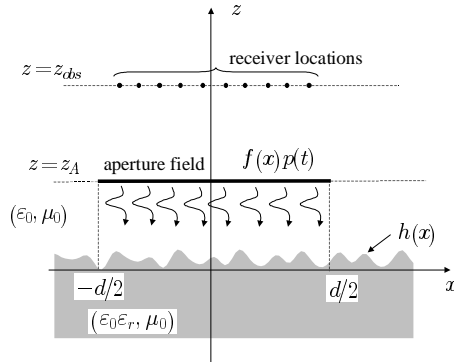


Fig. 1 - Problem geometry.

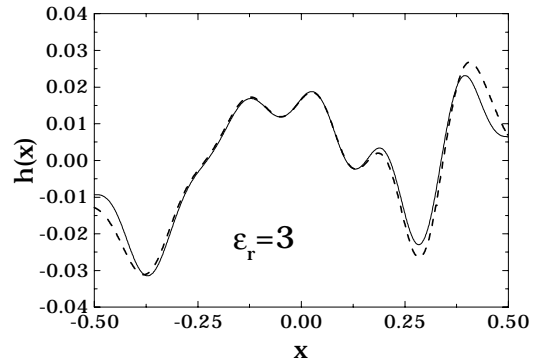


Fig. 2 - Reconstruction example for stepped frequency measurements.
— True; - - - Reconstruction.

REFERENCES

- [1] V. Galdi, D.A. Castañon, and L.B. Felsen, “Reconstruction of moderately rough interfaces via quasi-ray Gaussian beams,” submitted to *IEEE Trans. Geosci. Remote Sensing*, Sept. 2000.
- [2] V. Galdi, L.B. Felsen, and D.A. Castañon, “Time-domain 2D scattering by moderately rough dielectric interfaces via narrow-waisted Gaussian beams,” this Conference. Also, submitted to *IEEE Trans. Antennas Propagat.*, Jan. 2001.
- [3] V. Galdi, L.B. Felsen, and D.A. Castañon, “Narrow-waisted Gaussian beam discretization for two-dimensional time-dependent radiation from large apertures,” submitted to *IEEE Trans. Antennas Propagat.*, Sept. 2000.
- [4] Y. Leviatan and A. Boag, “Analysis of electromagnetic scattering from dielectric cylinders using a multifilament current model,” *IEEE Trans. Antennas Propagat.*, vol. 35, No. 10, pp. 1119-1127, Oct. 1987.
- [5] V. Galdi, L.B. Felsen, and D.A. Castañon, “Quasi-ray Gaussian beam algorithm for time-harmonic scattering by moderately rough interfaces,” *IEEE Trans. Antennas Propagat.*, 2001 (in print).

Gaussian beam algorithm for rough surface underground imaging

Vincenzo Galdi^{1,2}, Leopold B. Felsen^{1,3}, and David A. Castañon¹

(1) Department of Electrical and Computer Engineering, Boston University
8 Saint Mary's St., Boston, MA 02215, USA

(2) Waves Group, University of Sannio, I-82100 Benevento, Italy

(3) Department of Aerospace and Mechanical Engineering, Boston University
110 Cummington St., Boston MA 02215, USA

Also, University Professor Emeritus, Polytechnic University, Brooklyn, NY 11201, USA

E-mail: vgaldi@bu.edu, lfelsen@bu.edu, dac@bu.edu

Abstract— An adaptive framework is presented for wide-band ground penetrating radar imaging of low-contrast buried objects in the presence of a moderately rough air-soil interface. The proposed approach relies on recently developed Gabor-based narrow-waisted Gaussian beam algorithms as fast forward scattering predictive models. A prior nonlinear inverse scattering problem is solved to estimate the unknown coarse scale roughness profile. The Born-linearized underground imaging problem is subsequently solved compensating for the clutter effect produced by the estimated roughness profile. Preliminary outcomes indicate that the proposed framework is attractive as compared to standard statistical approaches.

I. INTRODUCTION

In ground penetrating radar (GPR) applications, modelling the distortion, introduced by a twice-traversed rough air-ground interface, in the interrogating signal on its way to and from the targets of interest is a key issue. Standard statistical approaches, which tend to model such distortion as an additive colored Gaussian noise, perform reasonably well in *detection* problems [1], but have been found to yield limited accuracy and robustness in *classification* and *reconstruction* (see, e.g., [2]).

In this paper, with particular reference to the problem of anti-personnel plastic land mine detection/classification via wide-band GPR, we present a novel *adaptive* approach, based on a prior estimation of the unknown coarse scale roughness profile and a subsequent *quasi-deterministic* compensation of the related clutter. The proposed approach is built on recently developed Gabor-based narrow-waisted Gaussian beam (GB) algorithms for short-pulse scattering from, and transmission through, moderately rough dielectric interfaces [3]. By exploiting these fast for-

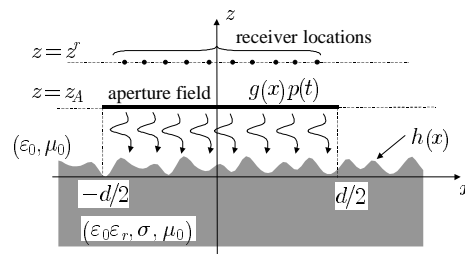


Fig. 1. Problem geometry.

ward models, the prior surface estimation is posed as a nonlinear optimization problem [4]. A beam-computed Green's function accounting for the distortion introduced by the estimated roughness profile is exploited in the subsequent underground imaging problem, which is linearized via Born approximation [5] and solved for the unknown dielectric permittivity contrast, using various *point-wise* and *object-based* reconstruction techniques.

II. ROUGH SURFACE FORWARD SCATTERING

The problem geometry is sketched in Fig. 1: in an (x, z) two-dimensional (2D) coordinate space, a lossy homogeneous dielectric half-space (soil) of known relative permittivity ϵ_r and conductivity σ , bounded by a moderately rough interface $z = h(x)$, is illuminated by a y -directed (TM-polarized) pulsed field generated by a truncated aperture field distribution of width d at $z = z_A$. We assume that the incident field can be approximated by a pulsed truncated tapered plane wave propagating in the negative z -direction (Fig. 1),

$$e^i(\mathbf{r}, t) \sim g(x)p[t - c^{-1}(z - z_A)], \quad (1)$$

where $\mathbf{r} \equiv (x, z)$, $p(t)$ is a pulse of length $T \ll d/c$, with c representing the speed of light, and $g(x)$ is the taper profile. In this preliminary investigation, we ignore the presence of buried targets.

The forward scattering model, detailed in [3], is based on the Kirchhoff Physical Optics (PO) approximation in conjunction with the Gabor-based narrow-waisted pulsed beam (PB) discretization of 1D aperture field distributions in [6], and is restricted to moderate roughness (both in height and slope) and slightly lossy soils. The PO equivalent currents, which generate reflected and transmitted fields, are first parameterized in the frequency domain in terms of x -domain discretized m -indexed Gabor basis functions with narrow width L , centered on the Gabor lattice points $x_m = mL$; these initial conditions generate *narrow-waisted*, quasi-ray, complex-source-point GBs propagating along the local reflection/refraction directions. For Rayleigh (i.e., differentiated Gaussian) pulses, the resulting time domain analytic Fourier inversion integral can be approximated by rapidly computable closed form expressions. One obtains the following approximate PO-PB expansions for reflected and transmitted fields e^r and e^t , respectively (see [3] for details),

$$e^\nu(\mathbf{r}, t) \sim \sum_{|m| \leq (d/2L)} a_m^\nu b_m^\nu(\mathbf{r}, t - t_m), \quad \nu = r \text{ or } t, \quad (2)$$

where the Gabor expansion coefficients a_m^ν and the time delays t_m are approximated by sampling the PO equivalent currents at the lattice points $x_m = mL$, and the PB propagators b_m^r and b_m^t are expressed in terms of rapidly computable confluent hypergeometric functions [3].

III. INVERSE SCATTERING

A. Rough interface estimation

The interface estimation problem has been addressed in [7] for frequency-stepped time-harmonic excitation and *sparse* data, and has been extended in [4] to pulsed excitation. Here, we briefly review the pulsed case. Due to the inherent *ill-posedness* of the problem, we exploit a *robust* inversion strategy based on an appropriate compact low-dimensional representation of the roughness profile, whose unknown parameters are estimated by fitting the PO-PB forward scattering prediction to the available (measured/simulated) data, i.e., minimizing a suitable *cost functional*. The resulting *nonlinear* optimization problem requires repeated evaluations of the reflected field at the receiver locations, and therefore its computational efficiency is strongly tied to the efficiency of the PO-PB algorithm. We found that a quartic spline parameterization

$$h(x) \approx \sum_{n=1}^N c_n s_n^{(4)}(x), \quad (3)$$

provides a good tradeoff between versatility and computational burden. In (3), c_n are unknown coefficients, and $s_n^{(4)}$ represents a quartic B-spline basis function. With $\hat{e}_{pq}^r = \hat{e}^r(x_p^r, z^r, t_q)$ denoting the y -directed reflected field *measured* at time t_q at receiver

location (x_p^r, z^r) (Fig. 1), and $e_{pq}^r = e^r(x_p^r, z^r, t_q; \underline{c})$ denoting the corresponding PO-PB *forward prediction* for surface profile coefficients $\underline{c} = \{c_1, \dots, c_N\}$, we formalize the *regularized* inverse scattering problem as finding the coefficient vector \underline{c} in (3) which minimizes the cost functional

$$J(\underline{c}) = \|\underline{e}^r(\underline{c}) - \hat{\underline{e}}^r\|^2 = \sum_{p=1}^{N_r} \sum_{q=1}^{N_t} \gamma_p (e_{pq}^r - \hat{e}_{pq}^r)^2, \quad (4)$$

where N_r is the number of receivers, N_t the number of time samples at each receiver, and γ_p are normalization coefficients. The cost functional in (4) is generally *non-convex* with respect to \underline{c} , and therefore the possible presence of local minima renders its minimization non-trivial. The resulting global optimization strategies and computational issues are discussed in [4].

B. Underground imaging

We again refer to the problem geometry in Fig. 1, but now include a buried plastic target with dielectric permittivity $\epsilon_r^{(tar)}$ and conductivity $\sigma^{(tar)} \approx 0$ occupying the region $D^{(tar)}$ in the half-space $z < h(x)$. The TM-polarized *total* field observed at the receiver location $\mathbf{r}^r = (x^r, z^r)$ can be written as

$$e(\mathbf{r}^r, t) = e^b(\mathbf{r}^r, t) + e^s(\mathbf{r}^r, t), \quad (5)$$

where e^b represents the *background field* (the field in the absence of the target, i.e., e^r in (2)), and

$$e^s(\mathbf{r}^r, t) = - \iint_{D^{(tar)}} \left\{ \frac{\Delta\epsilon_r(\mathbf{r}')}{c^2} \partial_t^2 [g_b(\mathbf{r}^r, \mathbf{r}', t, t') \otimes e(\mathbf{r}', t')] + \frac{\sigma}{c\epsilon_0} \partial_t [g_b(\mathbf{r}^r, \mathbf{r}', t, t') \otimes e(\mathbf{r}', t')] \right\} d\mathbf{r}' \quad (6)$$

is the field scattered by the target. In (6), the *object function* $\Delta\epsilon_r \equiv (\epsilon_r^{(tar)} - \epsilon_r)$ represents the unknown relative permittivity contrast, c is the speed of light, g_b denotes the space-time Green's function of the rough-interface dielectric half-space, e is the *total* field in the target region for pulsed excitation $p(t)$, and \otimes indicates time convolution. Assuming that an estimation of the rough interface profile is available, one can generate prediction of the background field contribution e^b in (5) for a given pulsed excitation, and hence *isolate* the contribution e^s from the target. For plastic anti-personnel land mines, of particular interest in our investigation, the dielectric properties of the target are typically close to those of the background soil, i.e., $\epsilon_r^{(tar)} \approx \epsilon_r$. Accordingly, one can exploit the *weak-scattering* Born approximation [5],

$$e(\mathbf{r}', t') \approx e^t(\mathbf{r}', t'), \quad \mathbf{r}' \in D^{(tar)}, \quad (7)$$

where the *total* field e inside the target is approximated by the *transmitted* field e^t in the target region *in the absence of the target*, thus yielding in (6) a *linear* model relating the scattered field at the receivers to the dielectric contrast $\Delta\epsilon_r$. The needed convolutions $g_b \otimes e^t$ and their time derivatives in (6) with (7)

can be efficiently computed using PO-PB expansions similar to (2).

Based on the linear forward scattering model in (6) and (7), and on a number of observations of the scattered field waveforms at the receivers, an inversion scheme can be constructed by discretizing the test domain $D^{(test)}$ to be imaged into a number of adequately small pixels. Assuming that N_t time samples of the scattered field e^s are collected at N_r receivers, and that the test domain is discretized into N_p pixels, the linear forward model in (6) and (7) can be discretized accordingly and cast into matrix form as

$$\underline{y} = \underline{A} \cdot \underline{x} + \underline{n}, \quad (8)$$

where \underline{y} is a column vector containing the $N_r \cdot N_t$ time samples of the *known term* in (6) and (7),

$$e^s(\mathbf{r}^r, t) + \frac{\sigma}{c\epsilon_0} \iint_{D^{(test)}} \partial_t [g_b(\mathbf{r}^r, \mathbf{r}', t, t') \otimes e^t(\mathbf{r}', t')] d\mathbf{r}', \quad (9)$$

\underline{x} is a N_p -element column vector containing the *unknown* dielectric contrast $\Delta\epsilon_r$ at each pixel, \underline{A} is a $(N_r \cdot N_t) \times N_p$ matrix containing the space-time discretization of the integral

$$-c^{-2} \iint_{D^{(test)}} \partial_t^2 [g_b(\mathbf{r}^r, \mathbf{r}', t, t') \otimes e^t(\mathbf{r}', t')] d\mathbf{r}', \quad (10)$$

and the *noise* vector \underline{n} accounts for measurement uncertainty and unmodeled effects. Due to the inherent ill-posedness of this inverse scattering problem, and the various approximations in our model, it is essential to introduce some regularization. In this connection, we have explored various *edge-preserving* regularization approaches such as *total variation* [2] and *curve evolution* [8]. In total variation regularization, the problem is posed as the minimization of the functional

$$\Phi(\underline{x}) = \|\underline{y} - \underline{A} \cdot \underline{x}\|^2 + \lambda_1 \|\underline{K} \cdot \underline{x}\|^2 + \lambda_2 \|\underline{D} \cdot \underline{x}\|^1, \quad (11)$$

where the parameters λ_1 and λ_2 are usually determined empirically, \underline{D} is a spatial gradient operator, and the operator \underline{K} penalizes negative values of the reconstructed object function (the sign of the object function $\Delta\epsilon_r$ is assumed to be known a priori)

$$[\underline{K} \cdot \underline{x}]_i = \begin{cases} x_i, & x_i < 0, \\ 0, & x_i \geq 0. \end{cases} \quad (12)$$

In (11), the L_1 norm on the spatial gradient term encourages piecewise smoothness in the reconstructed object function and, unlike the L_2 norm in standard Tikhonov regularization, allows sharp edges to form, yielding visually better (less blurred) reconstructions. Alternatively, one can exploit *object-based* reconstruction approaches, applicable to piece-wise smooth object functions, where instead of a *point-wise* reconstruction one focuses on the *direct estimation* of the target boundary and the dielectric contrast value, with a considerable saving in the number of unknowns and a consequent better conditioning of the problem.

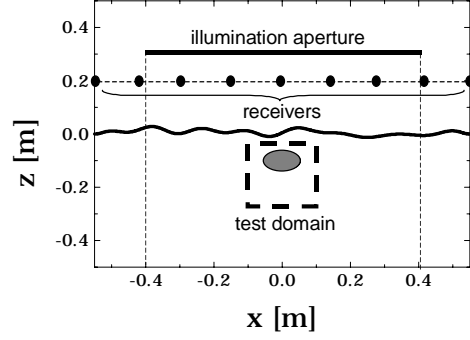


Fig. 2. Simulation geometry and parameters. Soil: $\epsilon_r = 3$, $\sigma = 0.01$ S/m. Target: $10\text{cm} \times 6\text{cm}$ ellipse with center at 10cm below nominal ground, $\epsilon_r^{(tar)} = 3.3$. Excitation: cosine-tapered aperture field distribution ($z_A = 30\text{cm}$, $d = 80\text{cm}$), fourth-order Rayleigh wide-band excitation (2.45 GHz center frequency, 1.4 GHz bandwidth). Test domain: $20\text{cm} \times 20\text{cm}$ square (30×30 pixels, i.e., $N_p = 900$). $N_r = 11$ equispaced receivers at 20cm above nominal ground.

This is usually accomplished via curve evolution techniques, i.e., by designing a gradient flow which attracts initial closed curves to the target boundary (see [8] for details).

IV. RESULTS

The simulation geometry and parameters are illustrated in Fig. 2. We consider a plastic mine-like $10\text{cm} \times 6\text{cm}$ elliptic target with relative permittivity $\epsilon_r^{(tar)} = 3.3$ shallowly buried in a homogeneous dielectric half-space with constitutive parameters chosen so as to simulate a class of realistic sandy soils ($\epsilon_r = 3$, $\sigma = 0.01\text{S/m}$). The rough surface realization, parameterized in this experiment with 20 B-spline basis functions, was generated randomly so as to mimic typical ($\sim 3 - 4$ cm peak-to-peak, maximum slope $\sim 30^\circ$) natural (moderate) roughness. A cosine-tapered transverse field distribution of aperture width $d = 80\text{cm}$ was chosen, with a fourth-order Rayleigh wide-band excitation (2.45 GHz center frequency, 1.4 GHz bandwidth). Synthetic field measurement data were generated via full-wave solution of the forward scattering problem (see [3] for details). A $20\text{cm} \times 20\text{cm}$ square test domain surrounding the target to be imaged was discretized in 30×30 pixels. In this experiment, $N_r = 11$ equispaced receivers at 20cm above nominal ground were used.

For the interface estimation problem, $N_t = 100$ time samples of the *early-time* response (i.e., excluding causal contributions from the test domain) at each receiver were used in the cost functional (4), thus avoiding any possible bias due to target scattering. The cost functional (4) was minimized using the conjugate gradient (CG) strategy described in [4], assuming as initial guess a flat interface at $z = 0$. A typical reconstruction example is shown in Fig. 3. The reconstruction is quite accurate throughout most of the interval, except near the edges of the illumina-

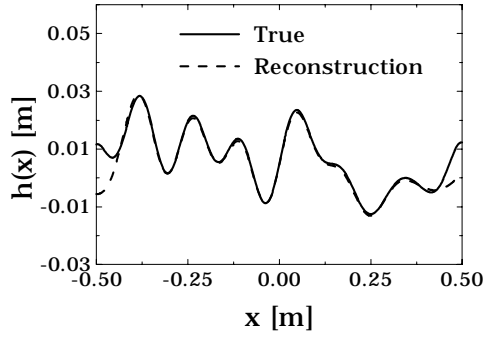


Fig. 3. Parameters as in Fig. 2. Rough interface profile reconstruction example ($N_t = 100$).

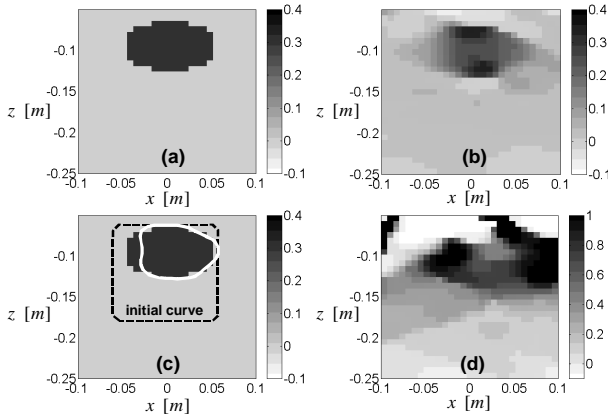


Fig. 4. Parameters as in Fig. 2. Underground imaging examples ($N_t = 300$). (a): Ground truth ($\Delta\epsilon_r = 0.3$). (b): Total variation reconstruction with adaptive compensation. (c): Curve evolution reconstruction with adaptive compensation; estimated target boundary (white curve) is superposed on ground truth; estimated dielectric contrast: $\Delta\epsilon_r = 0.32$. (d): Total variation reconstruction without compensation (flat interface at $z = 0$).

tion region. This kind of accuracy was observed in many numerical experiments, with CG convergence typically achieved in 30 – 40 iterations.

The obtained interface estimation was used to compute the background field e^b in (5), and hence isolate the target contribution e^s . For the underground imaging problem, $N_t = 300$ time samples of the *late-time* response (i.e., causally related to the test domain) were used. The data matrix $\underline{\underline{A}}$ and the known term \underline{y} in (8) were computed using a PO-PB algorithm similar to (2). Both total variation (via the iterative scheme in [2]) and curve evolution (using the algorithm in [8]) reconstructions were attempted. The underground imaging algorithms are still under investigation, and are not yet fully optimized and calibrated. Some preliminary results are shown in Fig. 4. Specifically, the true object function (ground truth) is shown in Fig. 4a, and the corresponding total variation and curve evolution reconstructions are shown in Figs. 4b and 4c, respectively. It is observed that total variation, though not yielding a highly accurate

point-wise reconstruction, provides a reasonably accurate target localization. On the other hand, curve evolution provides rather accurate direct estimations of *both* target boundary *and* dielectric contrast. Note that the limited viewing geometry renders the problem *more ill-posed* in the horizontal direction, resulting in less accurate horizontal localization in both cases. The effect of the interface roughness and the importance of adaptive compensation is highlighted in Fig. 4d, where a total variation reconstruction *without* any compensation (i.e., assuming a flat interface at $z = 0$) is shown. The poor quality of the reconstruction can be only slightly improved using statistical processing, as shown in [2], but is still not comparable to that of Figs. 4b,c.

V. CONCLUSIONS

An adaptive approach for rough surface underground imaging has been presented, which exploits short-pulse Gabor-based narrow-waisted Gaussian beam algorithms as fast forward scattering models. Preliminary 2D results, restricted to slightly lossy soils with moderate roughness and low-contrast targets, show that quite accurate estimations of the coarse scale roughness profile can be obtained from reflected field sparse data, and can be fruitfully exploited to enhance underground imaging, with reasonable computing time and resources.

ACKNOWLEDGEMENTS

This work was supported in part by ODDR&E under MURI Grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028, and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. The help of Haihua Feng (Boston Univ.) in the inverse scattering simulations is gratefully acknowledged.

REFERENCES

- [1] T. Dogaru and L. Carin, "Time-domain sensing of targets buried under a rough air-ground interface," *IEEE Trans. Antennas Propagat.*, Mar. 1998., Vol. 46, pp. 360–372.
- [2] H. Feng, D.A. Castañon, W.C. Karl, and E.L. Miller, "GPR imaging approaches for buried plastic landmine detection," *Proc. SPIE*, Aug. 2000, Vol. 4038, pp. 1485–1496.
- [3] V. Galdi, L. B. Felsen, and D. A. Castañon, "Quasi-ray Gaussian beam algorithm for short-pulse two-dimensional scattering by moderately rough dielectric interfaces," submitted to *IEEE Trans. Antennas Propagat.*, Apr. 2001.
- [4] V. Galdi, W.C. Karl, D. A. Castañon, and L. B. Felsen, "Approaches to underground imaging for object localization," *Proc. SPIE*, Aug. 2001, Vol. 4394 (in print).
- [5] W. C. Chew, *Waves and Fields in Inhomogeneous Media*, Oxford Press, Oxford, UK, 1996.
- [6] V. Galdi, L. B. Felsen, and D. A. Castañon, "Narrow-waisted Gaussian beam discretization for short-pulse radiation from one-dimensional large apertures," *IEEE Trans. Antennas Propagat.*, Sept. 2001, Vol. 9 (in print).
- [7] V. Galdi, D. A. Castañon, and L. B. Felsen, "Reconstruction of moderately rough interfaces via quasi-ray Gaussian beams," submitted to *IEEE Trans. Geosci. Remote Sensing*, Sept. 2000.
- [8] H. Feng, W. C. Karl, and D. A. Castañon, "A curve evolution approach to object-based tomographic reconstruction," submitted to *IEEE Trans. Image Processing*, 2000.

Underground Imaging Based on Edge-Preserving Regularization

Haihua Feng, David A. Castañón, W. Clem Karl

*Multi-Dimensional Signal Processing Laboratory
Department of Electrical and Computer Engineering
Boston University
8 Saint Mary's Street, Boston, MA 02215, USA
hfeng@bu.edu , dac@bu.edu, wckarl@bu.edu*

Abstract

This paper develops new approaches for imaging weak-contrast buried objects using data from a ground penetrating radar array. An approximate physical model relating the collected data to the underground objects is developed. This model uses ray optics to represent the air/soil interface, and a Born approximation to model the weak contrast back-scattering from buried objects. In order to address both modeling errors and ill-posedness, the proposed image reconstruction algorithms use regularization based on a total variation norm with orientation preference. The algorithms are tested on data generated by nonlinear finite difference time domain electromagnetic simulations.

1 Introduction

Problems of locating and reconstructing physically inaccessible objects arise in many different applications such as landmine detection and removal, seismic imaging, archaeological surveying, and hydrology. Many of these applications require techniques for accurate imaging that provide reliable information about buried objects' physical properties. This is particularly difficult when there is weak contrast between the properties of the buried object and the surrounding soil.

This paper focuses on underground imaging techniques for weak contrast buried objects using electromagnetic Ground Penetrating Radar(GPR). GPR is the preferred approach for detection and localization of plastic mines [4], where the traditional metal detectors fail. The approach proposed in this paper is based on the principles of diffraction tomography [5][6] for imaging weak-contrast dielectric objects. In diffraction tomography, the imaging problem is posed as an inverse scattering problem, and a physically-based linear model is used to define the relationship between collected data and buried objects' properties. Reconstructing an estimate of the objects' profiles is thus a linear inversion problem, for which many techniques have been developed based on singular value decompositions, regularization and optimization methods [7].

For the problems of interest in this paper, there are two major physical effects to consider: propagation through an air-ground interface, and backscattering from objects buried underground. The air/soil interface is modeled using ray optics. For backscattering, the physical models are based on the Born approximation in inverse scattering. This is because there is low contrast between the soil and mine permittivities ($\epsilon_{soil} = 2.6$, $\epsilon_{mine} = 2.9$), resulting in weak scattering. The resulting linear models relate the received signals to the transmitted signals and the underlying underground objects. This linear transformation is approximate and ill-conditioned, requiring algorithms which are robust to modeling errors and sensor noise.

Our proposed approach for inversion is based on robust regularization. Since edge information is important for delineating buried objects, we use total variation regularization [9], instead of conventional regularization techniques such as Tikhonov and singular value decomposition methods which result in blurred edges. Furthermore, we exploit the structure of the vertical illumination geometry for buried objects to introduce different levels of horizontal and vertical regularization. The rest of this paper is organized as

follows. Section 2 introduces the configuration of the underground imaging system. Section 3 contains the model of the backscattering through a flat air/soil surface. Section 4 describes the total variation based regularization method. Section 5 compares the performance of the proposed algorithms with that of standard regularization techniques on examples generated using high-fidelity nonlinear simulations.

2 Imaging System Configuration

The underground imaging system of interest consists of an array of GPR transmitters and a different array of receivers, both located above ground. Fig. 1 shows the geometry of the sensor and object configuration. To simplify the analysis, we consider a two-dimensional problem with transverse magnetic(TM) polarization mode for the signal, a flat interface between air and soil, and uniform relative permittivity for the background soil. The combined GPR transmitters generate a short pulse TM wave, approximated by a plane wave in the region of interest, propagating perpendicular to the soil. The pulse width is approximately 0.8 nanoseconds, corresponding to a 1.25 GHz bandwidth.

Interactions between the transmitted wave and the soil result in part of the plane wave penetrating the underground soil. This wave interacts with the buried object, generating backscattered signals which must propagate through the ground/air interface and up to the receivers.

The region enclosed by the dash lines is the region of interest for reconstruction. Since the TM mode is used, the magnetic field orientation is perpendicular to the planar geometry, so only its H_z component is non-zero. Similarly, the electric field has components only in the plane, so only E_x and E_y are non-zero. Although the transmitted signal has negligible E_y component because of its plane wave orientation, the backscattered signal will have both E_x and E_y components. We assume that the receivers are polarized, and receive only the contribution of the E_x field.

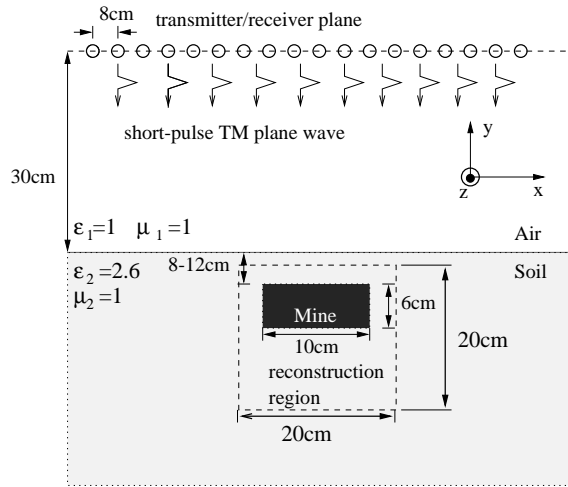


Figure 1: Geometry of the sensor and mine configuration

3 Problem Formulation

The basic mechanism for backscattering is due to a difference in the relative permittivities of the objects of interest and the surrounding background soil. We assume that the object of interest is described by the “object function” at position \vec{r} as

$$O(\vec{r}) = n(\vec{r})^2 / n_{soil}^2 - 1 \quad (1)$$

where $n(\vec{r}) = \sqrt{\mu(\vec{r})\epsilon(\vec{r})}$ is the square root of the product of the relative permeability and permittivity at \vec{r} . For the imaging problems of interest, the relative permeabilities of the object and background are close: $\mu(\vec{r}) \approx \mu_{soil} \approx 1$. Thus,

$$O(\vec{r}) = n(\vec{r})^2 / n_{soil}^2 - 1 = \epsilon(\vec{r}) / \epsilon_{soil} - 1 \quad (2)$$

The field $u(\vec{r}, t)$ satisfies the wave equation

$$\left[\nabla^2 - \frac{1}{v(\vec{r})^2} \partial_t^2 \right] u(\vec{r}, t) = 0 \quad (3)$$

where $v(\vec{r})$ is the wave propagation speed at position \vec{r} , which is related to the relative permittivity by

$$v(\vec{r}) = v_{soil} \sqrt{\frac{\varepsilon_{soil}}{\varepsilon(\vec{r})}} \quad (4)$$

where v_{soil} is the wave propagation speed in soil. Under the time-dependent Born approximation, the field scattered by the buried object can be expressed in the form

$$u(\vec{r}, t) \propto -\partial_t^2 \int d\vec{r}' \int dt' O(\vec{r}) u^i(\vec{r}', t') G(\vec{r}, t; \vec{r}', t') \quad (5)$$

where $u^i(\vec{r}', t')$ is the incident pulsed plane wave, and $G(\vec{r}, t; \vec{r}', t')$ is the time-dependent Green's function in the presence of the air/soil interface determining the field strength at \vec{r}, t from a point source at \vec{r}', t' . Note that, although the buried object is assumed to be in weak contrast with the soil, the air properties are strongly in contrast with that of the soil, and thus some care is required to compute the Green's function G . Fig 2 shows the reflection and refraction of the TM mode at the air/soil interface. Given source position a and destination position c , according to Snell's law, $\frac{\sin\theta_1}{\sin\theta_2} = \sqrt{\frac{\mu_2 \varepsilon_2}{\mu_1 \varepsilon_1}} \approx \sqrt{\frac{\varepsilon_2}{\varepsilon_1}}$, so we can determine the position b on the air/soil interface. Since the received measurement consists of E'_x , the field of interest $u(\vec{r}, t) = E'_x(\vec{r}, t)$ in our problem. The refraction coefficient of E'_x is given by

$$\Gamma = \frac{E'_x}{E} = \frac{2\cos\theta_1\sin\theta_2\cos\theta_2}{\sin(\theta_1 + \theta_2)\cos(\theta_1 - \theta_2)} \quad (6)$$

Using this model, the time-dependent Green's function in the presence of an air/soil interface can be

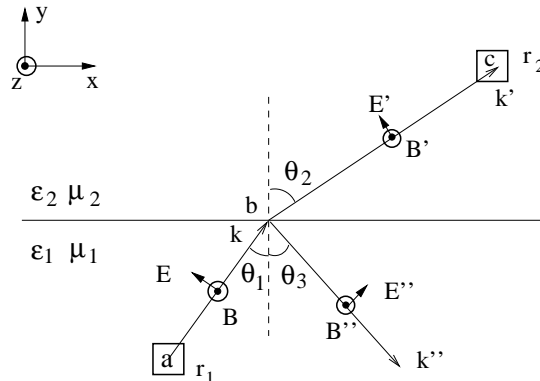


Figure 2: Reflection and refraction in TM mode

expressed as

$$G(\vec{r}, t; \vec{r}', t') = \frac{\Gamma \delta(t - t' - \overline{ab}/v_{soil} - \overline{bc}/v_{air})}{2\pi(\sqrt{\varepsilon_{soil}}\overline{ab} + \overline{bc})} \quad (7)$$

where \overline{ab} and \overline{bc} are the distances between a, b , and b, c respectively.

Based on the Born approximation presented in (5), we build the inversion scheme by sampling the measured E_x fields at each receiver and discretizing the object function. Including observation noise, the resulting model from (5) can be put into a matrix form as

$$y = Ax + n \quad (8)$$

where $y = [y_1^T | y_2^T | \cdots | y_N^T]^T$, and y_i is a column vector containing the sampled time signal collected by the i th receiver. The area of interest is discretized into M cells and x denotes a column vector which represents the object function in the area of interest. The matrix A is the discretized integral of (5), and n represents noise due to discretization or measurement error.

4 Regularization Method

The matrix A is both ill-conditioned, and based only on an approximation of the true physics. In order to obtain accurate reconstructions, we use regularization techniques to introduce additional information in the reconstruction algorithms. Since the object function is piecewise smooth and the mine's relative permittivity is larger than that of the soil, we use two different regularization functionals to encourage smoothness and positivity of the reconstructed object function. With the conventional Tikhonov regularization, the reconstruction problem minimizes the following objective:

$$\hat{x} = \arg \min_x \|y - Ax\|_2^2 + \lambda_1 \|Kx\|_2^2 + \lambda_2 \|Dx\|_2^2 \quad (9)$$

where D is the spatial gradient operator and K is defined by

$$[Kx]_i = \begin{cases} x_i & \text{if } x_i < 0 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

which penalizes negative values of the reconstructed object function.

To accurately locate a buried object, precise edge information is desired. However, with the conventional Tikhonov regularization (8), the edges in the reconstructed mine profiles tend to be blurred. As an alternative, we propose to use total variation regularization, which has better edge-preserving properties. In total variation regularization, the original L_2 norm on the piecewise smoothness penalty term is replaced with an L_1 norm, which penalizes large jumps less than the L_2 norm and thus allows sharp edges to form in reconstructed object profiles. The total variation based regularization objective is:

$$\begin{aligned} \hat{x} = & \arg \min_x \|y - Ax\|_2^2 + \lambda_1 \|Kx\|_2^2 \\ & + \lambda_2 \sum_{i=1}^M \sqrt{\alpha [D_x x]_i^2 + (2 - \alpha) [D_y x]_i^2} \end{aligned} \quad (11)$$

where D_x and D_y denote the first derivative operators in the horizontal and vertical directions respectively, and α controls the orientation preference of the smoothness regularization term. When $\alpha = 1$, there is no orientation preference. A regularization with orientation preference is desirable because the limited viewing geometry results in higher resolution in the vertical direction than in the horizontal direction. Thus, the problem is more “ill-posed” in the horizontal direction, and needs stronger regularization in that direction.

Equation (11) can be solved iteratively based on “half-quadratic” approximations[9], as follows. Let $x^{(k)}$ denotes the solution in the k th iteration step. Let

$$G^{(k)} = \text{diag} \left[\frac{1}{\sqrt{\alpha [D_x x^{(k)}]_i^2 + (2 - \alpha) [D_y x^{(k)}]_i^2 + \beta^2}} \right] \quad (12)$$

where β is a small smoothing parameter, which controls the the closeness to the true L_1 norm solution and the convergence speed. Let $P^{(k)} = \text{diag} [p_i^{(k)}]$, where $p_i^{(k)} = \begin{cases} 1 & \text{if } x_i^{(k)} < 0 \\ 0 & \text{otherwise} \end{cases}$. Then, the next solution $x^{(k+1)}$ is obtained by solving the following linear equation:

$$\begin{aligned} \left[A^T A + \lambda_1 P^{(k)} + \frac{\lambda_2}{2} \alpha D_x^T G^{(k)} D_x \right. \\ \left. + \frac{\lambda_2}{2} (2 - \alpha) D_y^T G^{(k)} D_y \right] x^{(k+1)} = A^T y \end{aligned} \quad (13)$$

5 Simulation Results

We use an accurate electromagnetic simulation software package, which uses the FDTD algorithm[8], to generate data. For the simulation presented in this paper, a flat air/soil interface is used; the soil was modeled as lossless, non-dispersive, and has uniform relative permittivity ($\epsilon_{soil} = 2.6$). A plastic mine was buried in the soil, with relative permittivity of 2.9. Both the soil and the mine's relative permeabilities are 1. The mine's upper surface is at a depth of 10cm below ground, and the mine is 6 cm thick, with a diameter

of 10cm. The region of interest is a 20cm by 20cm square, which is 5cm below the ground surface(see Fig. 1).

In the computation experiments, the full nonlinear simulation code is used to generate collected signals for two conditions: the case where the mine is present, and the case where only soil is present. The signals collected when no mine is present consists entirely of the ground reflection; these signals are used to subtract the ground reflection from the signals collected when the mine is present, leaving primarily the signals due to backscattering from the mine. Fig. 3 shows the signal collected by the middle receiver(after subtracting

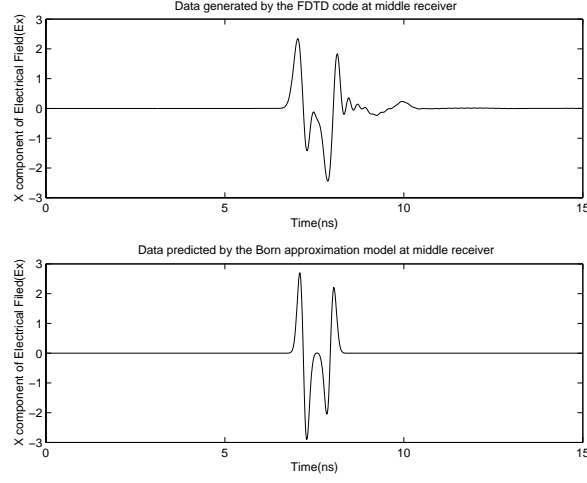


Figure 3: Data collected by the middle receiver

the ground reflection signal) and the signal predicted by the discrete Born approximation model. Fig. 4 and Fig. 5 illustrate how the data collected by different receivers decays with the receiver's offset from the central axis. To obtain the measured data, we corrupted the simulation outputs with 10dB of additional white Gaussian noise. The resulting signals were processed using the algorithms of the previous section to obtain the buried object profiles.

The reconstruction results are shown in Fig. 6. The total variation method with no orientation preference($\alpha = 1$) produces visually better results than the Tikhonov regularization, but lacks horizontal resolution. The total variation method with orientation preference($\alpha = 1.8$) gives the best reconstruction result, which significantly improves the horizontal resolution compared to the other two methods.

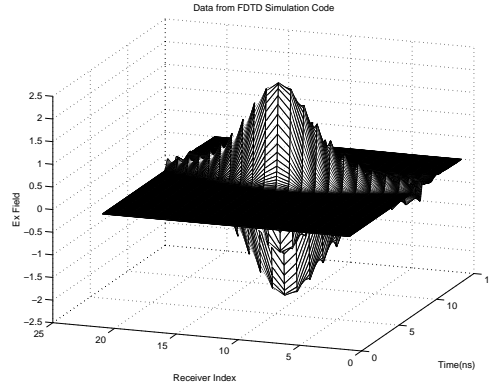


Figure 4: Data generated by FDTD code

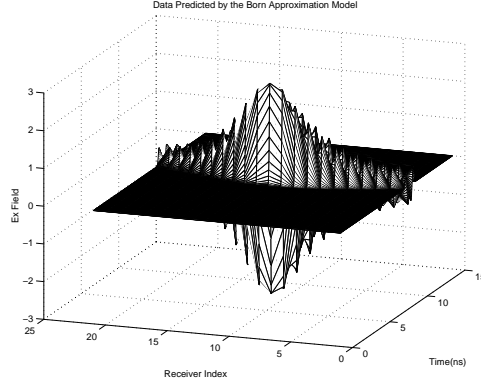


Figure 5: Data predicted by Born approximation model

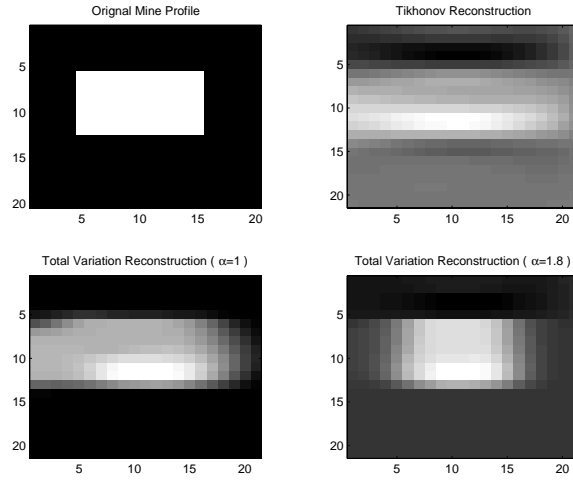


Figure 6: Original mine profile and reconstructions

6 Conclusion

We have developed an efficient reconstruction algorithm for weak-contrast buried objects based on the Born approximation model and edge-preserving regularization techniques. We have illustrated the advantages of this algorithm for an example including buried plastic mines imaged using ground-penetrating radar arrays.

The particular model used in this paper made important simplifying assumptions such as a flat air/soil interface and lossless, non-dispersive soil propagation. Our current research directions are exploring alternatives for incorporation of these effects into the reconstruction algorithms.

7 Acknowledgments

This work is partially supported by ODDR&E under MURI grants ARO DAAG55-97-1-0013 and AFOSR F49620-96-1-0028. We would also like to thank Professor L. B. Felsen for helpful discussions and the Center for Electromagnetic Research at Northeastern University for their assistance with the electromagnetic simulations.

References

- [1] T. Melamed, E. Heyman, and L. B. Felsen, “Local spectral analysis of short-pulse-excited scattering from weakly inhomogeneous media”, Technical Report, September, 1997.
- [2] A. Delaney, Y. Bresler, “Globally Convergent Edge-Preserving Regularized Reconstruction: An Application to Limited-Angle Tomography”, *IEEE Transactions on Image Processing*, vol 7, no. 2, February 1998.
- [3] R. A. Weisenseel, W. C. Karl, D. A. Cantañón, “Statistical fusion of GPR and EMI data”, *Proc. SPIE*, V 3710, SPIE, Orlando, 1999.
- [4] T.R. Witten, “Present state-of-the-art in ground penetrating radars for mine detection”, *Proc. SPIE*, V 3392, p14-22, SPIE, Orlando, 1998.
- [5] A. J. Devaney, “Geophysical diffraction tomography”, *IEEE Trans. Geosci. Remote Sensing*, vol GRS-24, pp. 654-662, Sept. 1986.
- [6] T. Melamed, Y. Ehrlich, E. Heyman, “Short-pulse inversion of inhomogeneous media: a time-domain diffraction tomography”, *Inverse Problems* vol 12, pp. 977-993, 1996.
- [7] P. Chaturvedi, R. G. Plumb, “Electromagnetic Imaging of Underground Targets Using Constrained Optimization”, *IEEE Trans. on Geosci. Remote Sensing*, vol 33, No. 3, May 1995.
- [8] W. H. Weedon, C. M. Rapport, “A General Method for FDTD Modeling of Wave Propagation in Arbitrary Frequency-Dispersive Media, *IEEE Transactions on Antennas and Propagation*, Vol 45, No. 3, March 1997.
- [9] C. R. Vogel and M. E. Oman, “Fast, robust total variation-based reconstruction of noisy, blurred images”. *IEEE Transactions on Image Processing*, 7(7): 813:824, July 1998.
- [10] J. D. Jackson, *Classical Electrodynamics*, John Wiley & Sons, Inc., 1962

Mine Detection Under Rough Ground Surfaces Using 2-D FDTD Modeling and Hypothesis Testing

He Zhan, Carey Rappaport, Magda El-Shenawee and Eric Miler

Northeastern University, 235 Forsyth Building, Boston MA 02115

(617) 373-4874

hzhhan@ece.neu.edu

ABSTRACT

This paper attempts to investigate hypothesis test performance in mine detection, based on numerically simulating the wideband scattering of experimentally measured GPR signals by realistic, dispersive soil interfaces. We employ a 2D finite difference time domain (FDTD) method to analyze the delay and amplitude characteristics of ground-scattered waves as a function of roughness parameters. In addition, we apply binary hypothesis tests to the signals obtained using physics-based signal processing techniques to investigate the presence of the target at certain depth. We quantify the detection performance in terms of the spatial distribution of transmitter and receiver.

INTRODUCTION

Detecting buried dielectric targets—such as nonmetallic antipersonnel mines—with ground penetrating radar (GPR) is important and difficult, because the dielectric constant of nonmetallic mine targets are similar to those of the surrounding soil, and because their size is comparable to the thickness of soil layer above it. In addition, the soil dielectric constant may not be well characterized, and the ground surface will usually be rough, often with roughness of the order of the target burial depth. We simulate the effects of random rough ground surface on the GPR signal using multiple Monte Carlo runs of 2D TM FDTD calculations. A typical bistatic geometry (Fig. 1a) is used along with a measured excitation signal generated by the Geo-Centers TEMR GPR antenna (Fig. 1b).

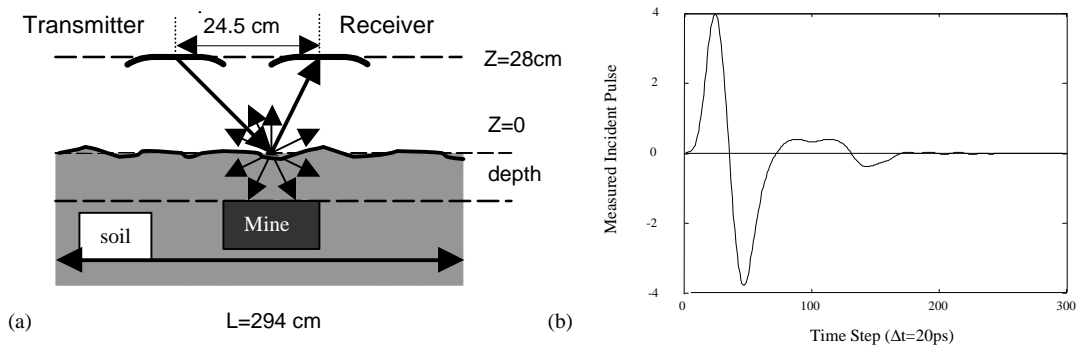


Figure 1: (a) Rough surface computational geometry; (b) incident measured waveform

In the FDTD code, the time step is $\Delta t = 20$ ps and $\Delta = 1.22$ cm. Simulations are done for 500 surface realizations with and without a mine target at various depths below the nominal surface level for a variety of roughness statistics. The probability density function for the height and the surface profiles spectrum are both assumed Gaussian [1,2]. The soil model is Puerto Rican clay loam with 10% moisture and 1.4g/cc density, with average dielectric constant $\epsilon' = 6.2$ [3].

SIGNAL AND CLUTTER SEPARATION MODELING AND PROCESSING

The background averaging procedure is used to remove the ground clutter signal, when the target is small, shallow, and of low contrast, [2,4]. The cross-correlation function and relative scaling factor are introduced. The cross-correlation function between reference signal ‘ f ’ and any realization signal ‘ i ’ is:

$$C_{fi}(m) = \sum_{n=1}^{N-|m|} S_f(n) S_i(n+m) / \left(\sqrt{\sum_{n=1}^N |S_f(n)|^2} \sqrt{\sum_{n=1}^n |S_i(n)|^2} \right) \quad \text{for } m > 0 \quad (1a)$$

$$C_{fi}(m) = C_{if}(-m) \quad \text{for } m < 0 \quad (1b)$$

The relative scaling is defined as:

$$A_i = \sqrt{\sum_{n=1}^N |S_i(n)|^2} / \max(C_{fi}) \quad (1c)$$

Where $i = 1, 2, \dots, M$ is the rough surface realization index, M is the size of Monte Carlo sample, and N is the total number of time steps. Compared to the scaling definition that is relative to the signal energy, the background averaging process will end up with a better output signals using the definition in (1c). The nominal background signal is found by aligning the time index of its maximum cross-correlation $m_{\max i}$, scaling by A_i , and averaging. To reduce ground clutter, this nominal background signal is again shifted back by $m_{\max i}$, scaled by A_i , and subtracted from each S_i signal. These background-subtracted signals are once again realigned to the expected target position [2,4]. Fig. 2 shows the average of 500 aligned background-subtracted signals (solid line) and the ± 1.0 standard deviation confidence interval (dotted lines).

SIMPLE BINARY HYPOTHESIS TESTING

We simplify the mine detection problem to be: whether there is mine at certain depth with acknowledgement of the roughness of the surface. Each of the two answers corresponds to a hypothesis:

H_1 corresponds to the presence of a target (*i.e.* mine presence)

H_0 corresponds to no target (*i.e.* no mine)

Applying the likelihood ratio test, which is derived to be [5]:

$$\frac{1}{2} (\mathbf{R}^T - \mathbf{m}_0^T) \mathbf{Q}_0 (\mathbf{R} - \mathbf{m}_0) - \frac{1}{2} (\mathbf{R}^T - \mathbf{m}_1^T) \mathbf{Q}_1 (\mathbf{R} - \mathbf{m}_1) \underset{H_0}{\overset{H_1}{<}} \text{threshold} \quad (2a)$$

where \mathbf{R} , \mathbf{m}_0 , and \mathbf{m}_1 are vectors with N elements, \mathbf{Q}_0 and \mathbf{Q}_1 are $N \times N$ covariance matrices. \mathbf{R} represents the individual background-subtracted (target) signals. \mathbf{m}_1 is the average of background-subtracted (target) signals shown in Fig. 2, and \mathbf{m}_0 is the average of unaligned no target (clutter) signals, *i.e.* the no-target signal. The covariance matrices \mathbf{Q}_0 and \mathbf{Q}_1 are the inverses of the diagonal matrices of standard deviations of the target and clutter signals. 100 out of 500 target signals and 100 out of 500 clutter signals are grouped as test signals. The mean signals \mathbf{m}_0 and \mathbf{m}_1 , and standard deviations are obtained from the remaining 400 target signals and 400 clutter signals. The likelihood ratio test is applied to each individual test signal associated with a set of trial thresholds.

Sequentially, the probability of *false alarm* P_F (i.e. we say the target is present when it is not) and the probability of detection *alarm* P_D (i.e. we say the target is present when it is) are obtained. The sample performances of the detections for several rough surfaces with Gaussian height S_c and correlation length L_h parameters, and depths are shown in Fig. 3.

As the depth of buried mine increases within some region and the surface becomes more flat, the performance improves. By relating signals associated with certain depth to several mean signals for different depths, it has been found that the best performance always occurs for the correct depth estimate. An example is shown in Fig. 4. The test signals with depth 8.5cm were related to the mean signals belonging to the same roughness group. Thus, the correct target depth will be obtained by testing sample signals with stored mean signals for a given ground surface roughness. Interestingly, the ROC curves for close but not exact depths are worse than those for much deeper or shallower depth estimates.

The previous results are based on the geometry shown in Fig. 1a, with the target centered beneath the transmitter-receiver (TR) pair. As the pair moves away from the mine, the test performance degrades. Also, the amplitude of the mean target signal will drop, as shown in Fig. 5. However, the combination of several TR views improves the detection performance. Although performance increases with the number of views, the combination of three TR positions, each spaced 10 spatial steps apart (i.e. $10 \times 1.22 \text{ cm} = 12.2 \text{ cm}$) to the left and right, appear to give excellent results. The improvements of using three and five bistatic TR pairs are shown in Fig. 6.

CONCLUSION

Using the physics-based signal processing, the background-subtracted and clutter signals have been analyzed. Applying binary hypothesis tests on these signals, the mine can be localized at certain depths even under rough ground surfaces. Furthermore, the detection performance is improved by combining multiple TR pair combinations.

ACKNOWLEDGEMENT

This work is supported by the Army Research Office grant No. DAAG55-97-0013.

REFERENCES

- [1] P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, (New York), 1963
- [2] M. El-Shenawee and C. Rappaport, "Quantifying the Effects of Different Rough Surface Statistics for Mine Detection Using the FDTD Technique," *Proc. SPIE*, Apr. 2000.
- [3] J. Hipp, "Soil Electromagnetic Parameters as Functions of Frequency, Soil Density and Soil Moisture," *Proceedings of the IEEE*, vol. 62, no. 1, Jan. 1974, pp. 98-103.
- [4] C. Rappaport and M. El-Shenawee, "Modeling GPR Signal Degradation from Random Rough Ground Surface," *IGARSS*, Honolulu, HI, July 2000.
- [5] H. Van Trees, *Detection, Estimation, and Modulation Theory*, Wiley (New York), 1968.

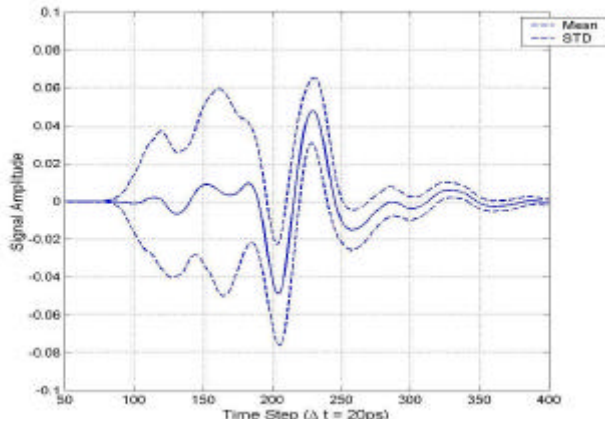


Figure 2: Average aligned background-subtracted signal with ± 1.0 standard deviation confidence interval for $\sigma_h = 1$ cm, $L_c = 10$ cm.

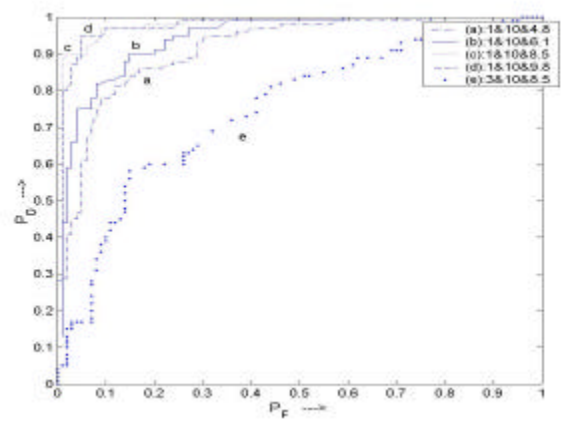


Figure 3: Receiver Operating Characteristic (ROC) curves, for $(\sigma_h, L_c, \text{depth}) = (1, 10, 4.8), (1, 10, 6.1), (1, 10, 8.5), (1, 10, 9.8),$ and $(3, 10, 8.5)$ cm.

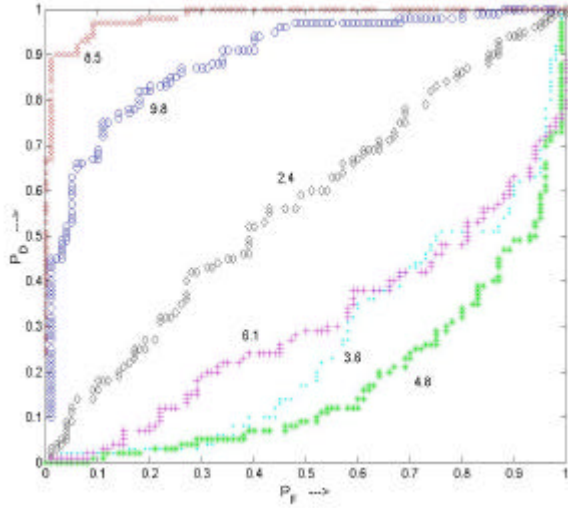


Figure 4: ROC curves, for $(\sigma_h, L_c) = (1, 10)$ cm, with target at depth 8.5 cm, using various depth test hypothesis.

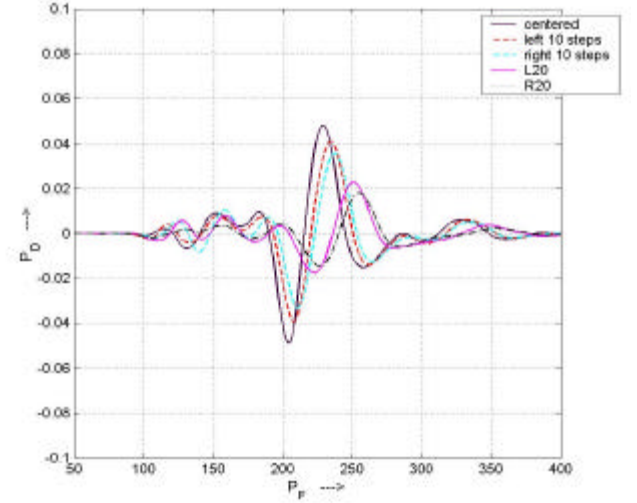


Figure 5: Average aligned background-subtracted signals for centered, ± 12.2 cm, and ± 24.4 cm TR pair positions.

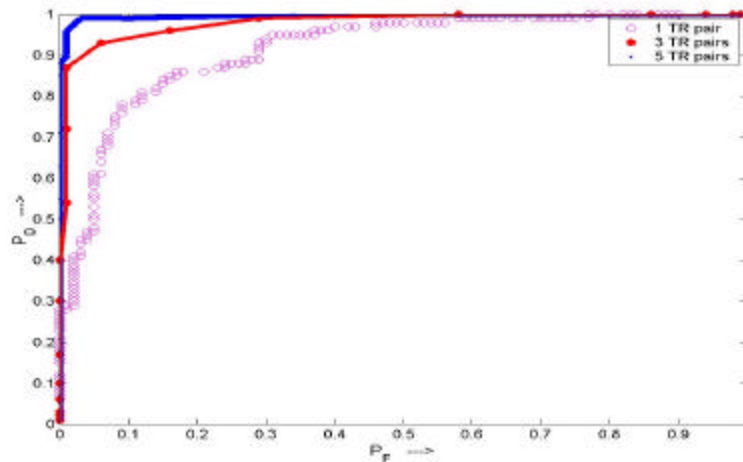


Figure 6: ROC curves for combinations of 1, 3 and 5 TR pairs



Northeastern University Humanitarian Demining MURI Efforts on Clutter Modeling and Inverse Scattering Methods for GPR

Carey Rappaport, Eric Miller, Magda El-Shenawee, & Misha Kilmer
Dept. Electrical and Computer Engineering
Northeastern University, Boston MA

This work was supported by an OSD MURI on Demining under
Grant DAAG55-97-1-0013

Clutter Modeling: Rough Ground Surface

Outline

- Necessity of modeling soil background environment in mine detection
- Frequency domain discrimination of non-metallic target shapes for smooth/rough ground
- Dispersion in the time domain
- Monte Carlo FDTD modeling of rough ground effects.

Numerically Modeling EM Wave Propagation in Soil to Aid Sensor Design

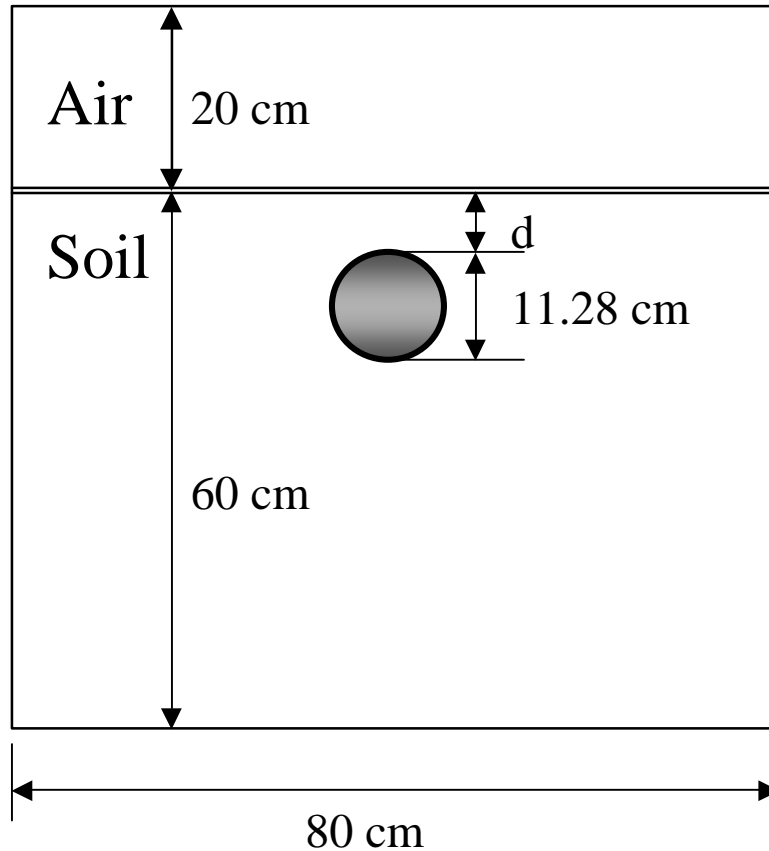
- Only with knowledge of the environment surrounding mines can efficient advanced sensors be developed.
- Soil environment is extremely varied and inhomogeneous.
- Propagation is quite different in different soils.
- Numerical modeling is the only viable means of testing candidate sensing concepts on the wide variety of conditions.
- Simple flexible models that capture the essential electrical characteristics are best.
- Visualization of wave propagation assists in sensor synthesis.

FDFD Model of Plane Wave Incidence on Various Buried Non-Metallic Target Shapes

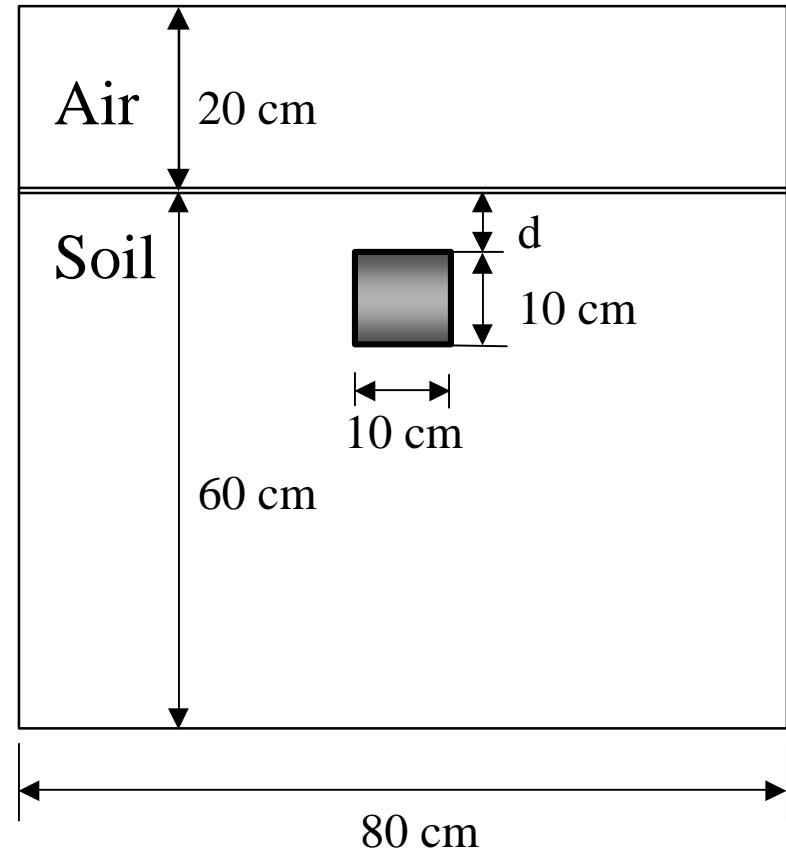
Sandy soil: $\epsilon_s = 2.5$, $\sigma_s = 0.01$

TNT mine: $\epsilon_m = 2.9$, $\sigma_m = 0.004$

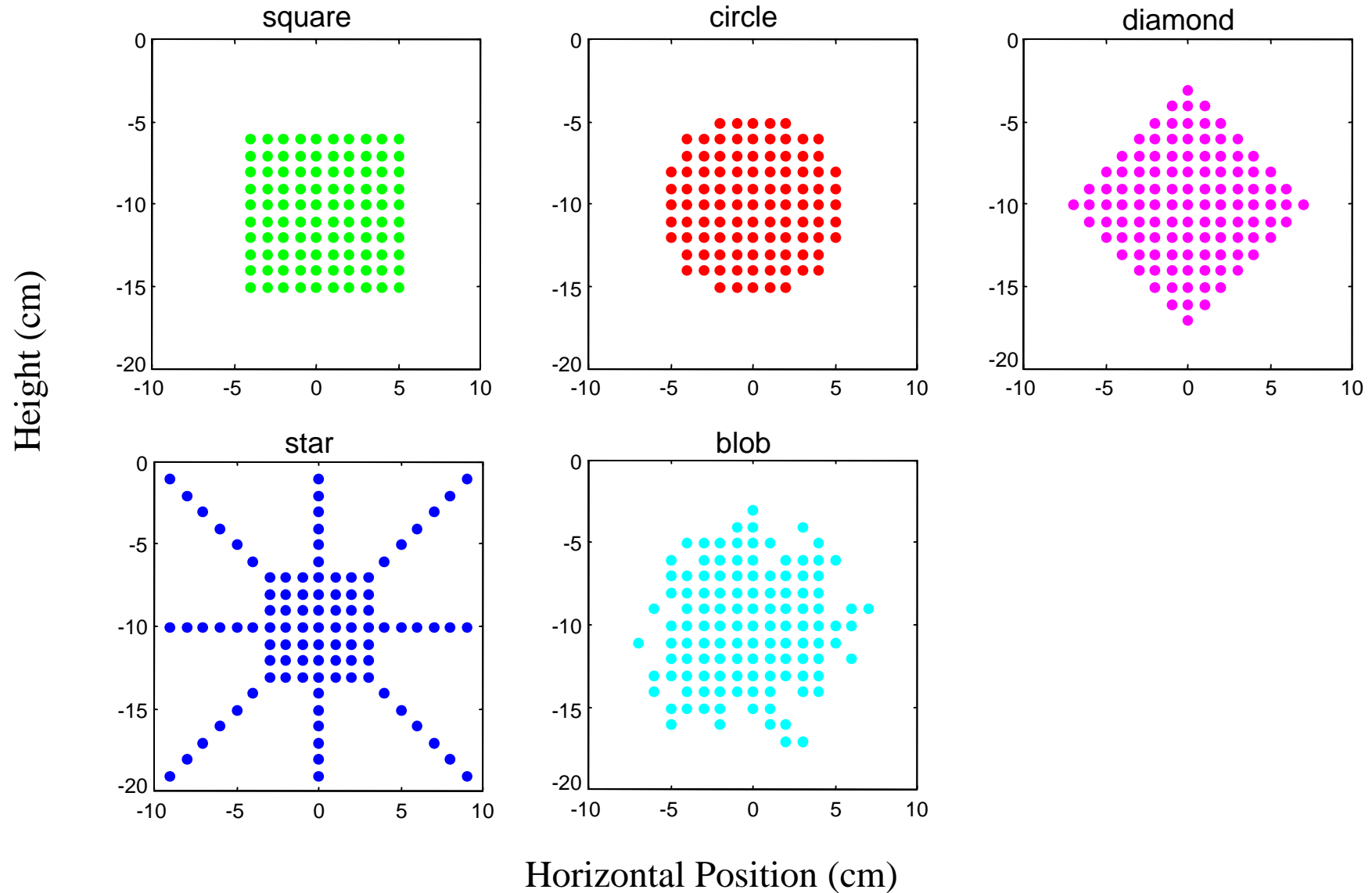
Circular Target



Square Target

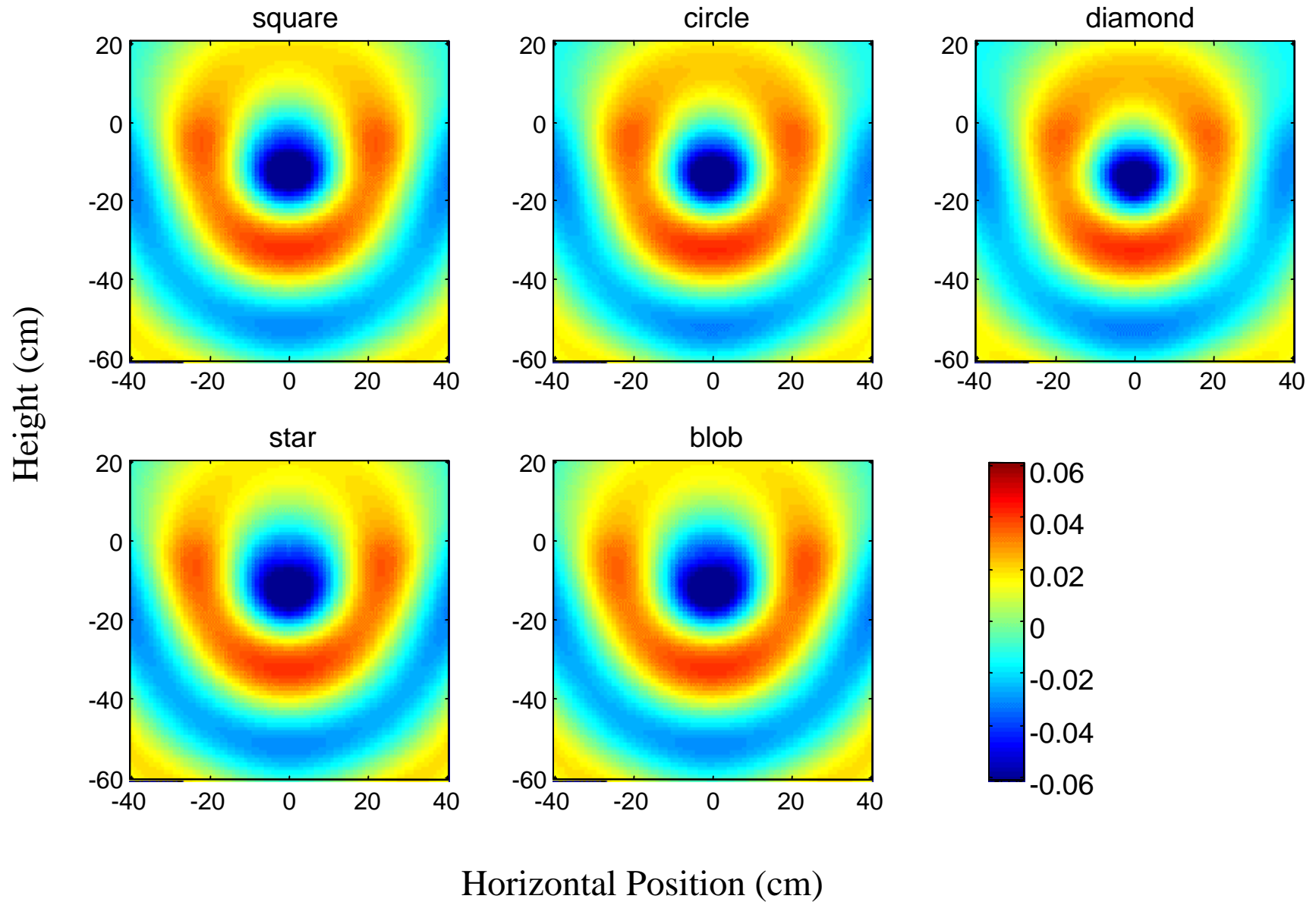


Different Buried Test Target Shapes



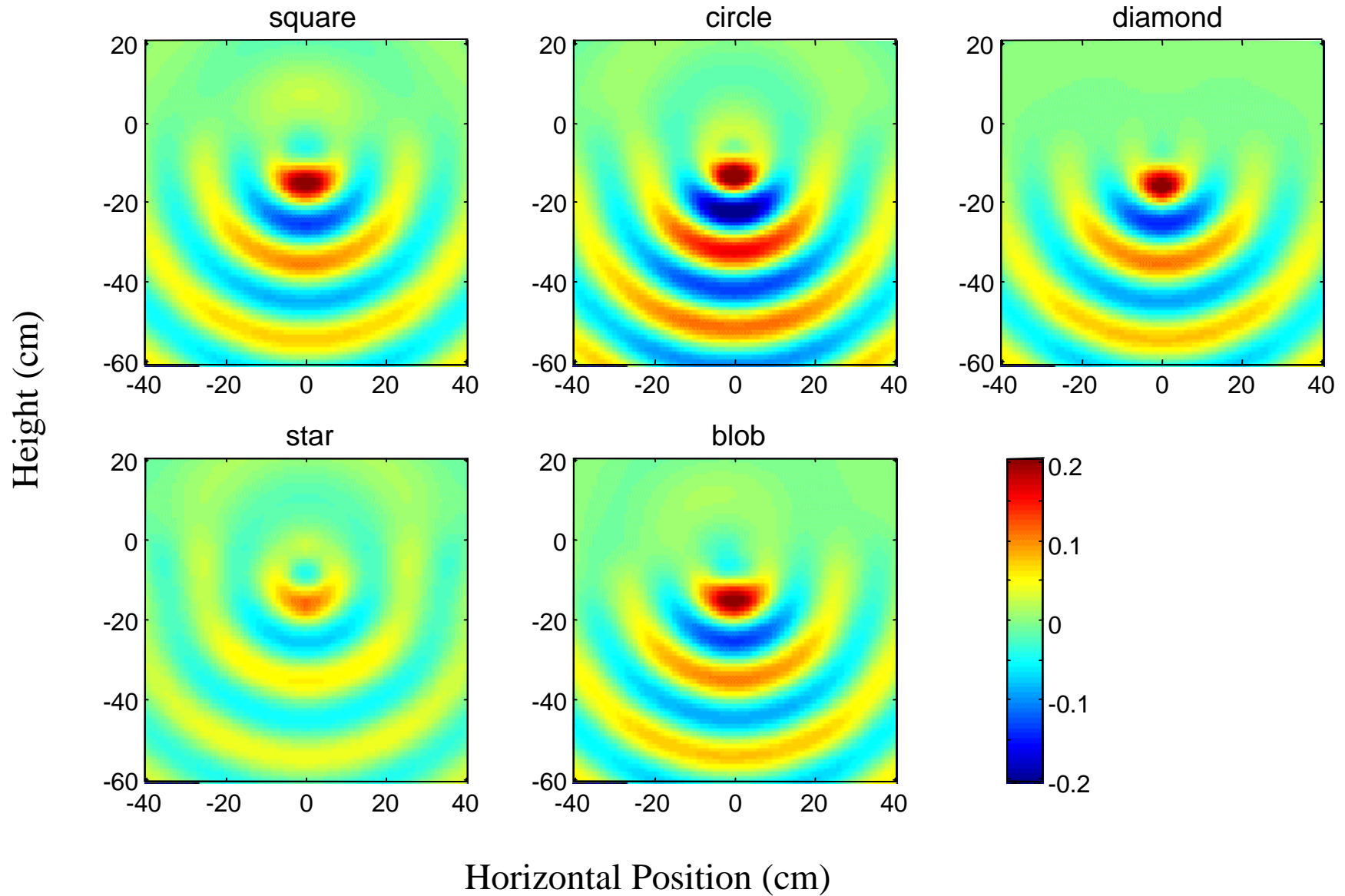
Scattered Field - Real Part

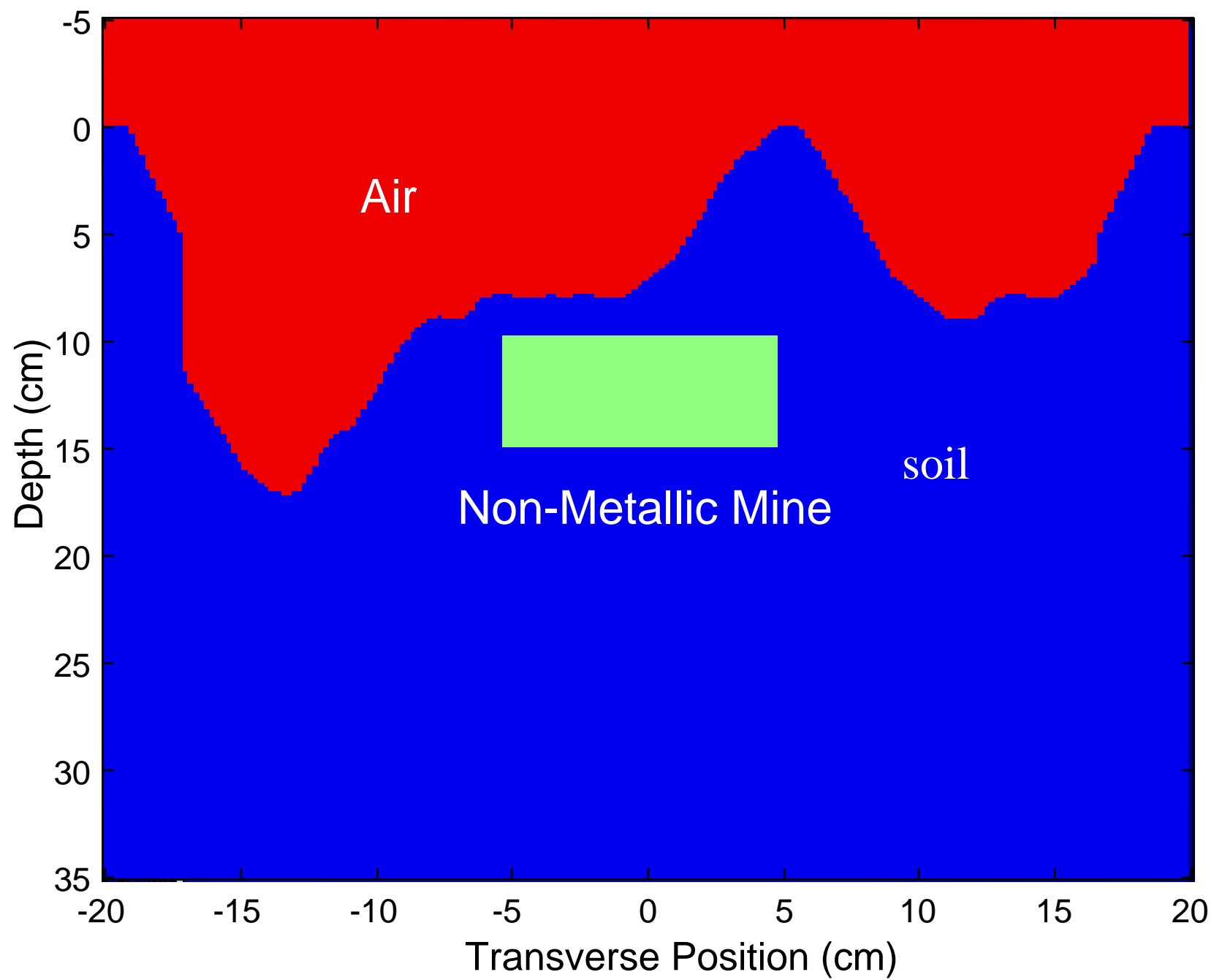
500 MHz, depth = 5 cm



Scattered Field - Real Part

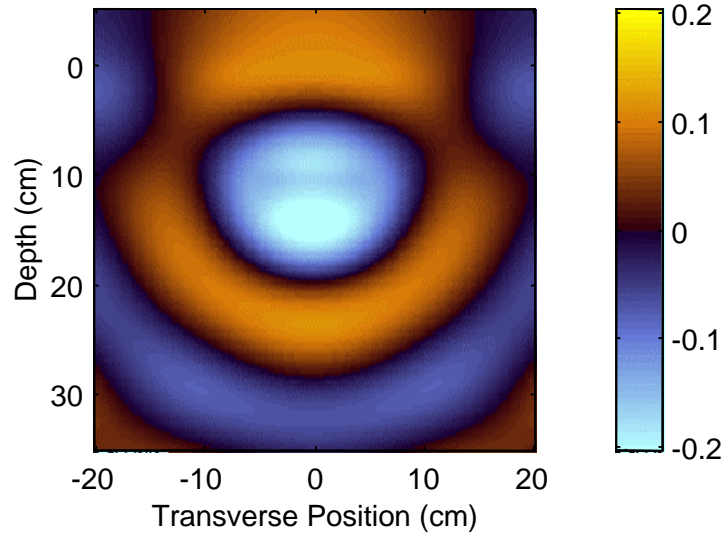
1000 MHz, depth = 5 cm



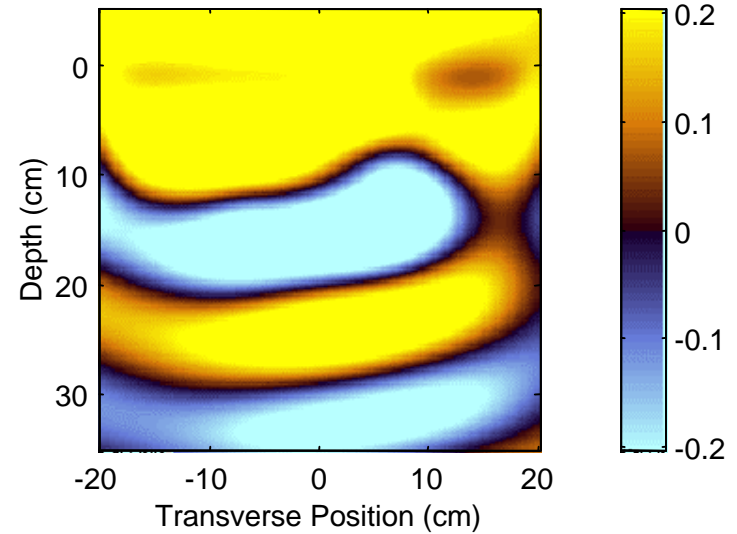


480 Mhz

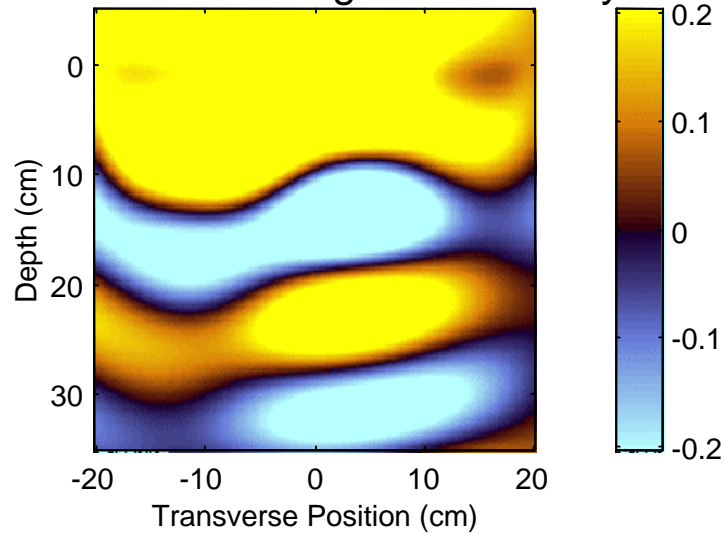
Mine scattered field: smooth surface



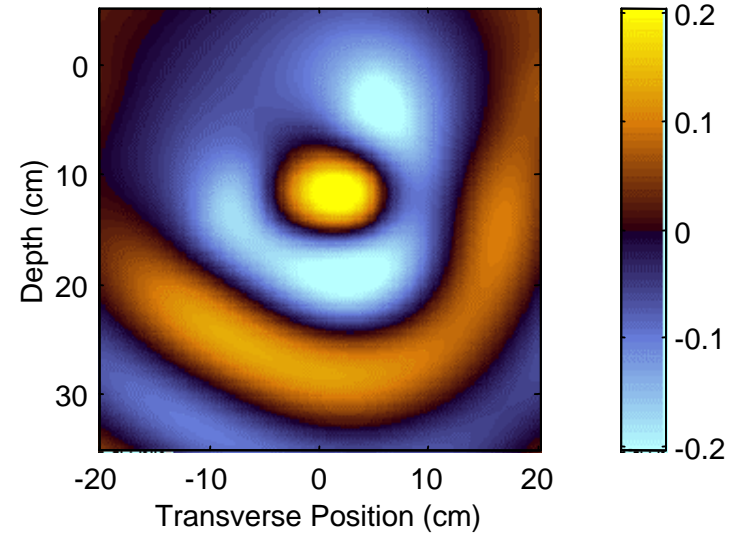
Scattered field: rough surface with mine



Scattered field: rough surface only

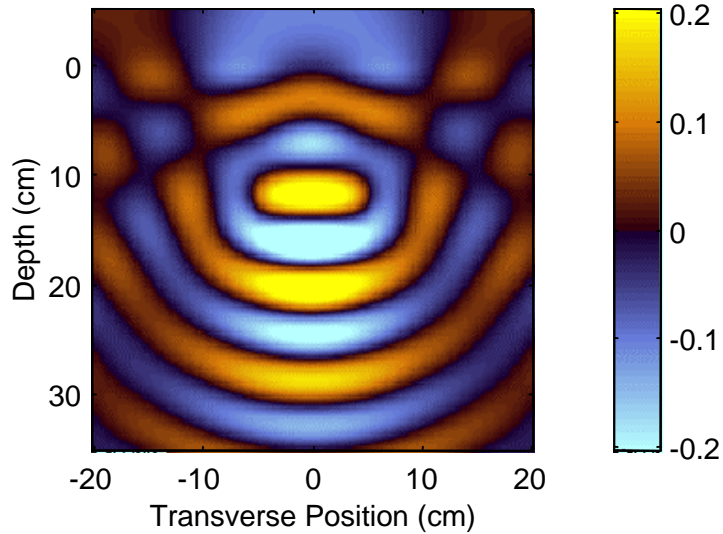


Mine scattered field: rough surface

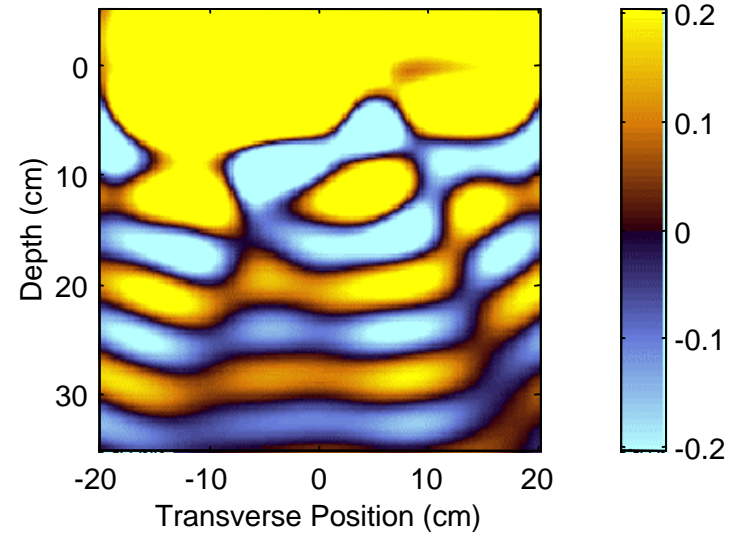


960 Mhz

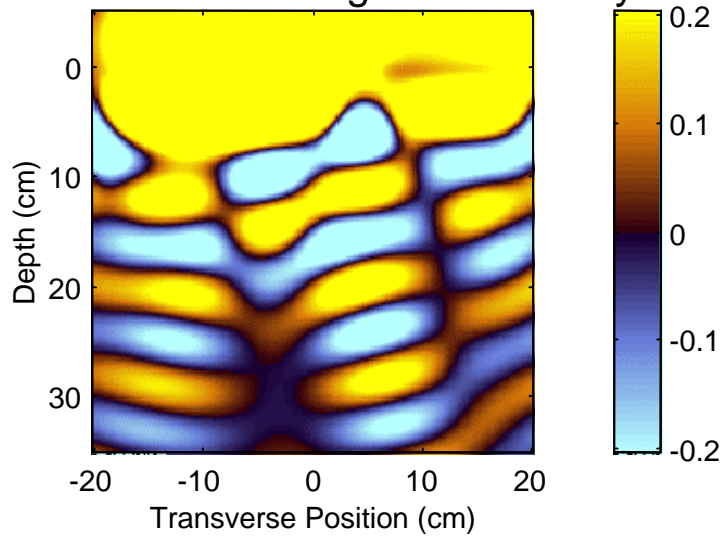
Mine scattered field: smooth surface



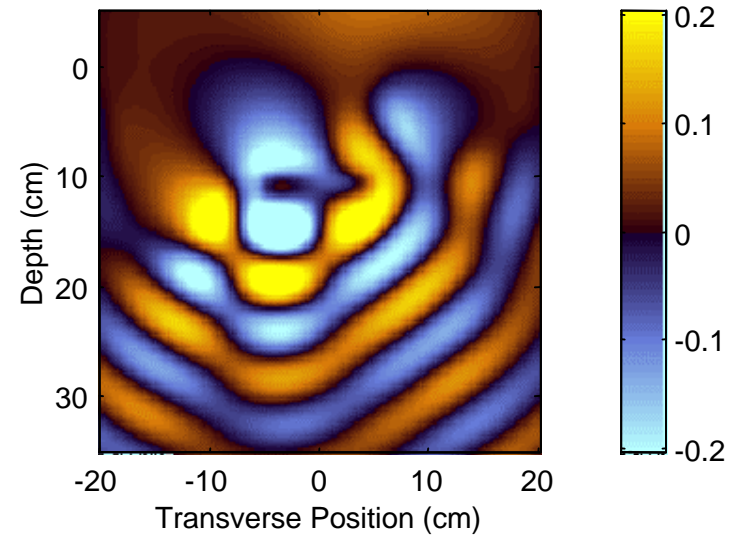
Scattered field: rough surface with mine



Scattered field: rough surface only

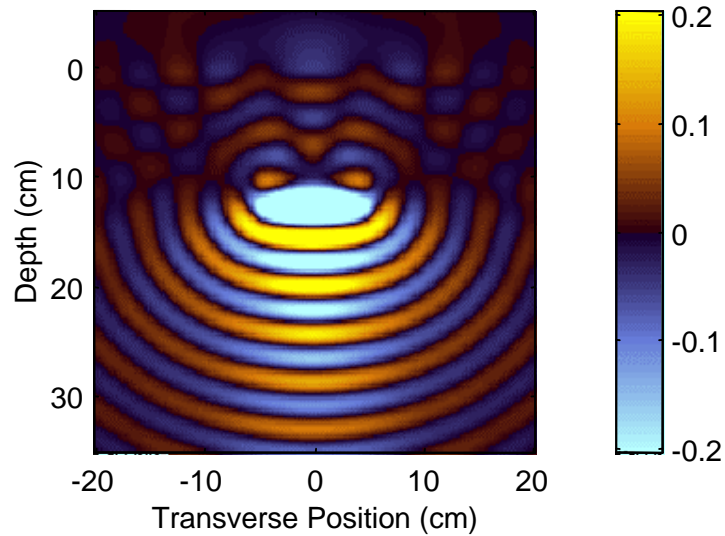


Mine scattered field: rough surface

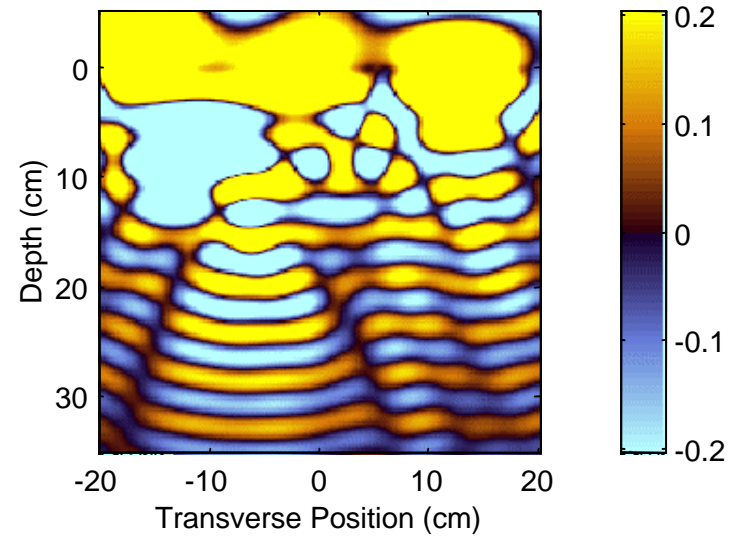


1920 Mhz

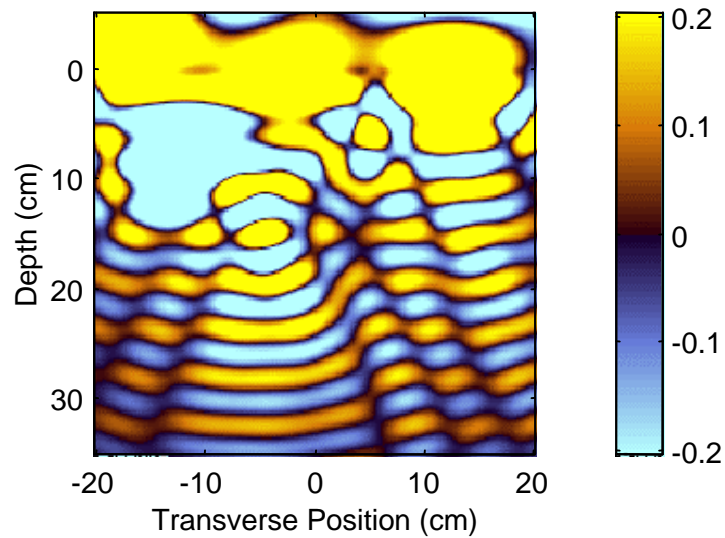
Mine scattered field: smooth surface



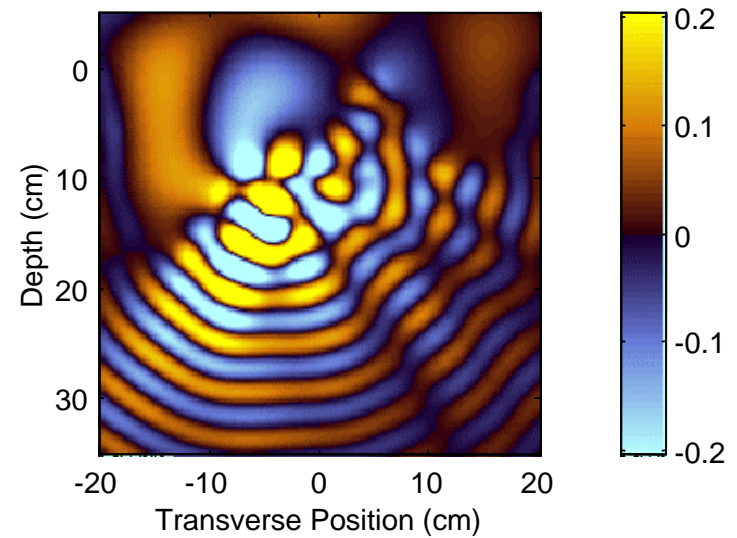
Scattered field: rough surface with mine



Scattered field: rough surface only

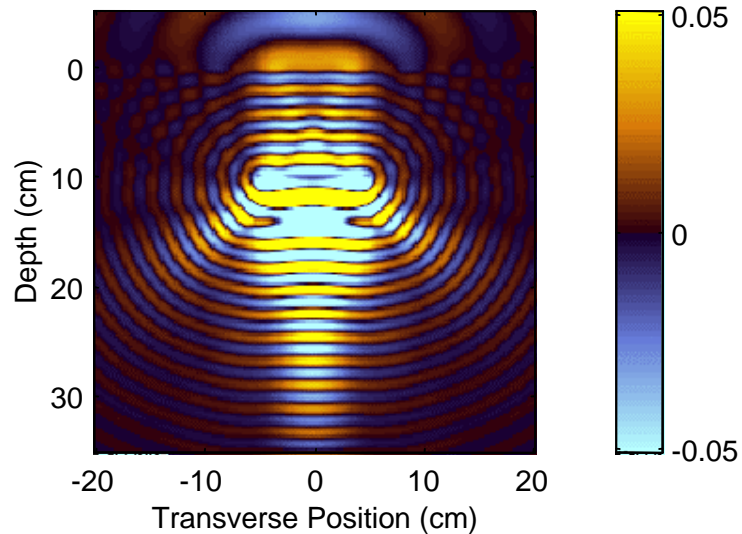


Mine scattered field: rough surface

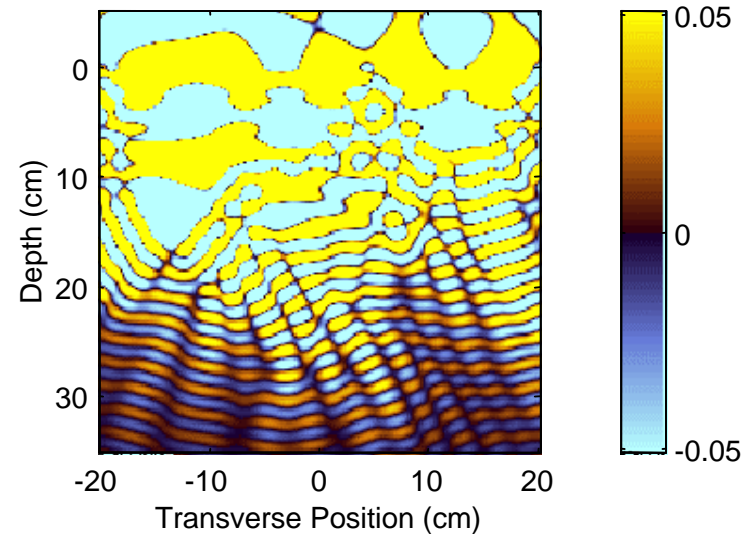


3840 Mhz

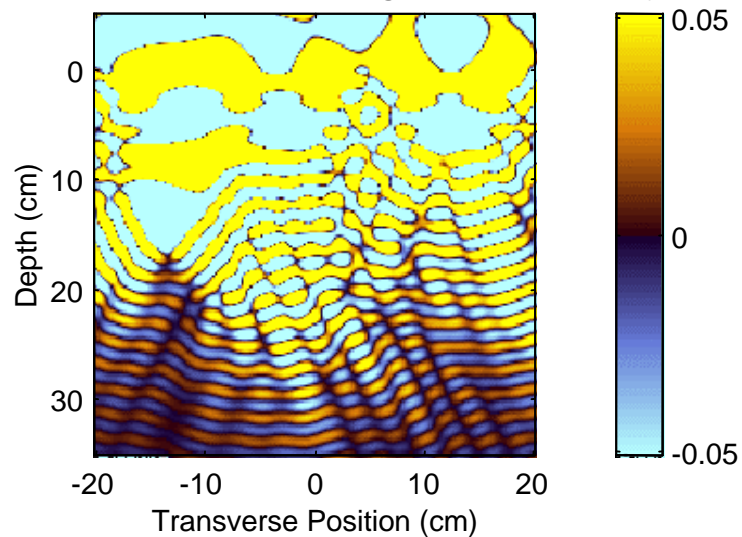
Mine scattered field: smooth surface



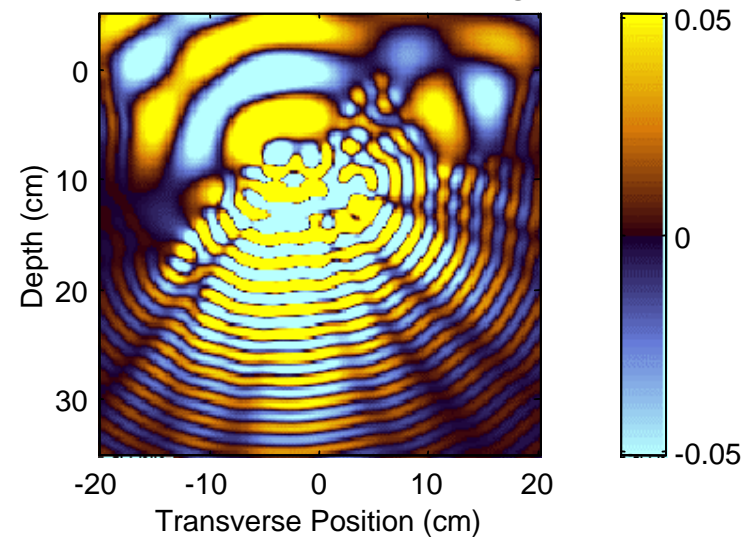
Scattered field: rough surface with mine

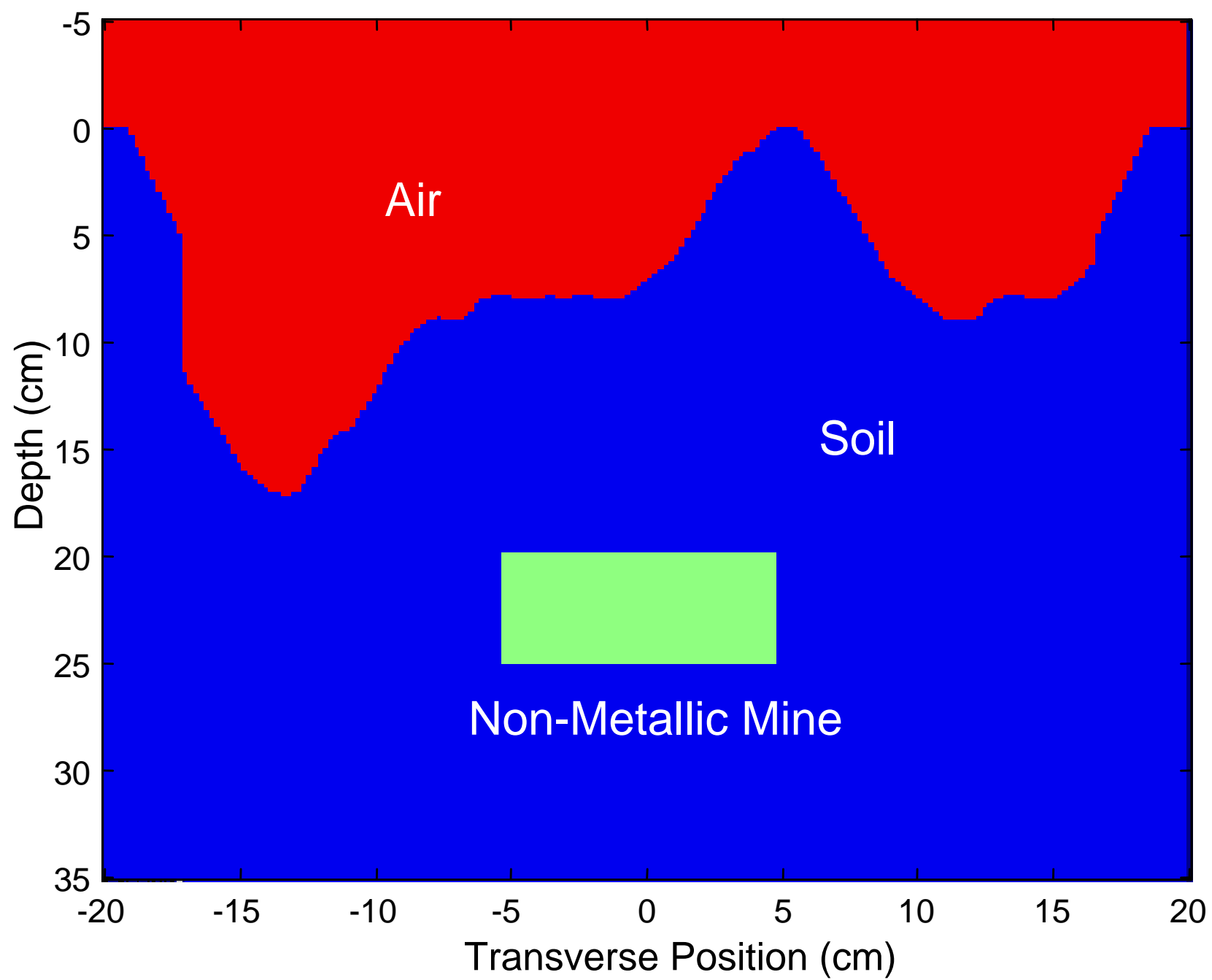


Scattered field: rough surface only



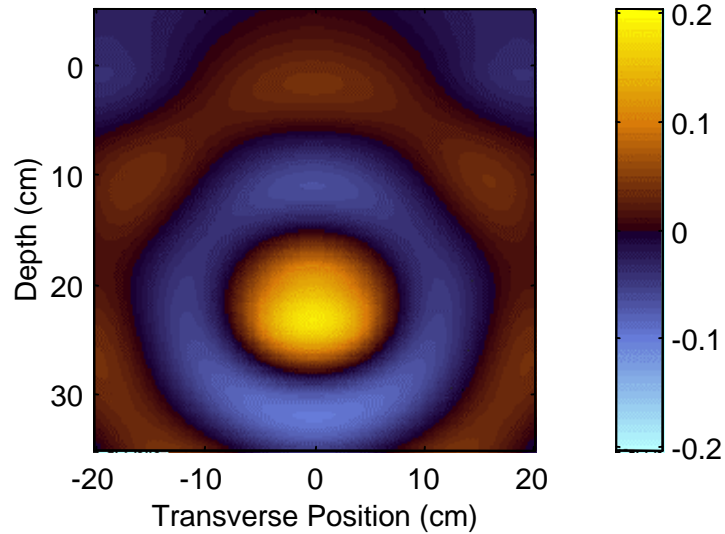
Mine scattered field: rough surface



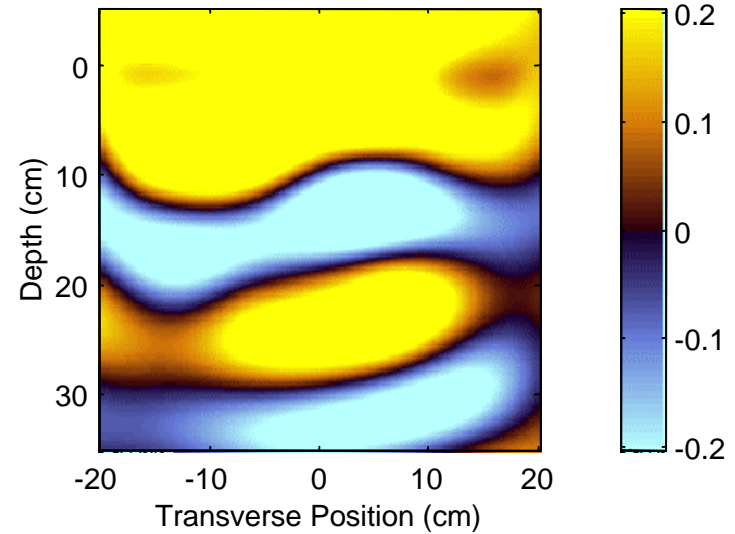


480 Mhz

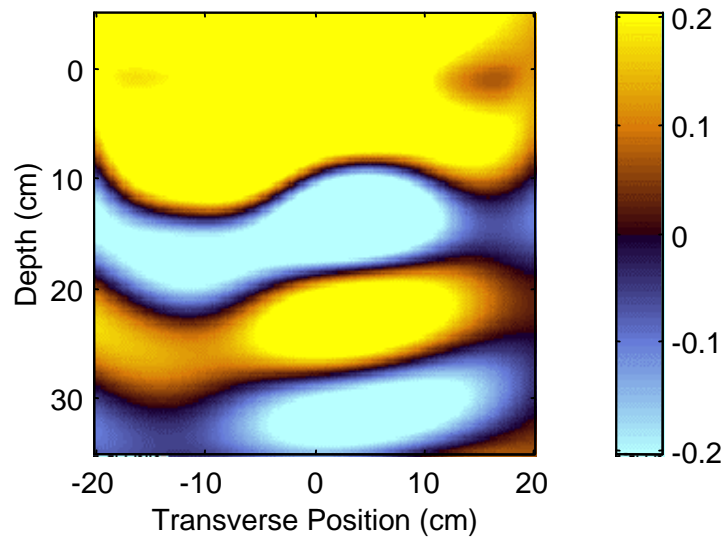
Scattered field from mine with smooth surface



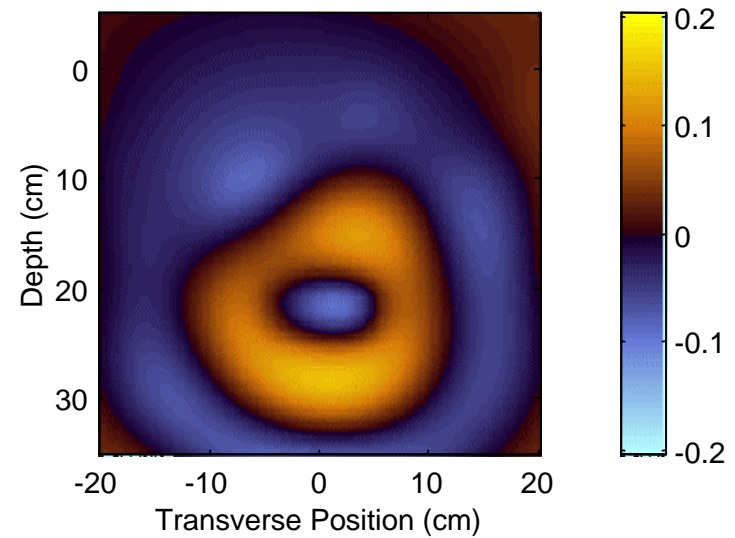
Scattered field from rough surface and mine



Scattered field from rough surface

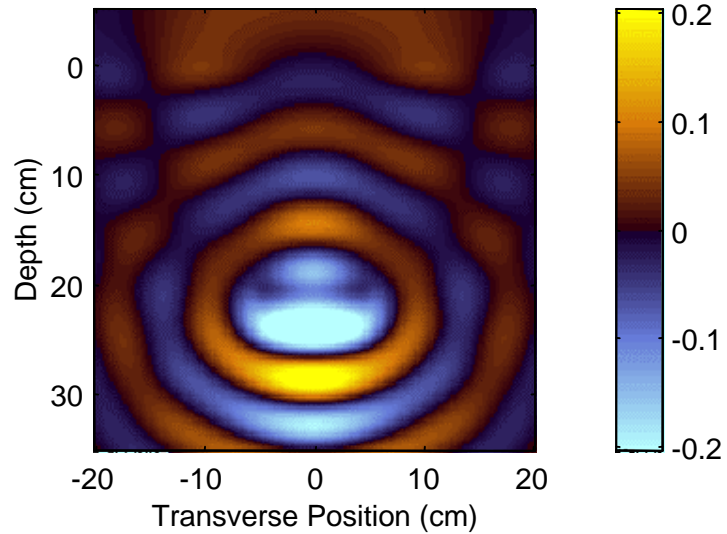


Scattered field from mine with rough surface

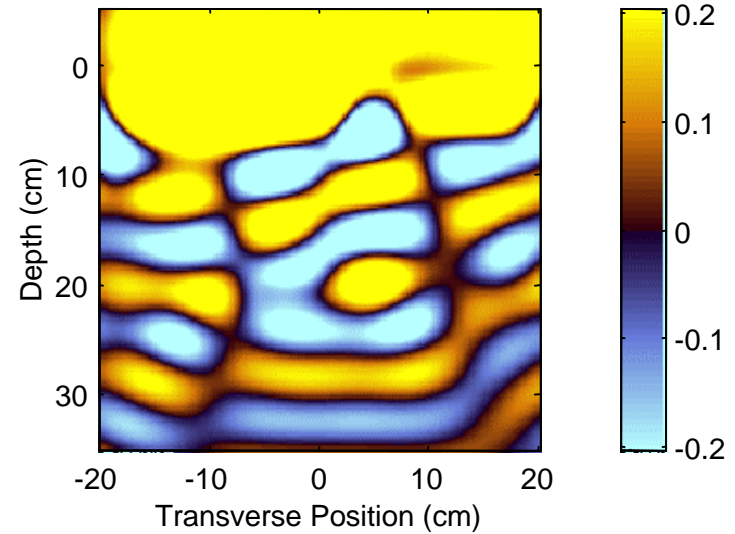


960 Mhz

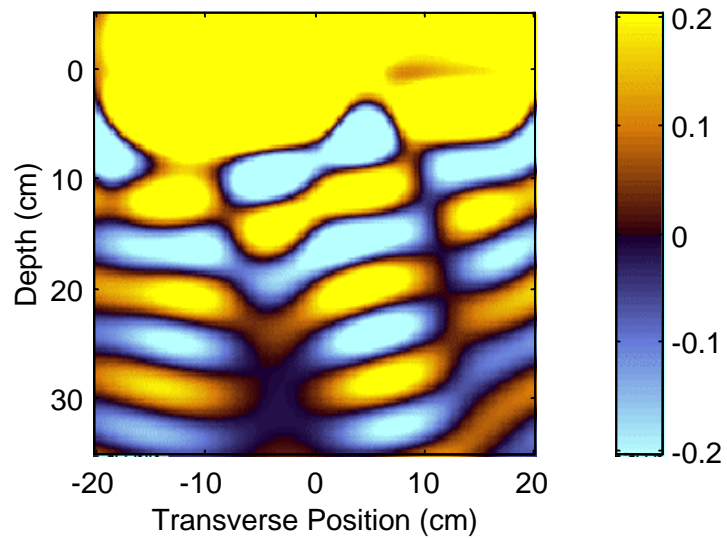
Scattered field from mine with smooth surface



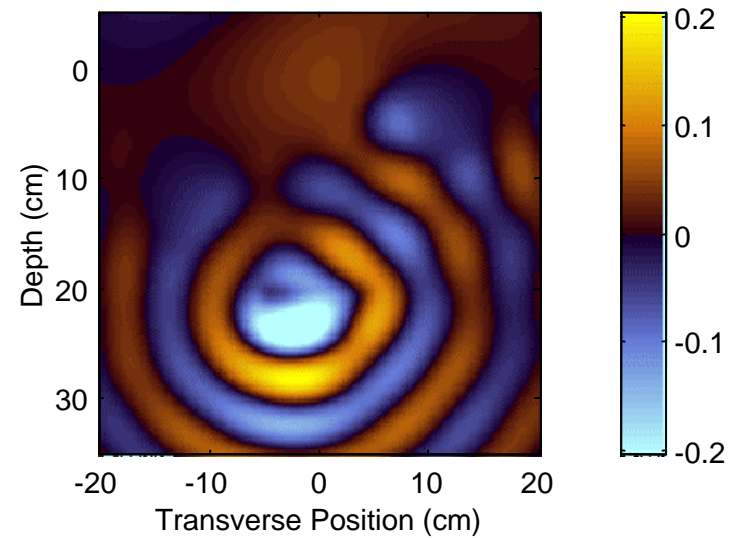
Scattered field from rough surface and mine



Scattered field from rough surface

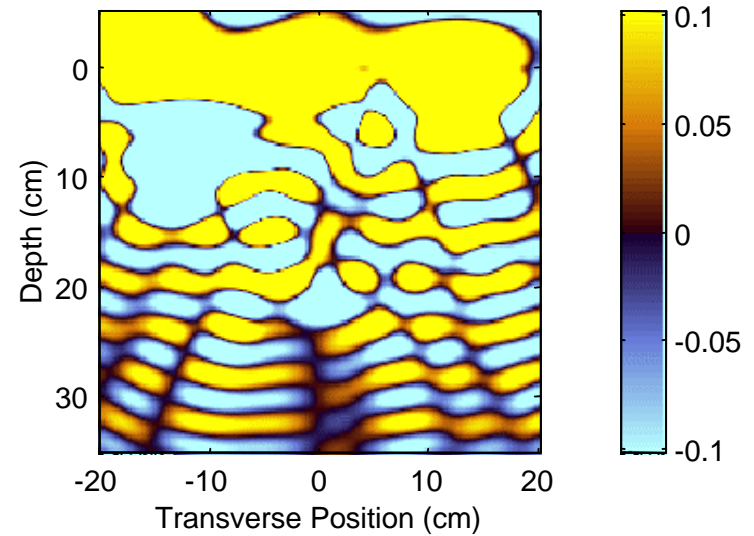
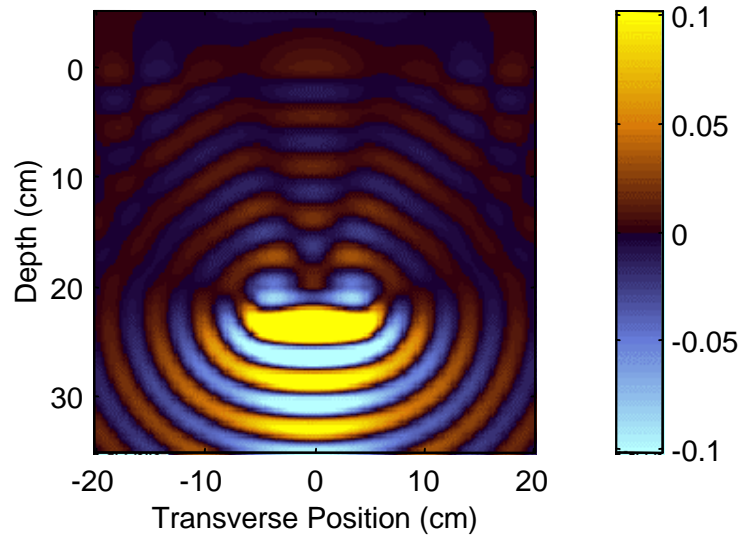


Scattered field from mine with rough surface

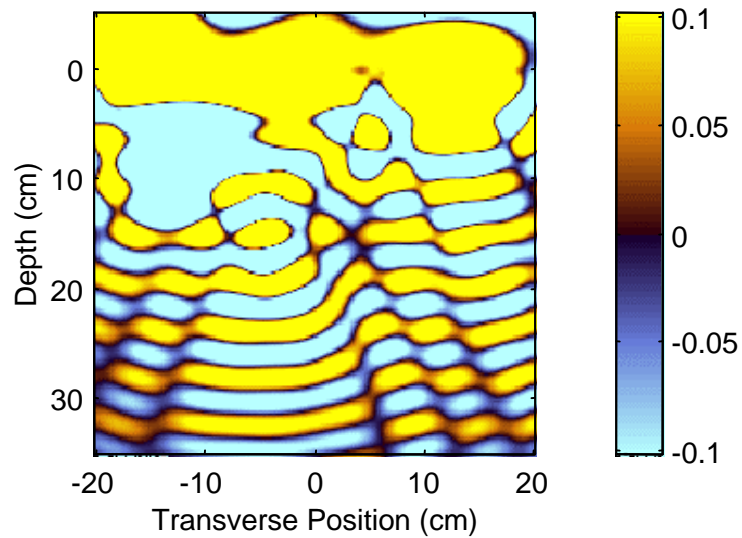


1920 Mhz

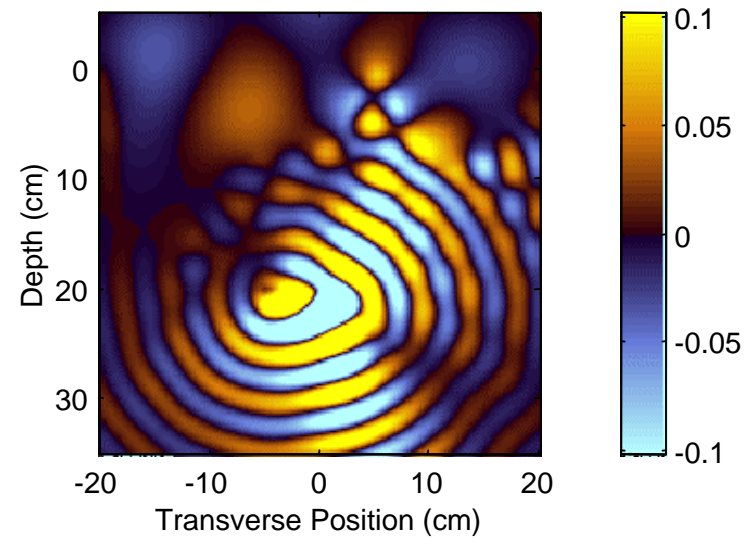
Scattered field from mine with smooth surface Scattered field from rough surface and mine



Scattered field from rough surface

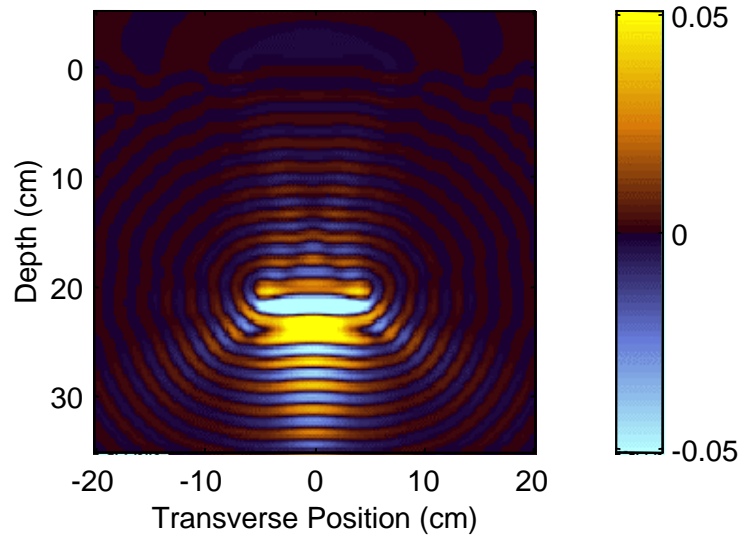


Scattered field from mine with rough surface

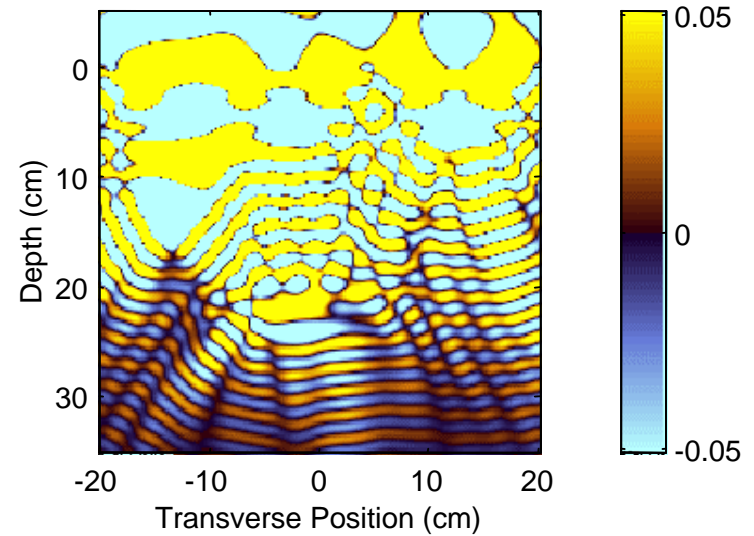


3840 Mhz

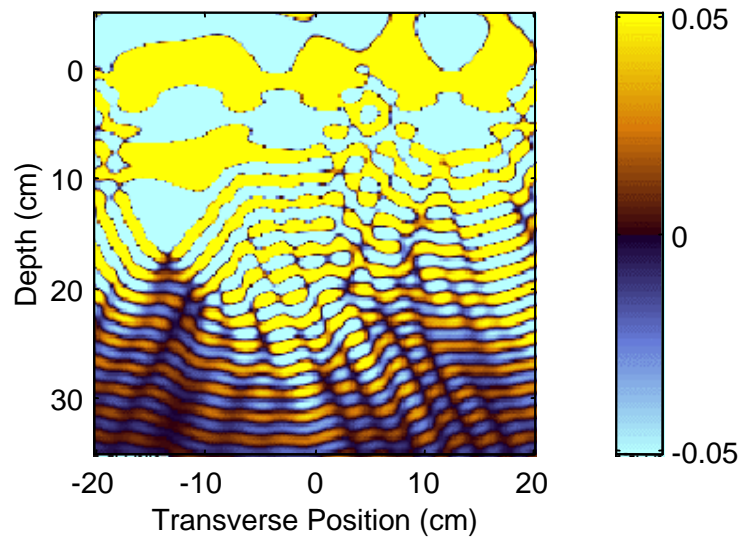
Scattered field from mine with smooth surface



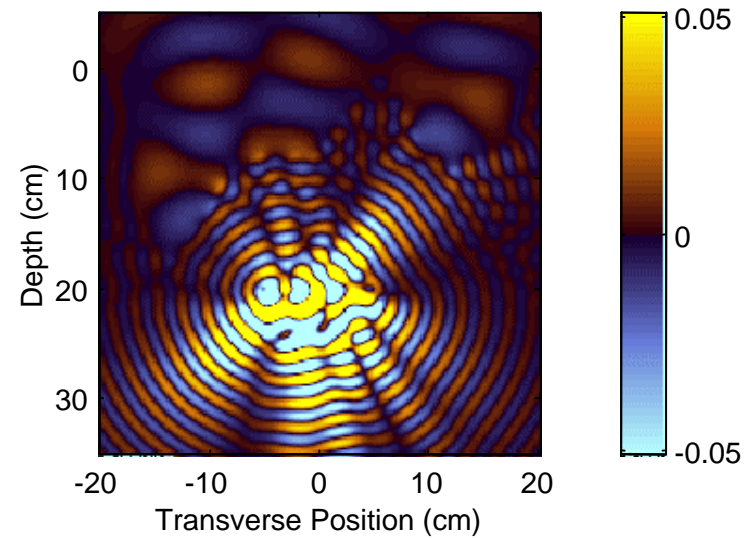
Scattered field from rough surface and mine



Scattered field from rough surface



Scattered field from mine with rough surface



Propagation in Soil is Frequency Dependent

$$k(f) = b(f) - ja(f) = \frac{2\pi f}{c} \sqrt{\epsilon'(f) - j \frac{\sigma(f)}{2\pi f \epsilon_0}}$$

Modeling Dispersion for Easy Transformation to Time Domain

Debye (Teixeira)

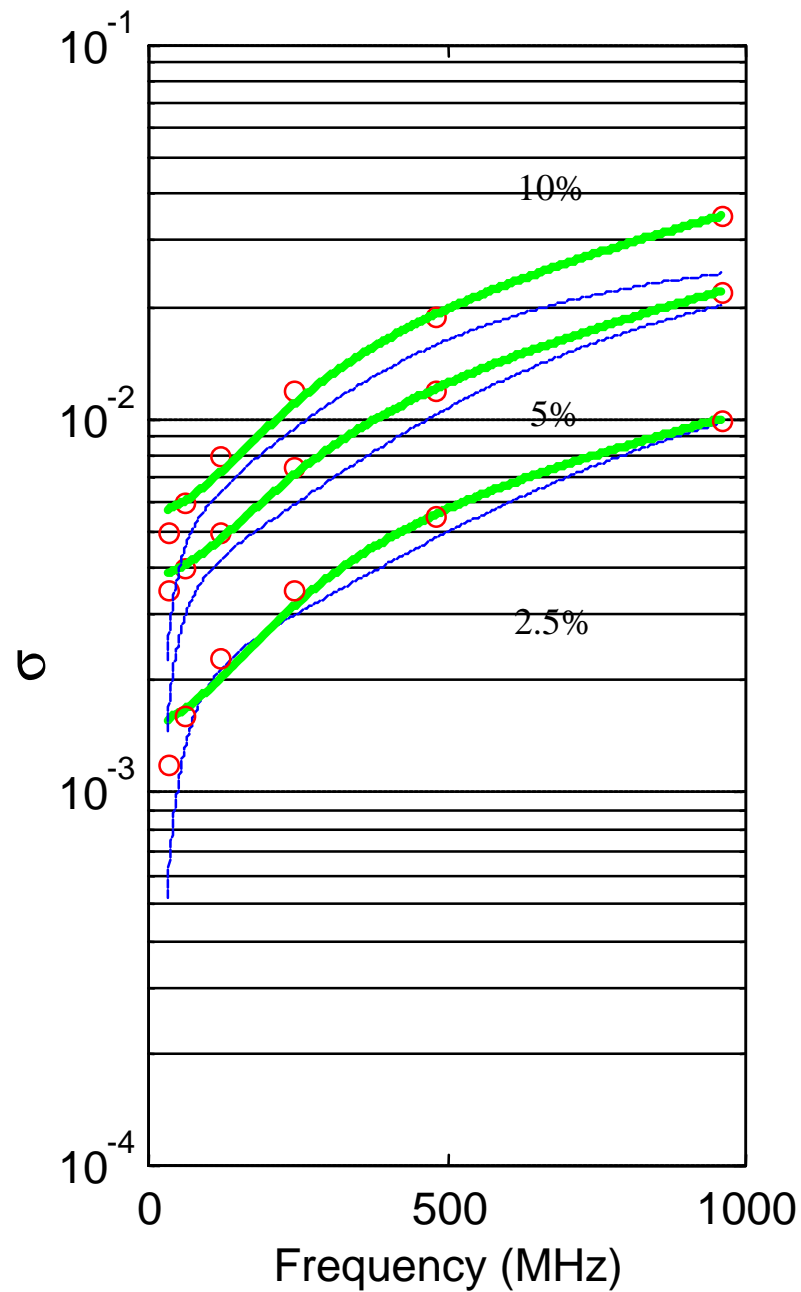
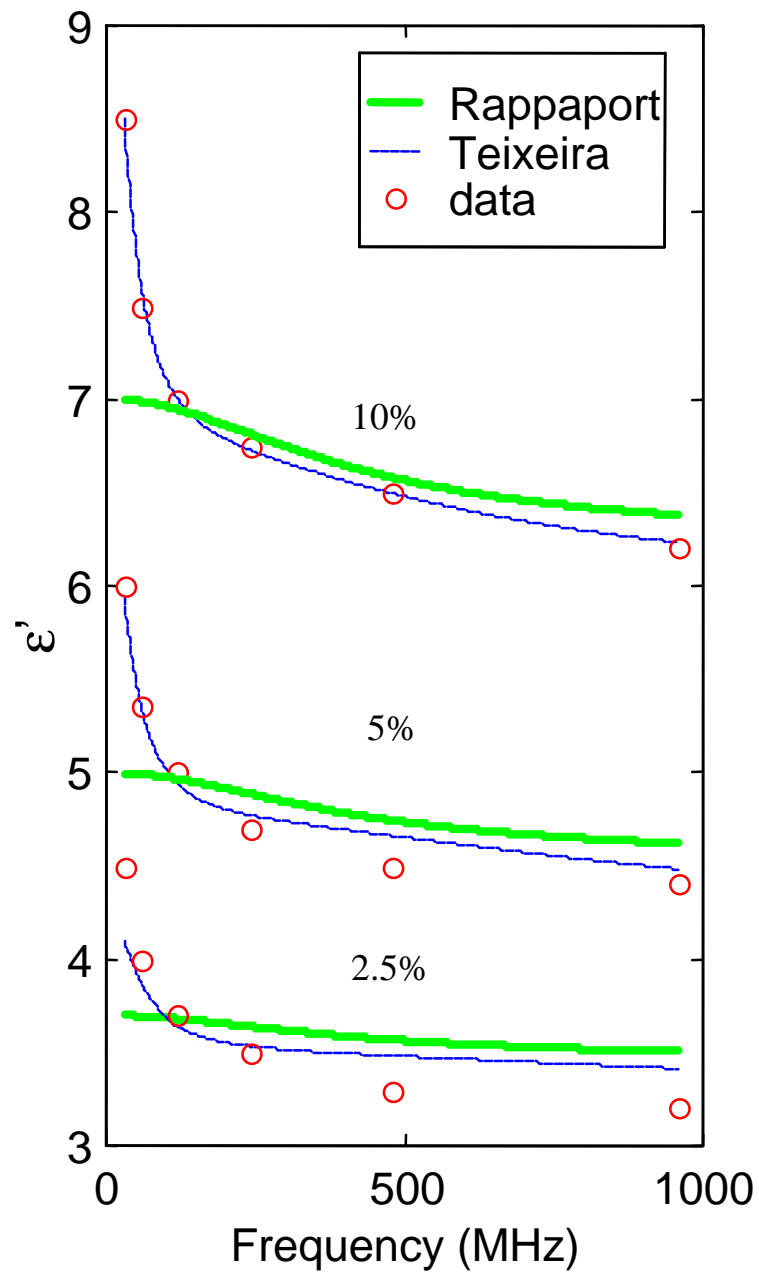
$$\mathbf{e} = \mathbf{e}_0 \mathbf{e}_\infty + \mathbf{e}_0 \sum_{p=1}^N \frac{A_p}{1 - i\omega t_p} \quad N=2$$

Rappaport

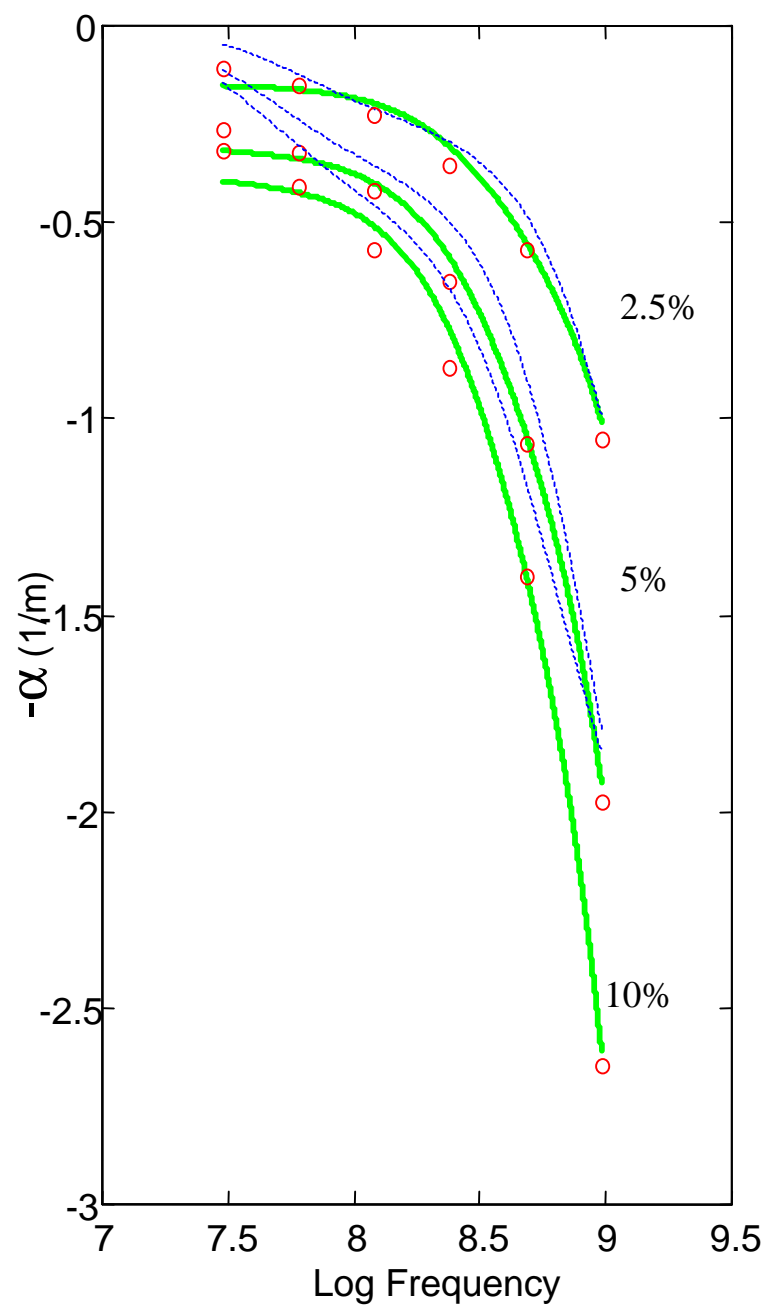
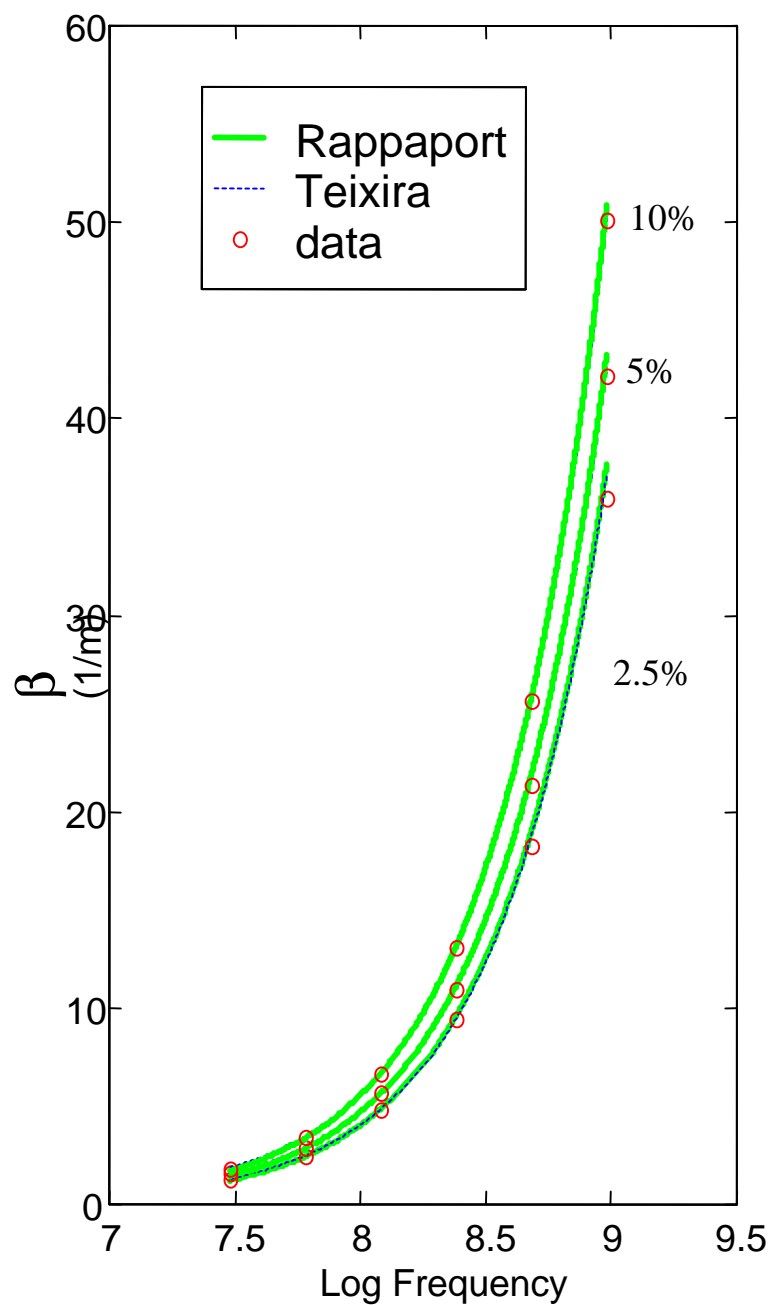
$$\epsilon' = \text{Constant}, \quad \mathbf{s} = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2}}{1 + a_1 Z^{-1}}, \quad Z = e^{i\omega \Delta t}$$

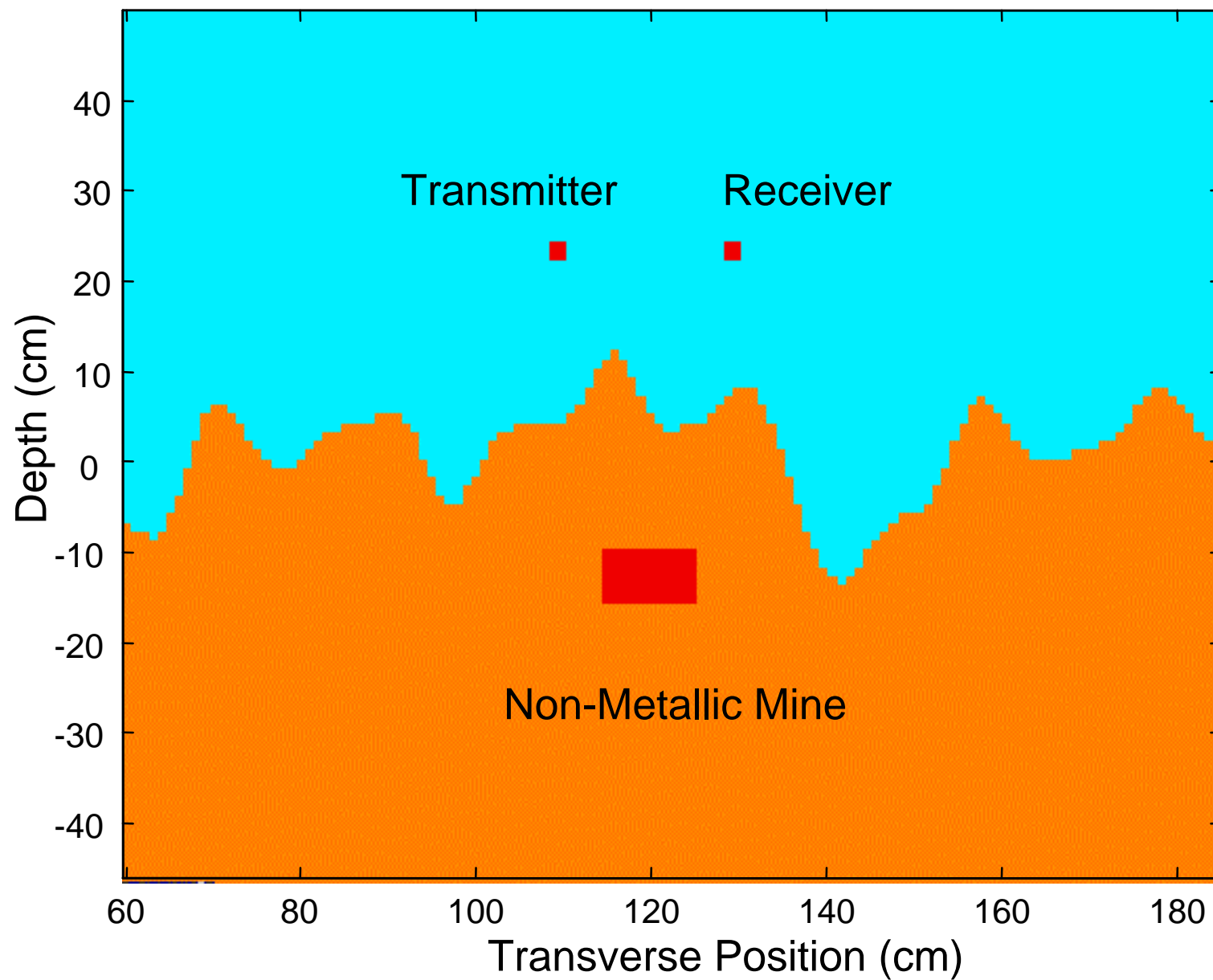
M o i s t u r e	2.5 %	5 %	10 %
b 0	. 5 8 1	1 . 3 6	2 . 1 5
b 1	- 1 . 1 2	- 2 . 6 4	- 4 . 1 8
b 2	. 5 4 6	1 . 2 8	2 . 0 3
a 1	- 0 . 9 5	- 0 . 9 5	- 0 . 9 5
ϵ'	2 . 1 8	1 . 5 2	1 . 4 6

Dielectric Constant and Conductivity for Puerto Rican Clay Loam (1.2 g/cc)

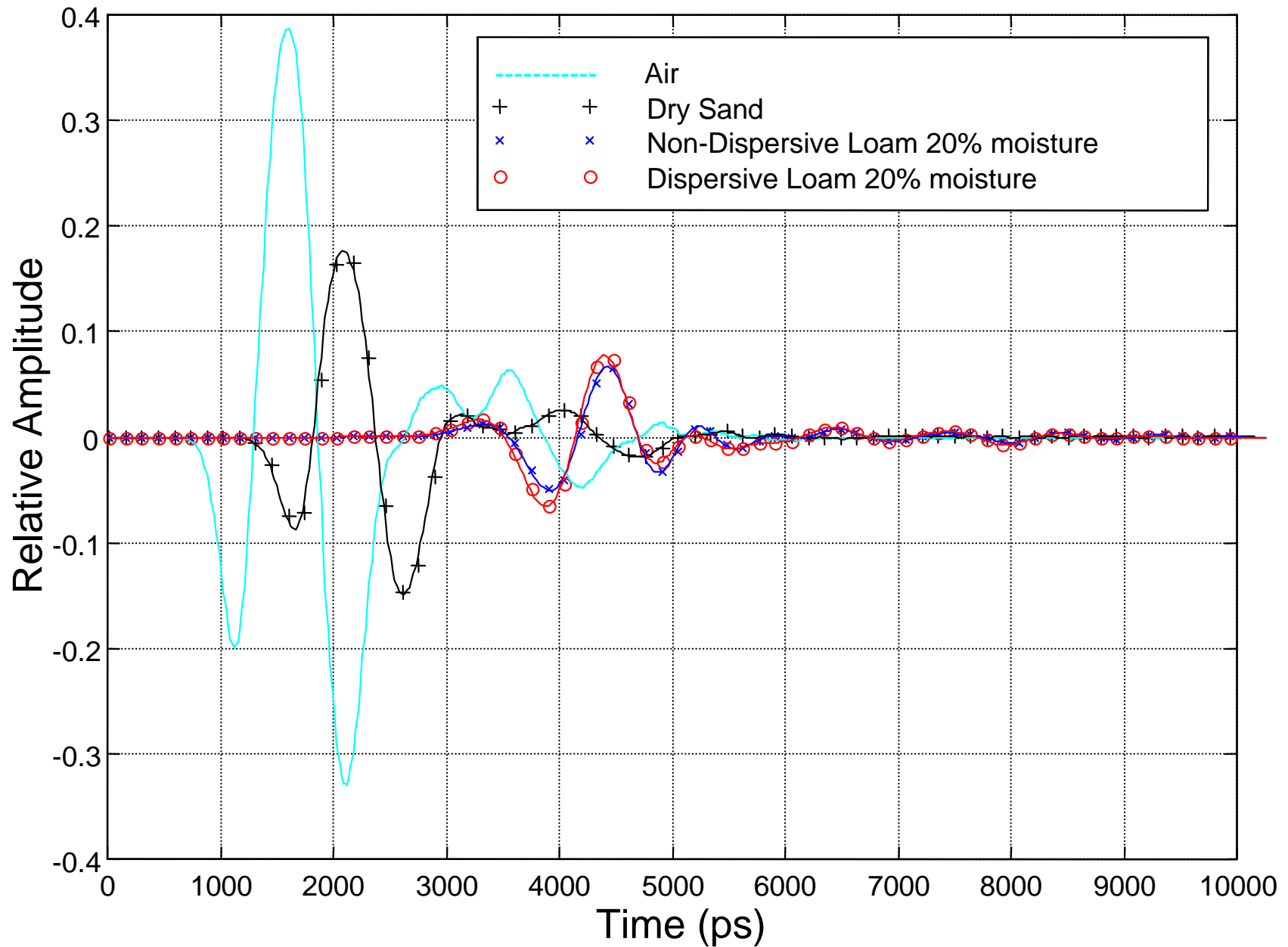


Real and Imaginary Wave Number for Puerto Rican Clay Loam (1.2g/cc)

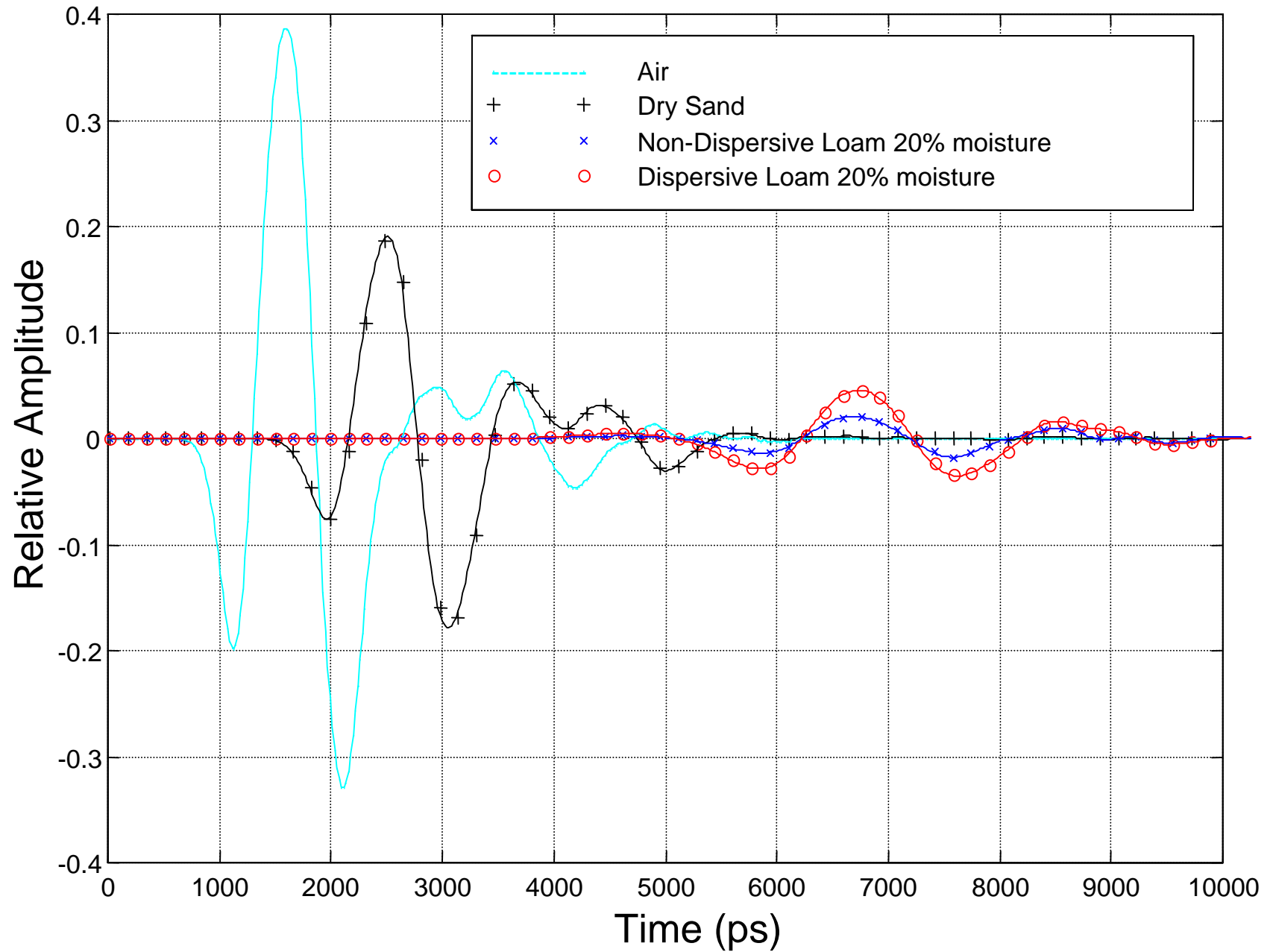




Non-Metallic Mine Scattered Field (about 10 cm burial) - Smooth Surface

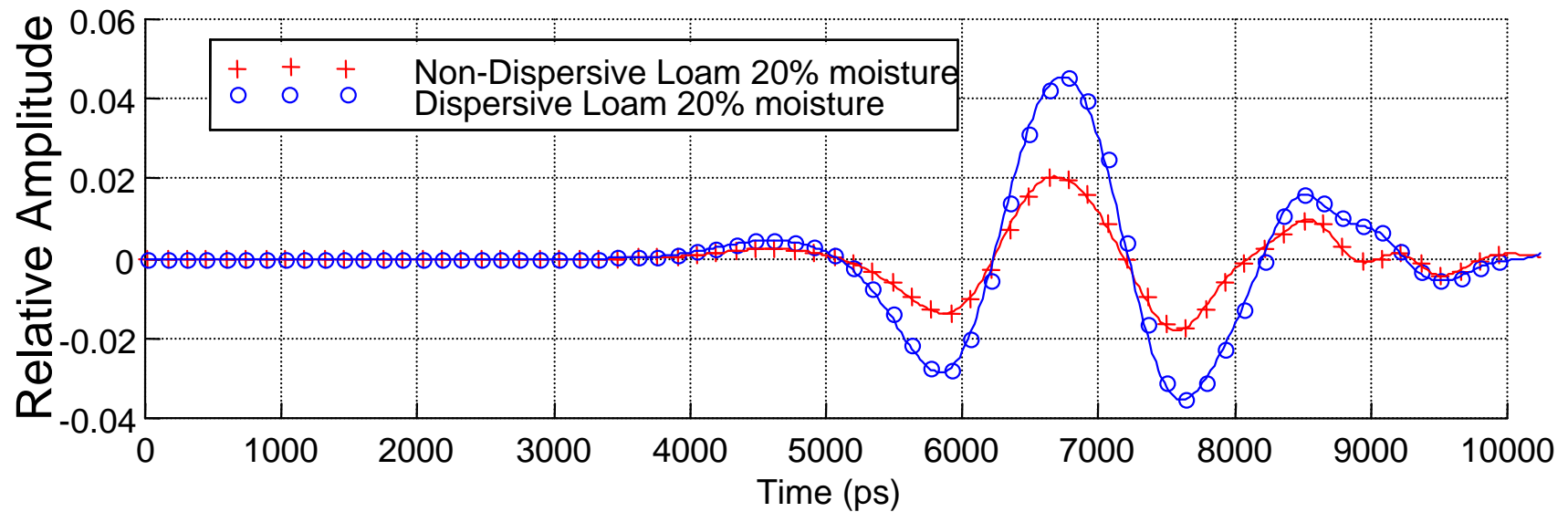
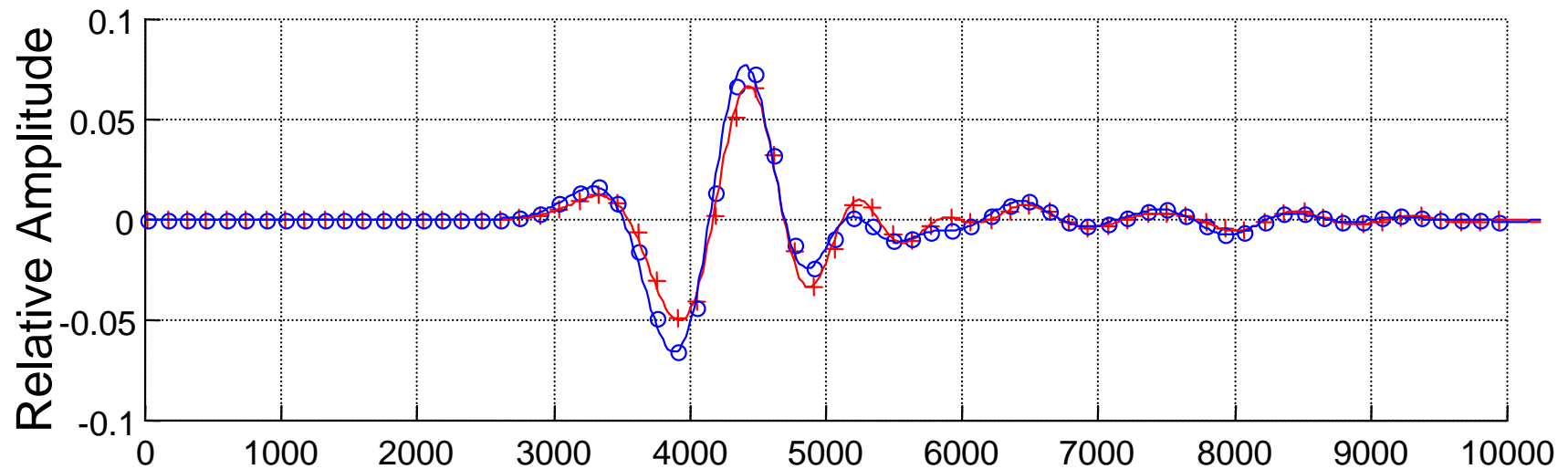


Non-Metallic Mine Scattered Field (about 10 cm burial) - Rough Surface

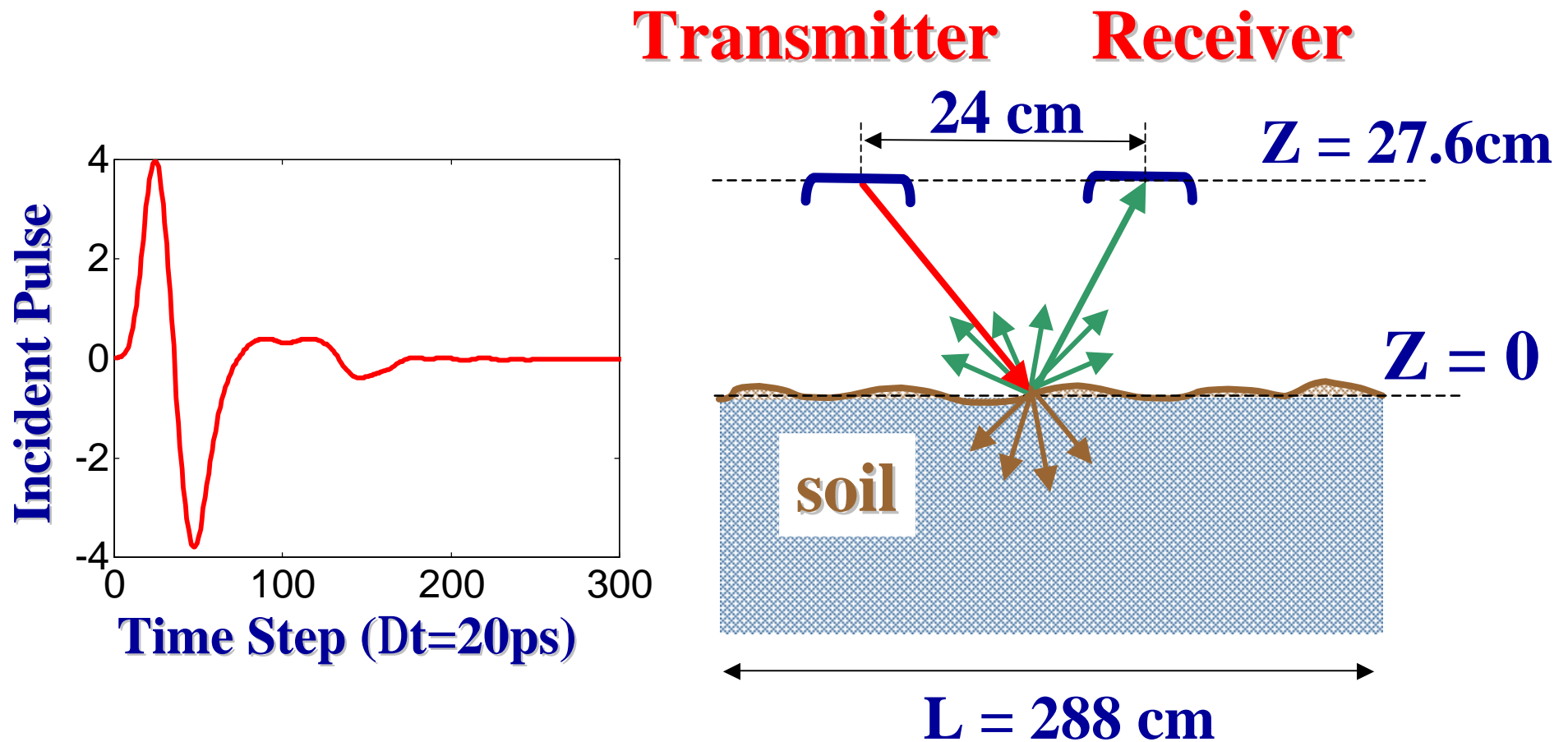


Non-Metallic Mine Scattered Field (10 cm depth)

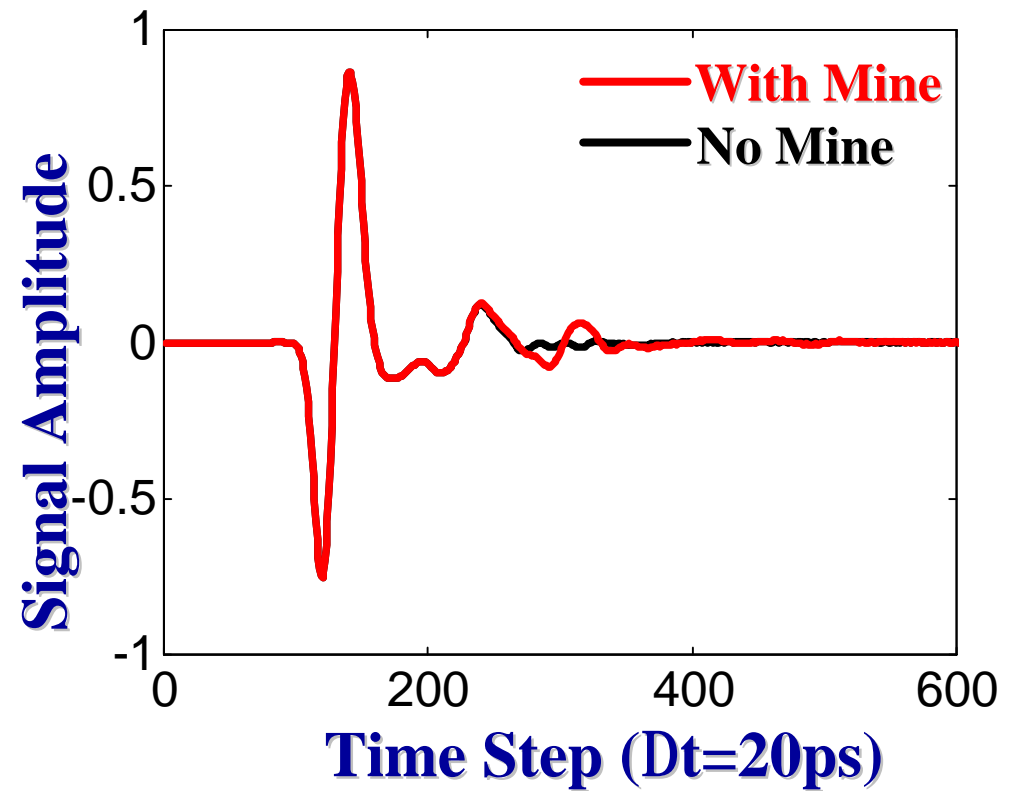
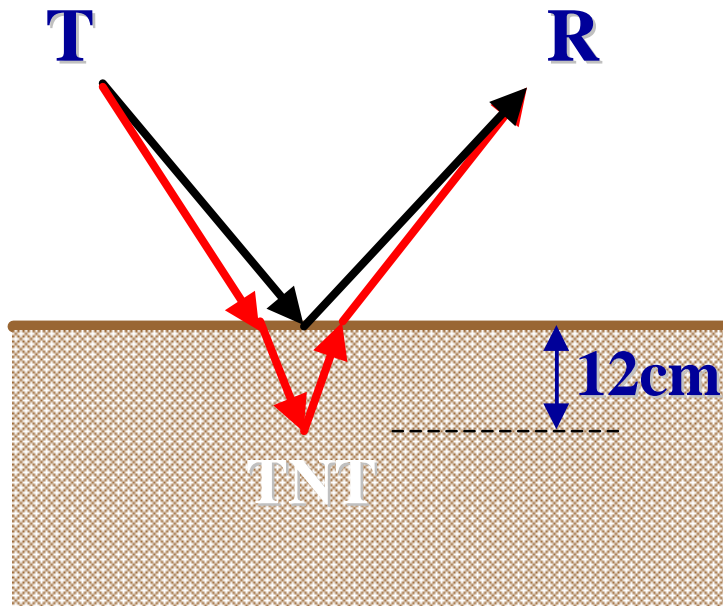
a) Smooth Surface, b) Rough Surface



Rough Surface Sensing Geometry



Flat Ground Scattering



Rough Surface Characterization

↑ **Gaussian joint PDF for surface heights with zero mean and r.m.s = S_h :**

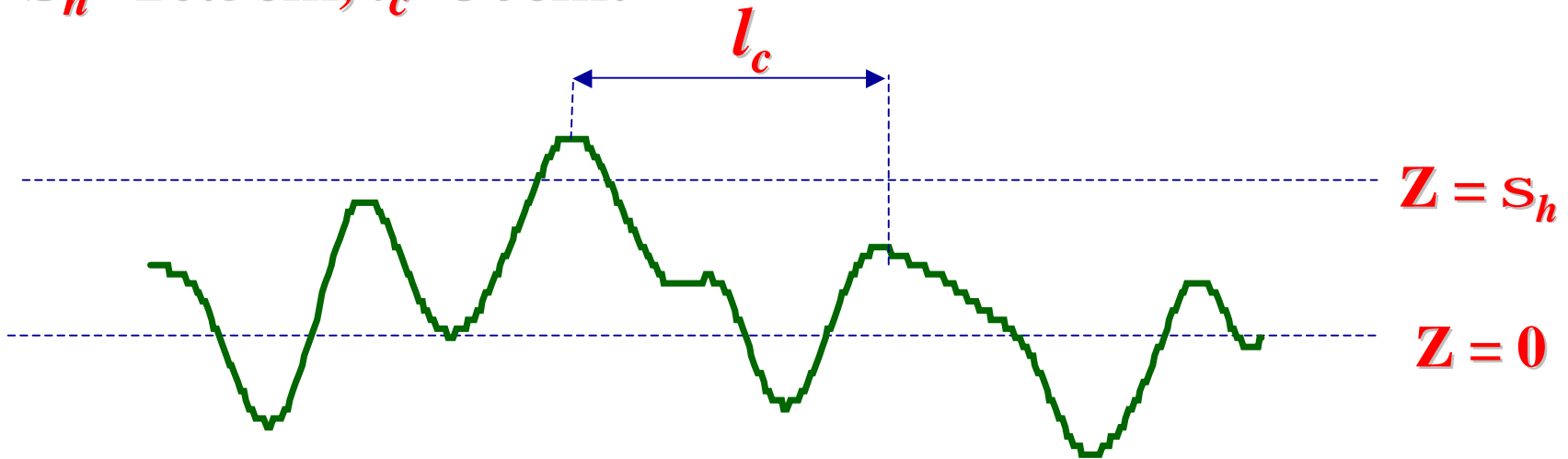
$$p(z_1, z_2) = \frac{1}{2\pi S_h^2 \sqrt{1 - R^2(l_c)}} \exp\left(-\frac{z_1^2 - 2R(l_c)z_1z_2 + z_2^2}{2S_h^2 [1 - R^2(l_c)]}\right)$$

↑ **Surface height auto-correlation function:**

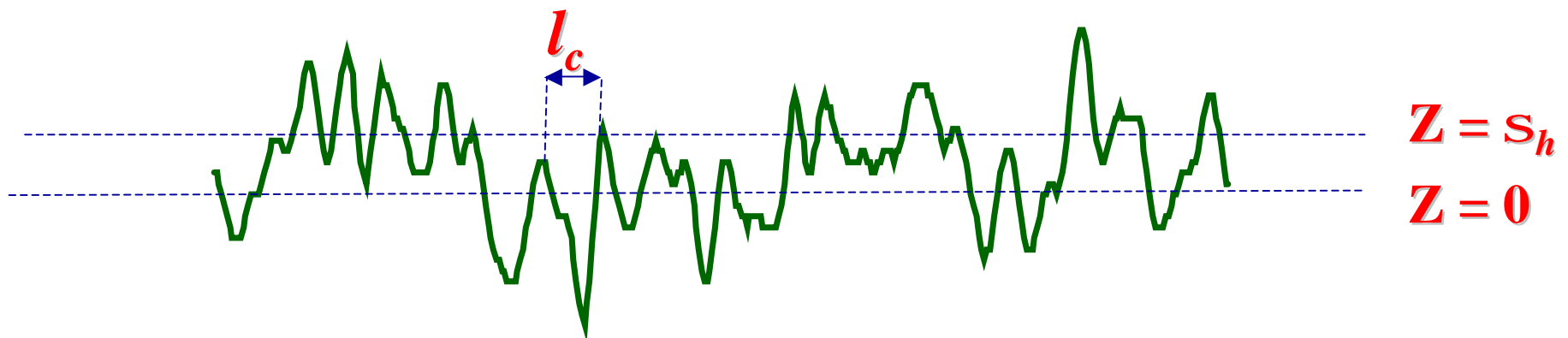
$$R(l_c) = \frac{S_h^2}{S_h^2} \exp\left(-\frac{(x_1 - x_2)^2}{l_c^2}\right)$$

Rough Surface Parameters

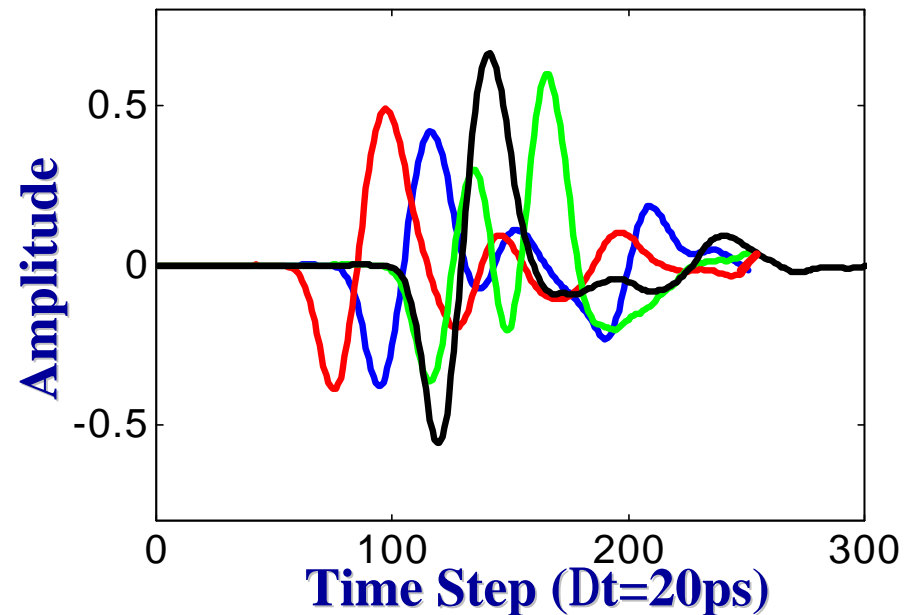
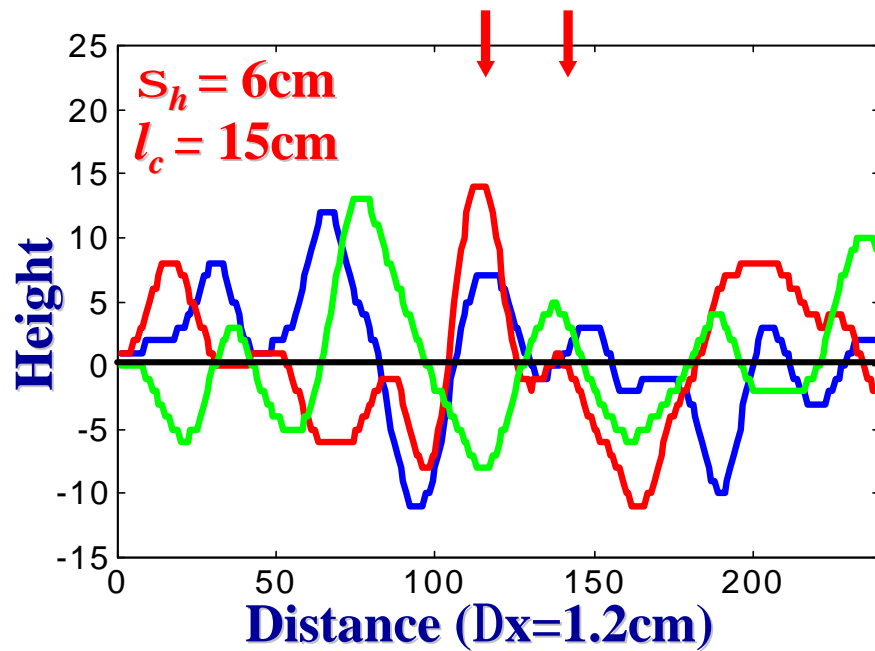
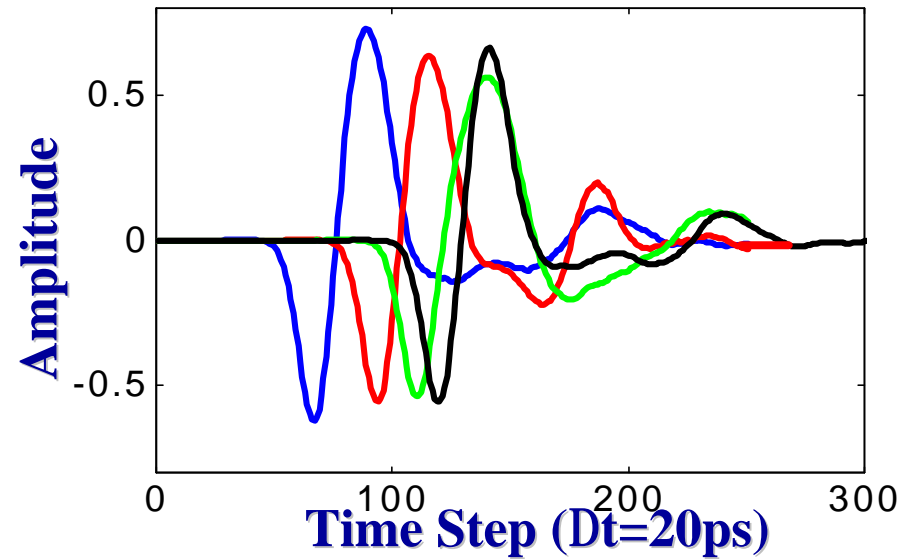
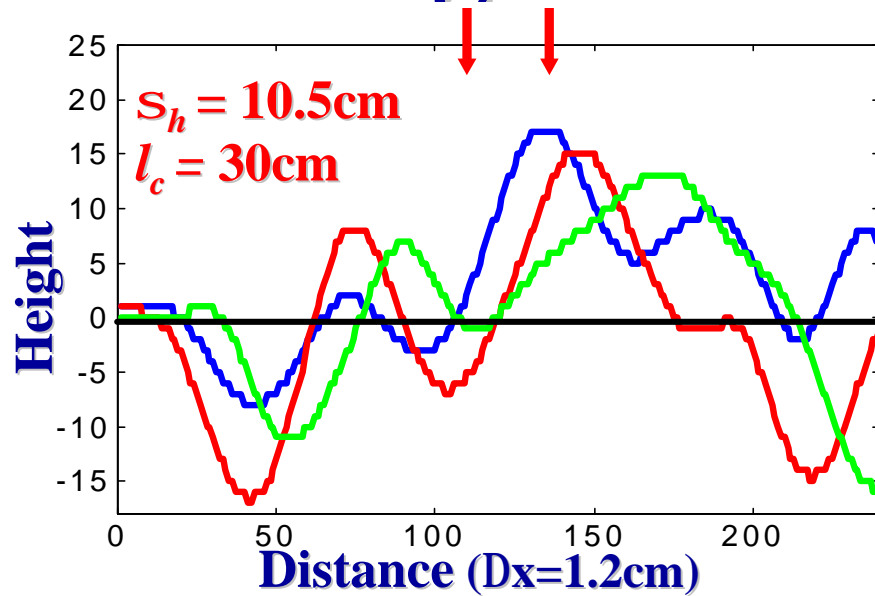
↑ $S_h = 10.5\text{cm}$, $l_c = 30\text{cm}$:



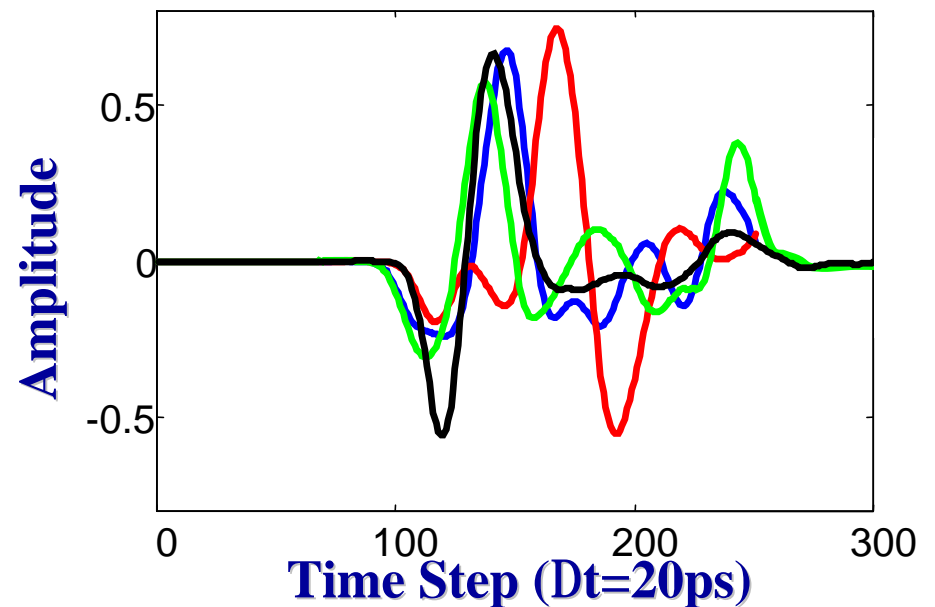
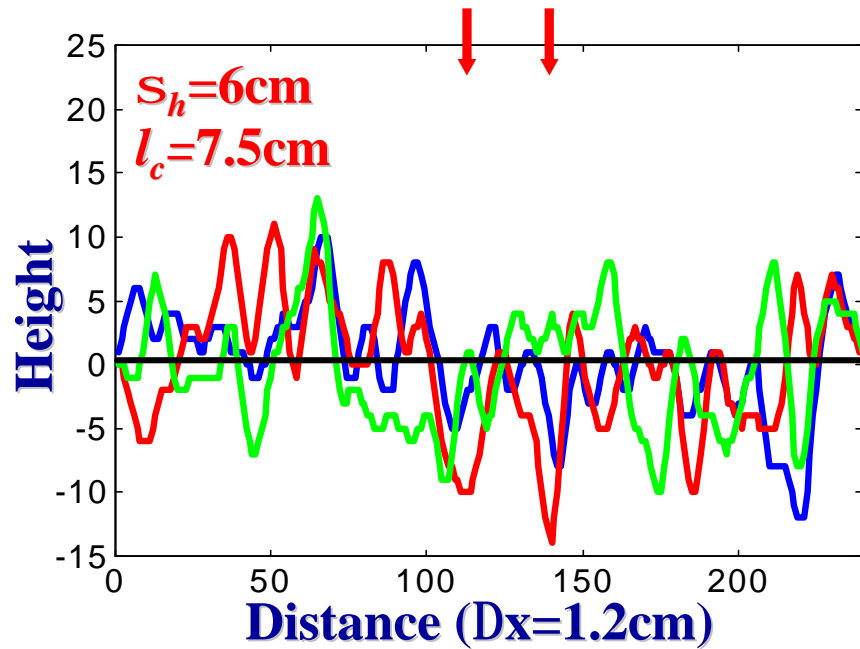
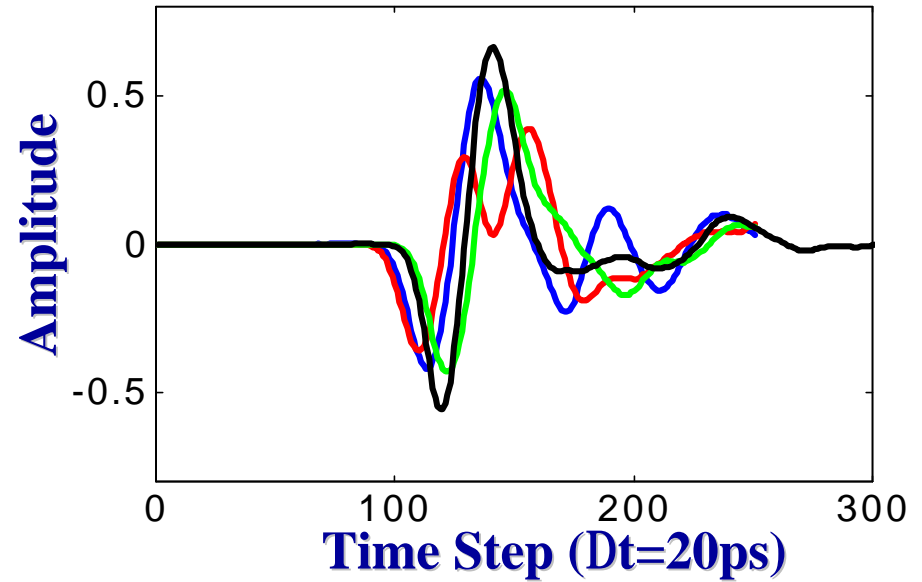
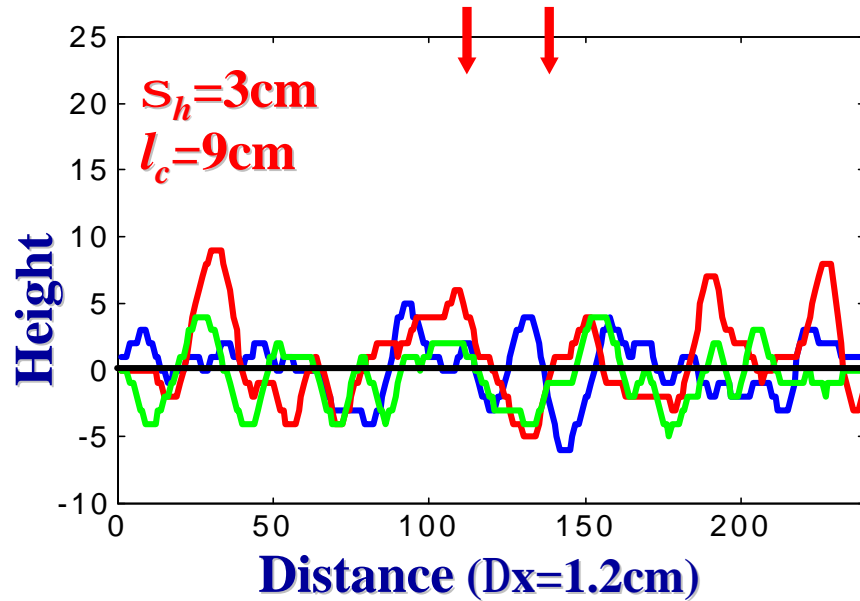
↑ $S_h = 6\text{cm}$, $l_c = 7.5\text{cm}$:



Rough Surface Scattering (1)



Rough Surface Scattering (2)



Cross-Correlation Function

↑ **Normalized Cross-Correlation Functions between Flat 'f' and Rough 'i' Surface:**

$$C_{fi}(m) = \sum_{n=1}^{N-|m|} S_f(n) S_i(n+m) / \sqrt{\sum_{n=1}^N |S_f(n)|^2} \sqrt{\sum_{n=1}^N |S_i(n)|^2} \quad \text{For } m > 0$$

$$C_{fi}(m) = C_{if}(-m) \quad \text{For } m < 0$$

↑ **Relative Amplitude:**

$$A_i(m) = \sum_{n=1}^N |S_i(n)|^2 / \sum_{n=1}^N |S_f(n)|^2$$

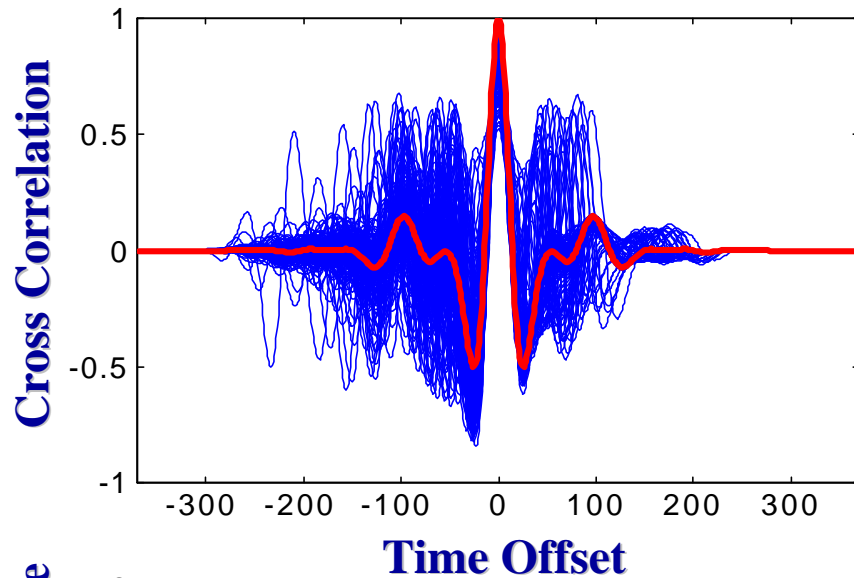
where $i = 1, 2 \dots M$.

M = number of surface realizations.

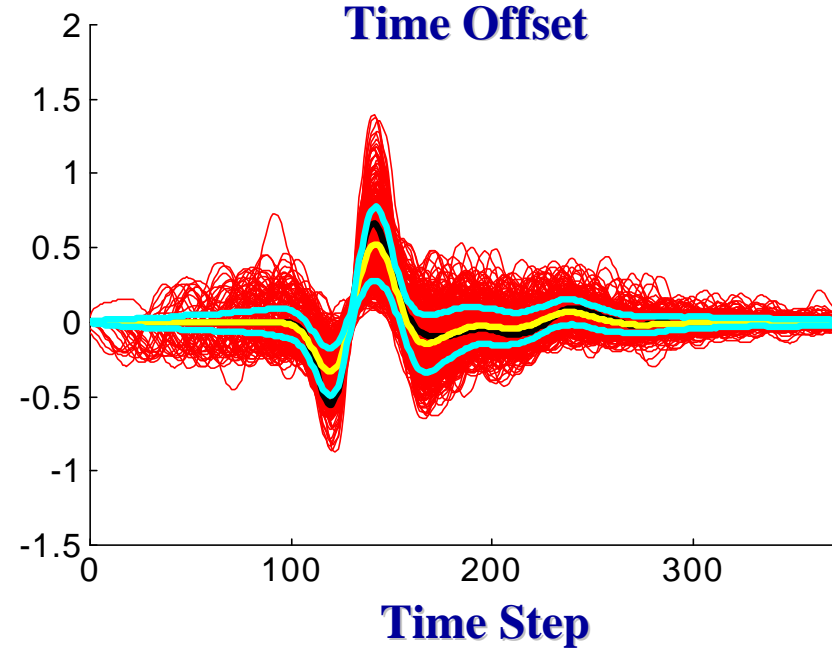
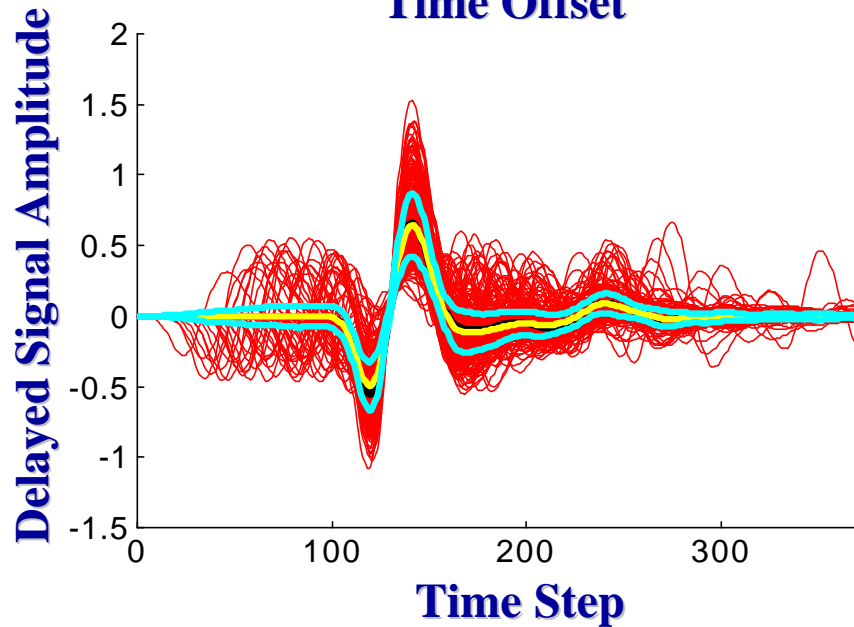
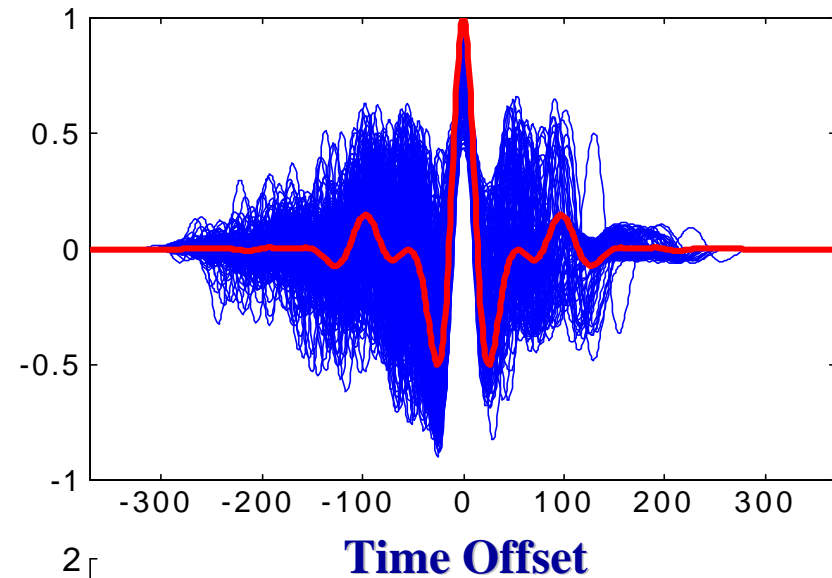
N = number of time steps in each signal.

Ground Surface Scattering

$s_h = 10.5\text{cm}, l_c = 30\text{cm}$

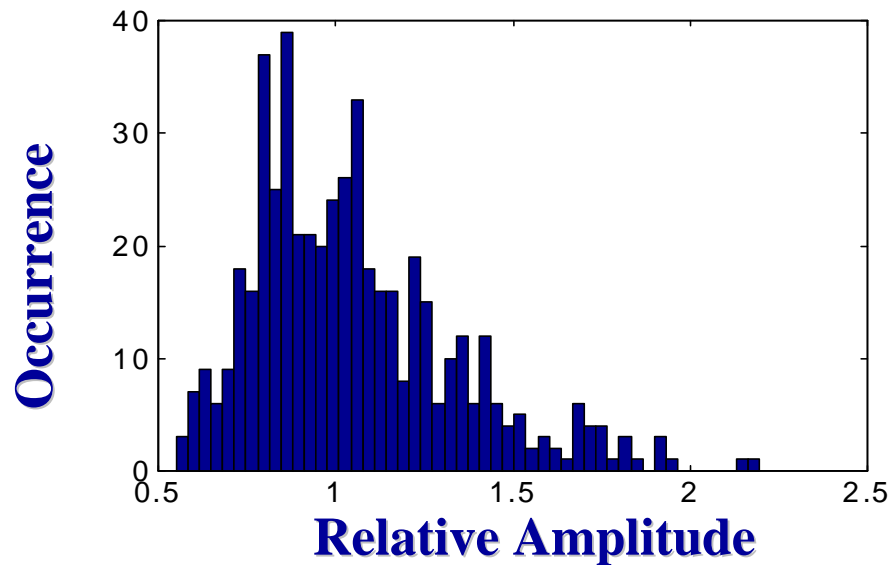


$s_h = 6\text{cm}, l_c = 7.5\text{cm}$

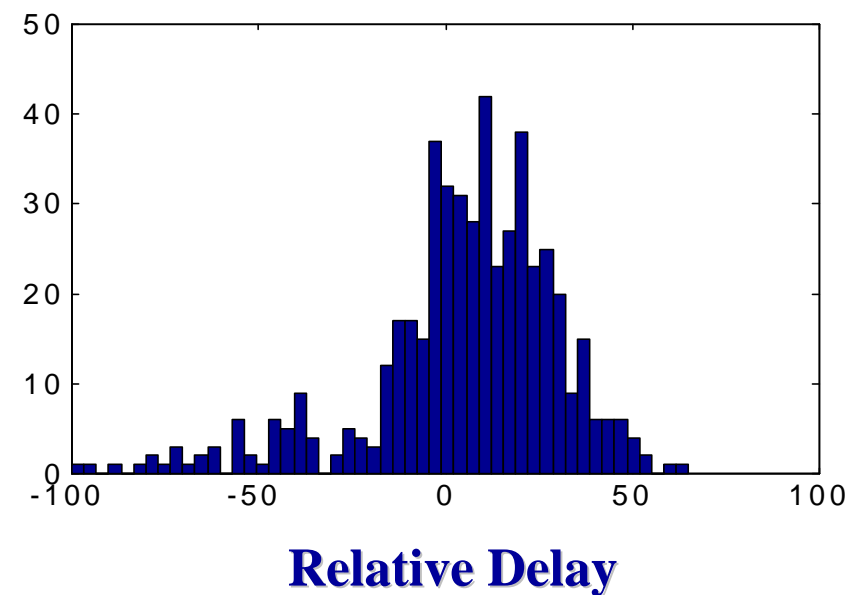
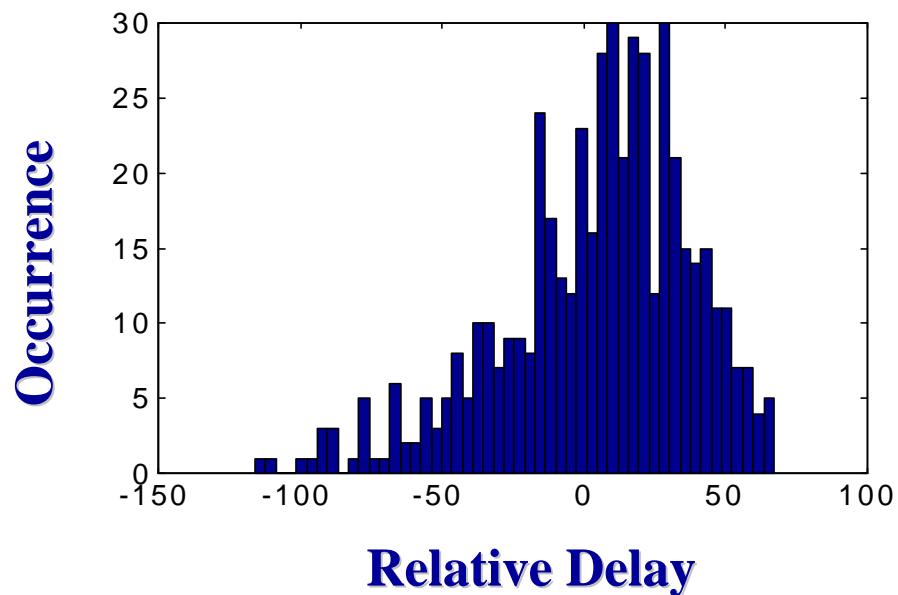
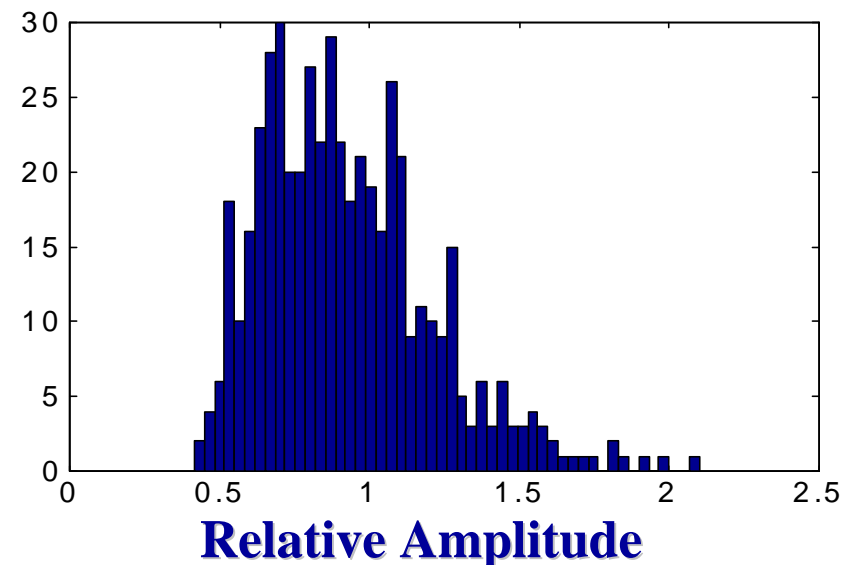


Scattering Signal Statistics

$s_h=10.5\text{cm}, l_c=30\text{cm}$

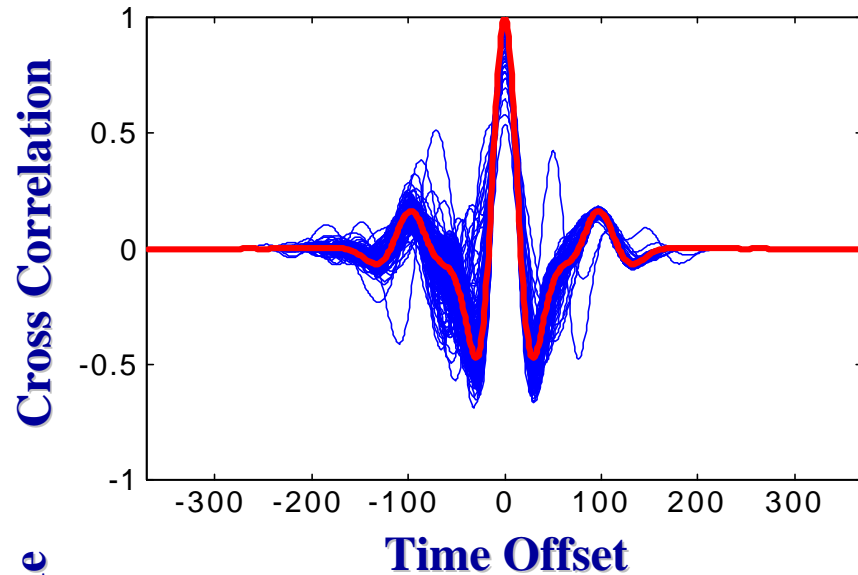


$s_h=6\text{ cm}, l_c=7.5\text{cm}$

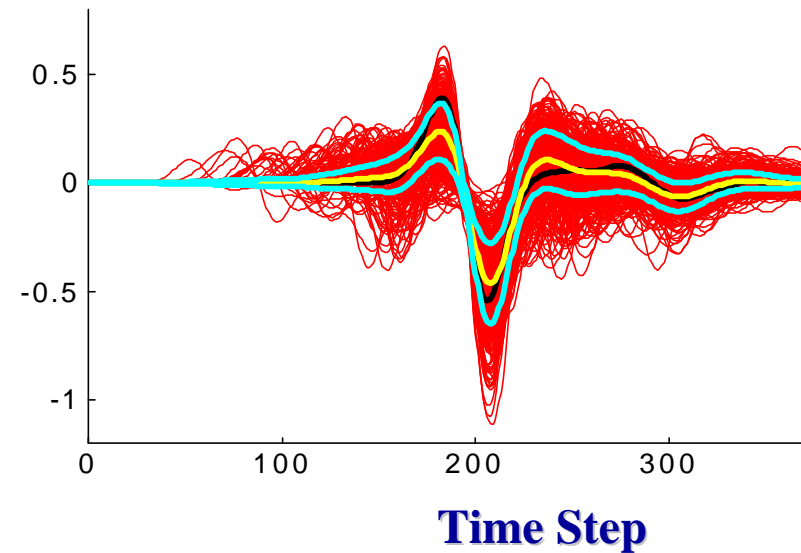
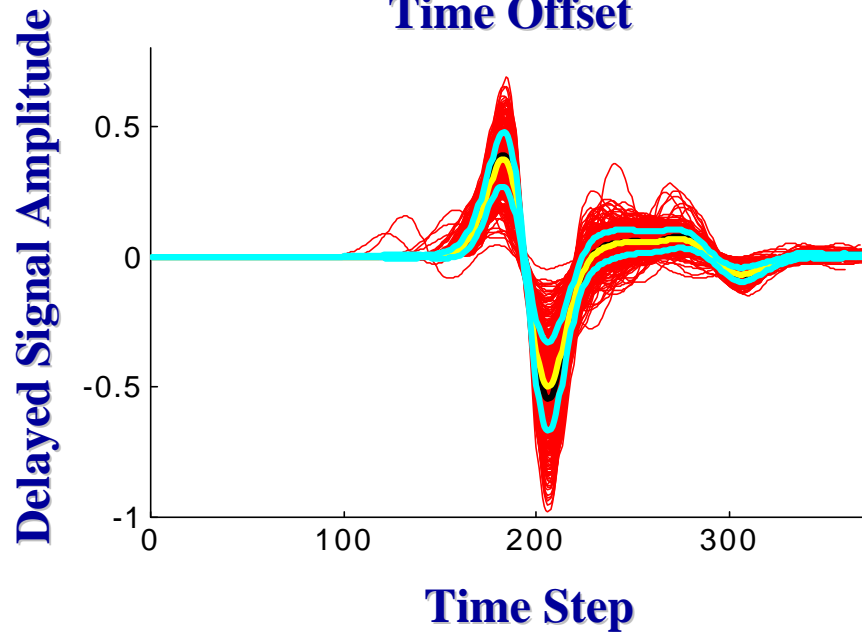
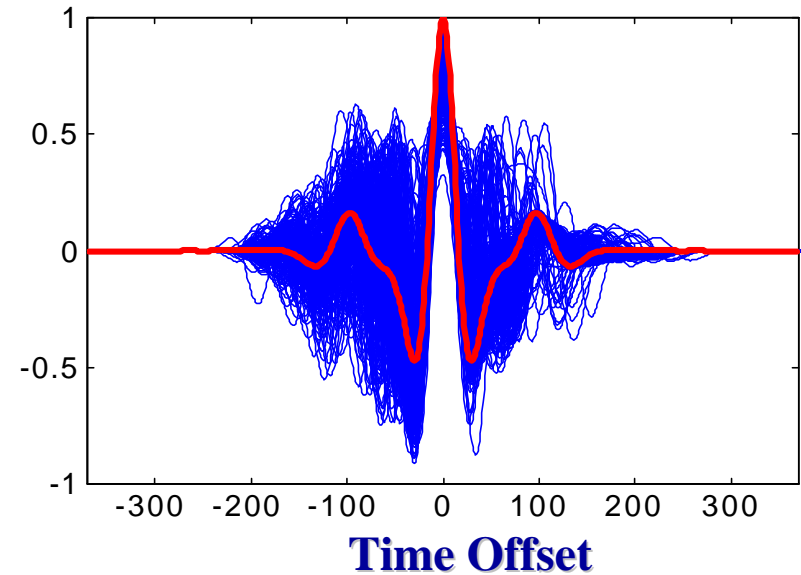


Ground Surface Transmission

$S_h=10.5\text{cm}, l_c=30\text{cm}$

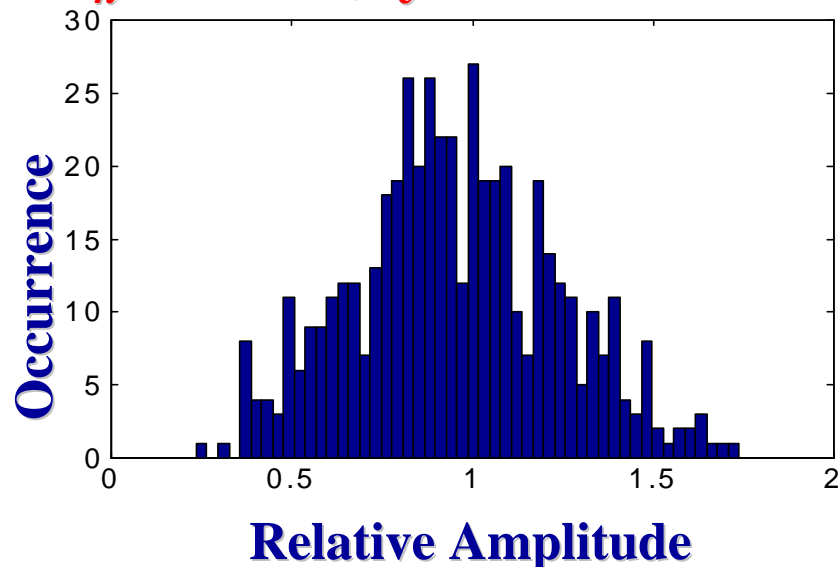


$S_h=6\text{cm}, l_c=7.5\text{cm}$

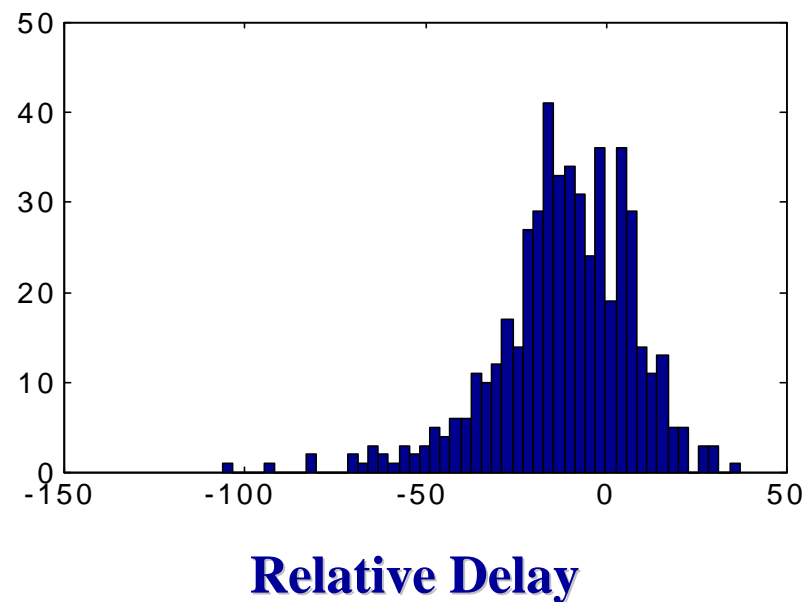
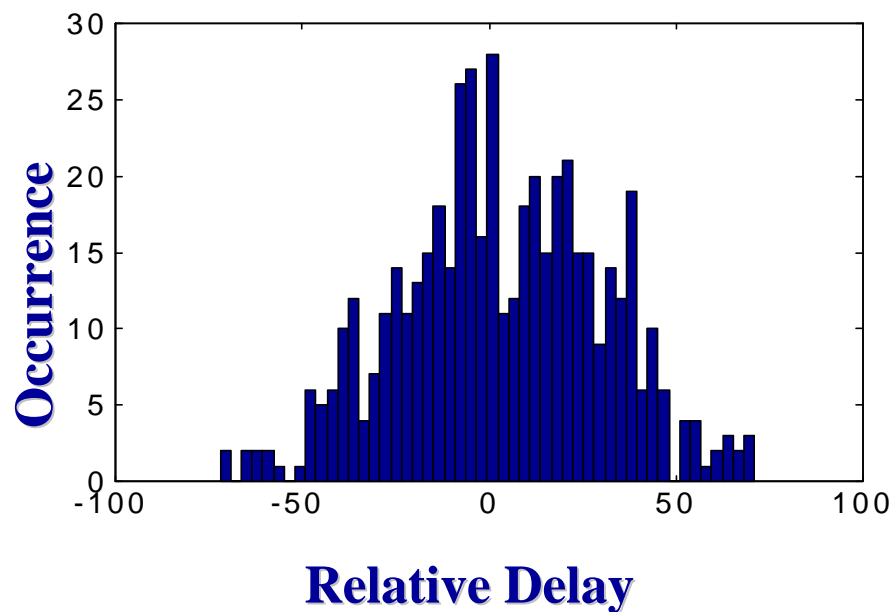
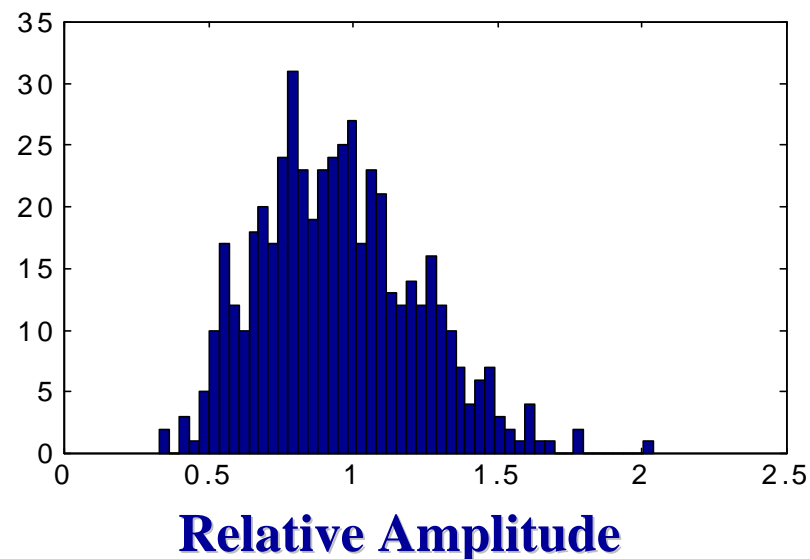


Transmission Signal Statistics

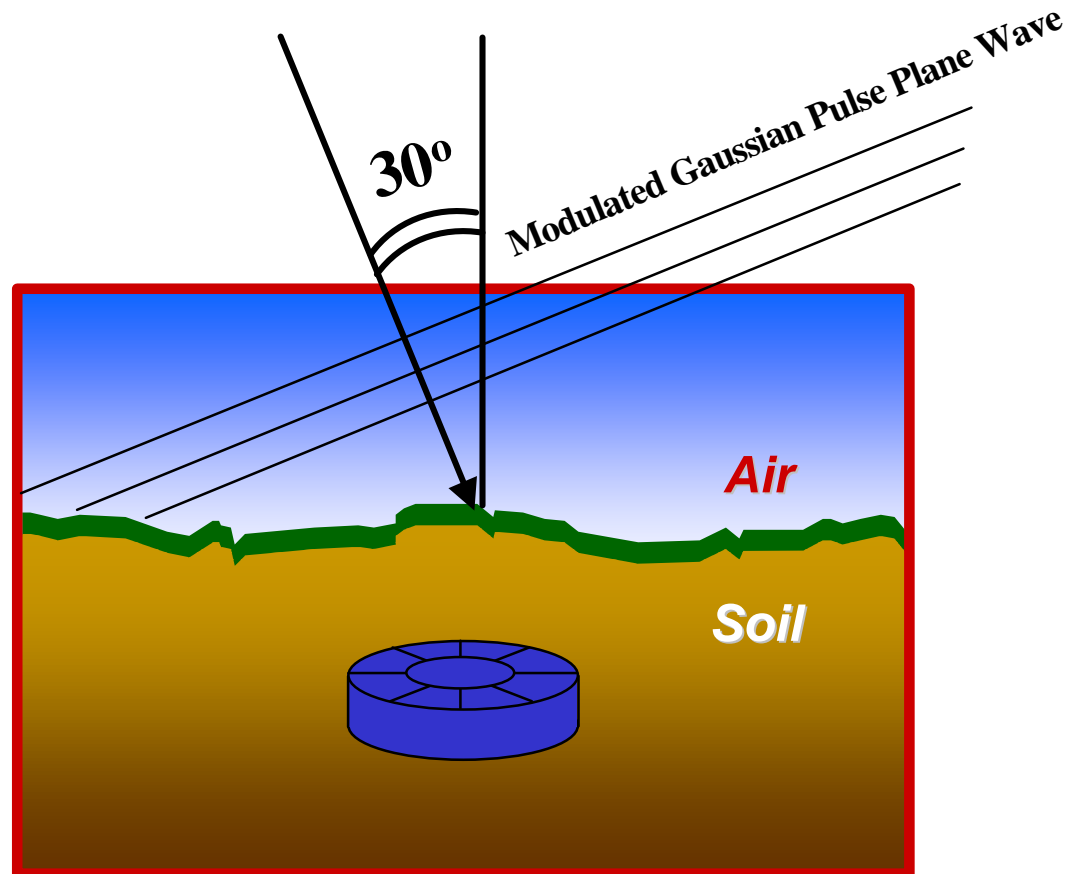
$S_h=10.5\text{cm}, l_c=30\text{cm}$



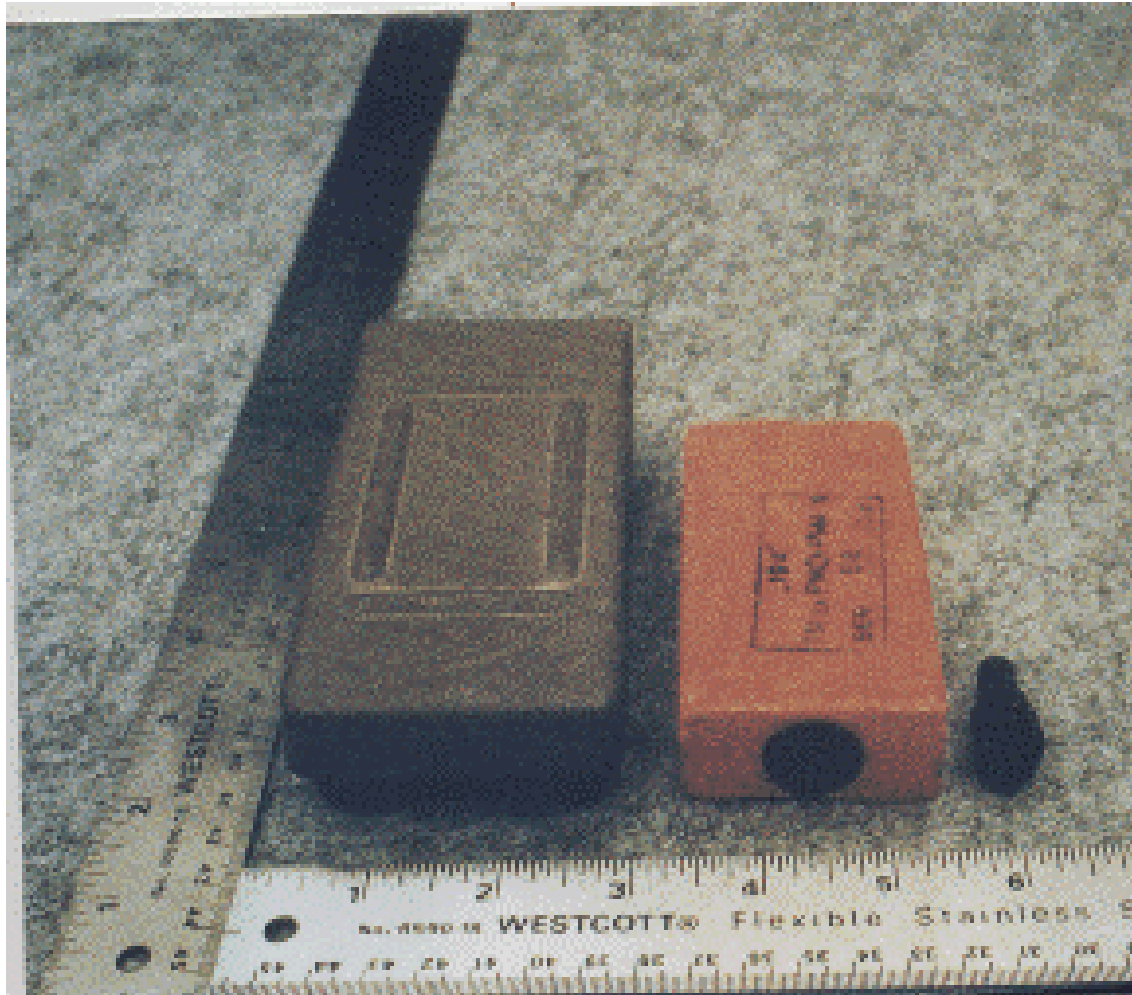
$S_h=6\text{cm}, l_c=7.5\text{cm}$



Short Pulse GPR Interaction with Rough, Dispersive Ground / Mine

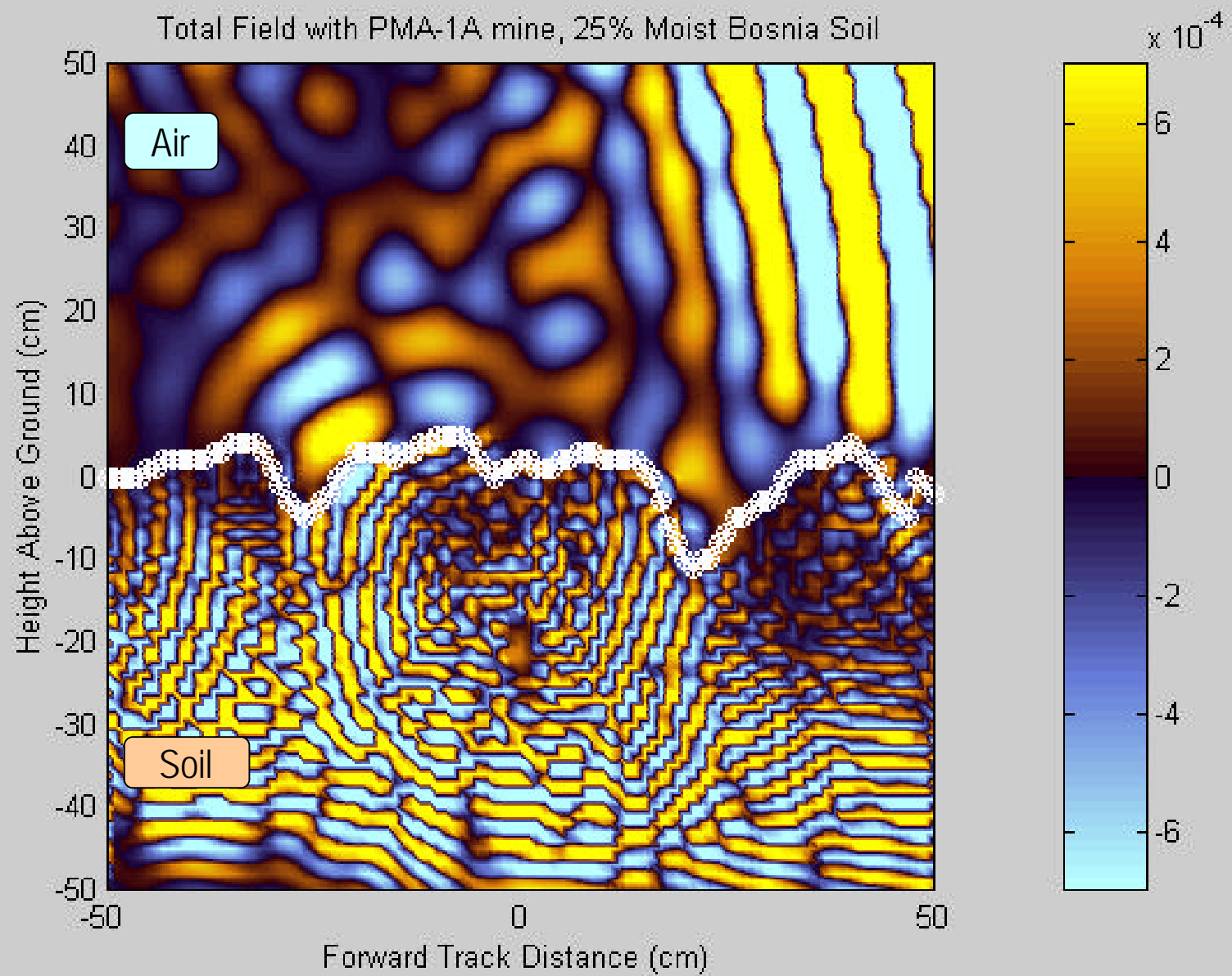


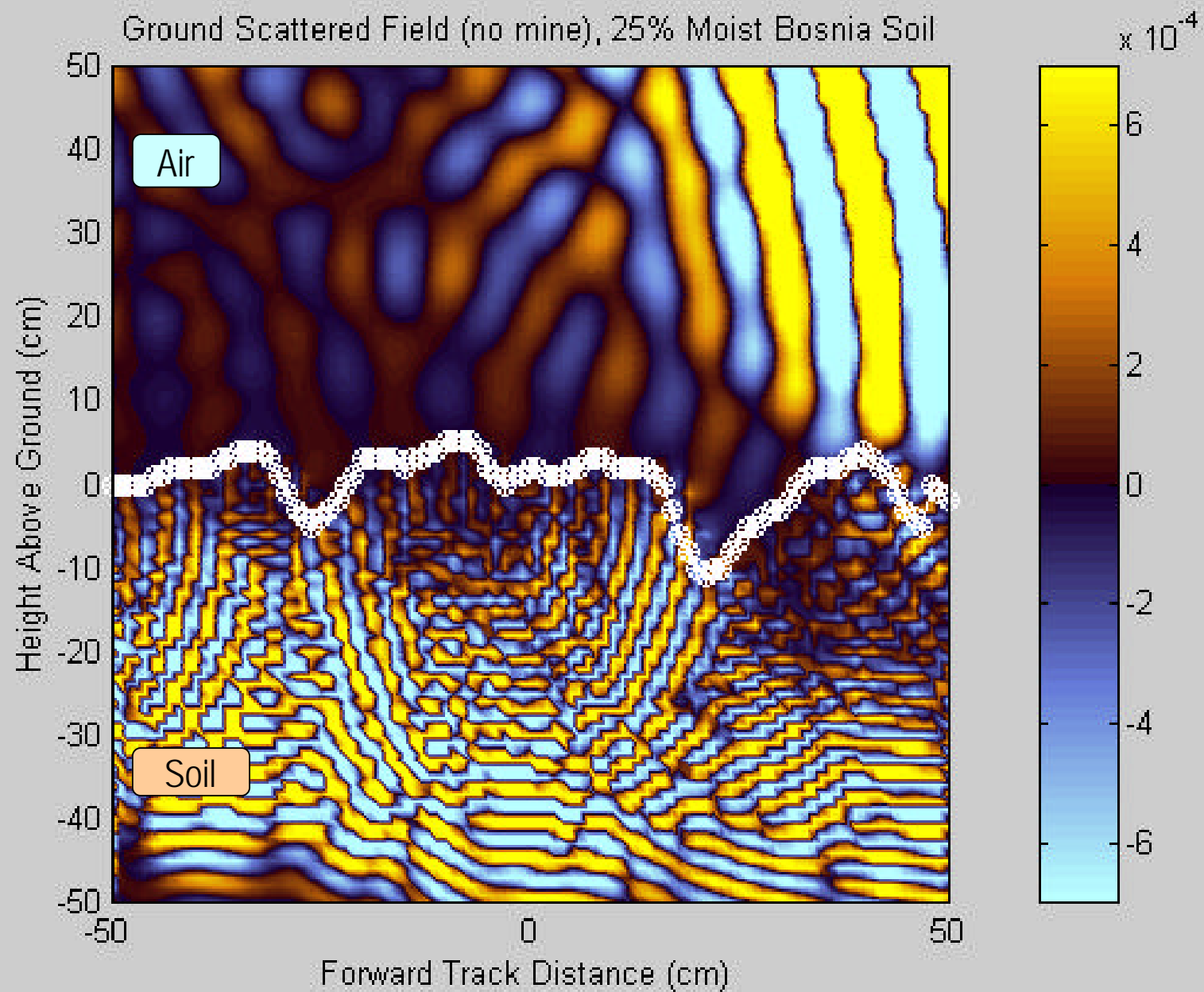
PMN-1A Non-Metallic AP Mine



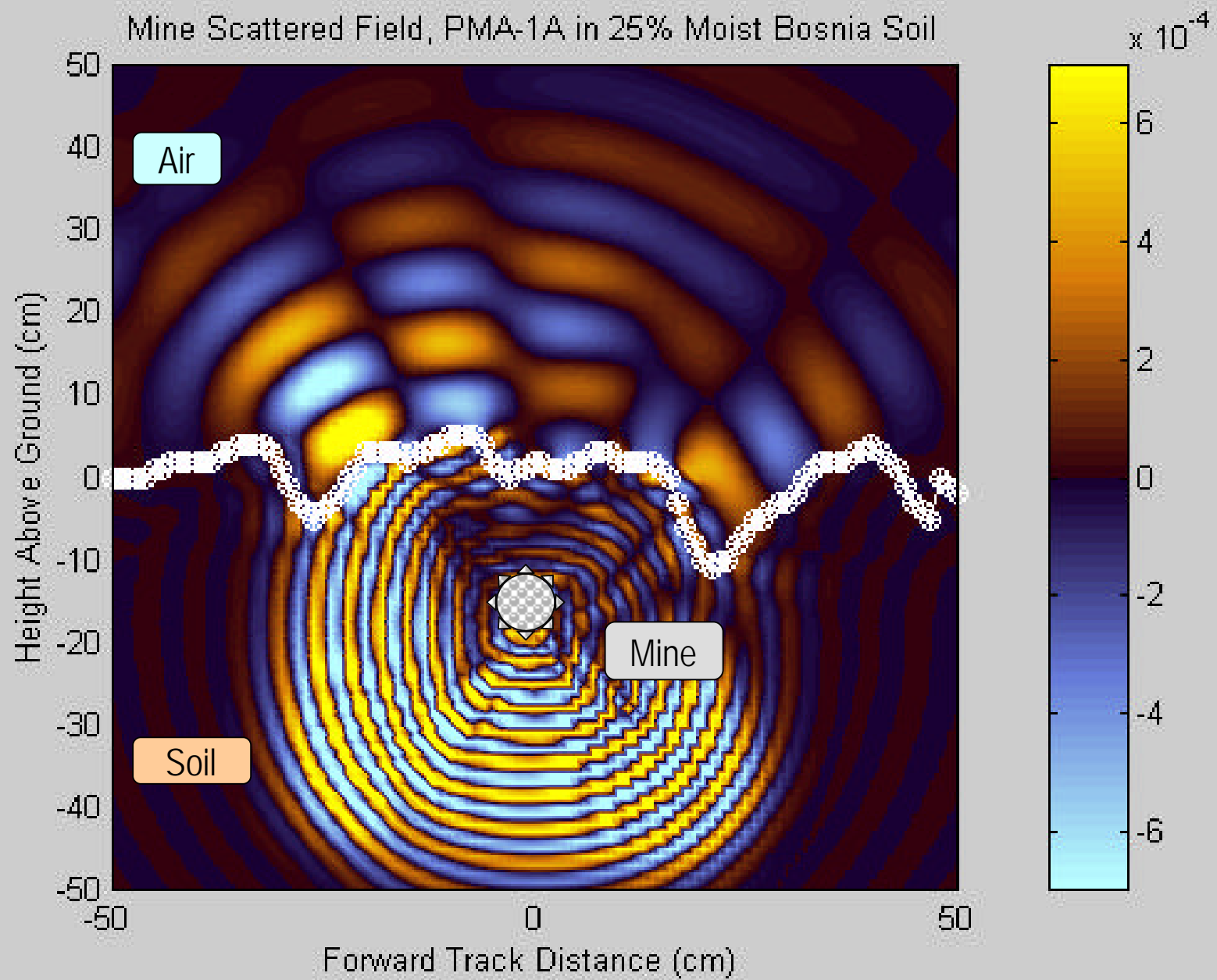
From MineFacts, version 1.2, National Ground Intelligence Center

Total Field with PMA-1A mine, 25% Moist Bosnia Soil





Mine Scattered Field, PMA-1A in 25% Moist Bosnia Soil



Conclusions

- Rough ground surface is the main difficulty with GPR detection/discrimination of mines
- For a single frequency, extraction of target signal from rough ground scattering is extremely difficult, even with ideal detectors
- Increasing sensing frequency improves detection resolution, but also greatly increases clutter signal
- Dispersion is an important effect for time domain modeling, and can be included with supplemental difference equation.
- Monte Carlo FDTD modeling of rough scattering indicates wide variety of signal corruption: time shifting, scaling, distortion, which depend considerably with roughness statistics
- With shifting and scaling, expected value of ground scattered signals can be well characterized

Physics-based forward models for mine detection signal processing

Carey Rappaport and Eric Miller
Northeastern University
Boston, MA 02115

To effectively detect non-metallic buried antipersonnel mines, sophisticated signal processing that identifies possible target anomalies in the presence of widely varying artificial and environmental clutter are essential. These methods rely on precise, yet robust models of wave propagation and field distribution. The competing requirements of flexibility of including inhomogeneities in background soil and ground surface, wide frequency ranges for sensing, and computational efficiency demand careful model specification.

A key issue in this effort is the synthesis of computational forward models that are well suited for the processing task at hand. Ideally these models should be explicitly parameterized in terms of the information one wishes to extract from the data. If the goal of the algorithm is the determination of the locations, shapes, and electrical properties of an object in the receiver's field of view, then it is generally best to employ a model in which these parameters appear explicitly. In particular, for such a problem a T matrix based method would be ideal as this class of forward model exploits heavily the geometry of the scattering objects. However, T matrix methods model inhomogeneities poorly. Alternatively, a high dimensional finite difference scheme could include a rough ground surface and buried rocks, but is less effective for target recognition since the sampling points have no easily defined relationship to the objects' geometric characteristics.

In addition to the parameterization of the model, it is also of import that the model possess low computational complexity. Generally, any inversion scheme requires repeated use of the forward model to evaluate the suitability of the current estimate and determine how best to improve this guess. When the number of such evaluations is of the order of hundreds or thousands, clearly it is critical that the underlying model be as fast as possible. Continuing the previous example, because the T matrix approach admits a recursive implementation for multi-object scattering problems, in general it is far more efficient than a finite difference or finite element approach which requires the solution of a large system of linear equations.

Mine Detection Using a Handheld Parabolic Antenna: Analysis of Experimental Data



Bo Yang and Carey Rappaport
Center for Subsurface Sensing and Imaging Systems
Northeastern University
Boston MA

PIERS, Cambridge MA, July 5 2000

Outline

- Novel parabolic mine detector
- Result of the 2-D FDTD simulation of parabolic reflector in mine detection
- Experiment and data analysis
- Conclusion

Demining MURI Supported by ARO

Problem: Rough Ground Surface is Greatest Contributor to GPR Clutter

- For standoff GPR detection of small antipersonnel mines, the largest scattered signal comes from the ground surface.
- In the real world, the ground is never flat, the soil type is never known exactly, and the height of the antenna is never fixed.
- This clutter can be reduced by:
 - Time gating
 - Signal processing (matched filtering, deconvolving)
 - Spatially separating target/clutter signals

Advantages of Plane Wave Excitation Over Currently Used Mine Detection Technology

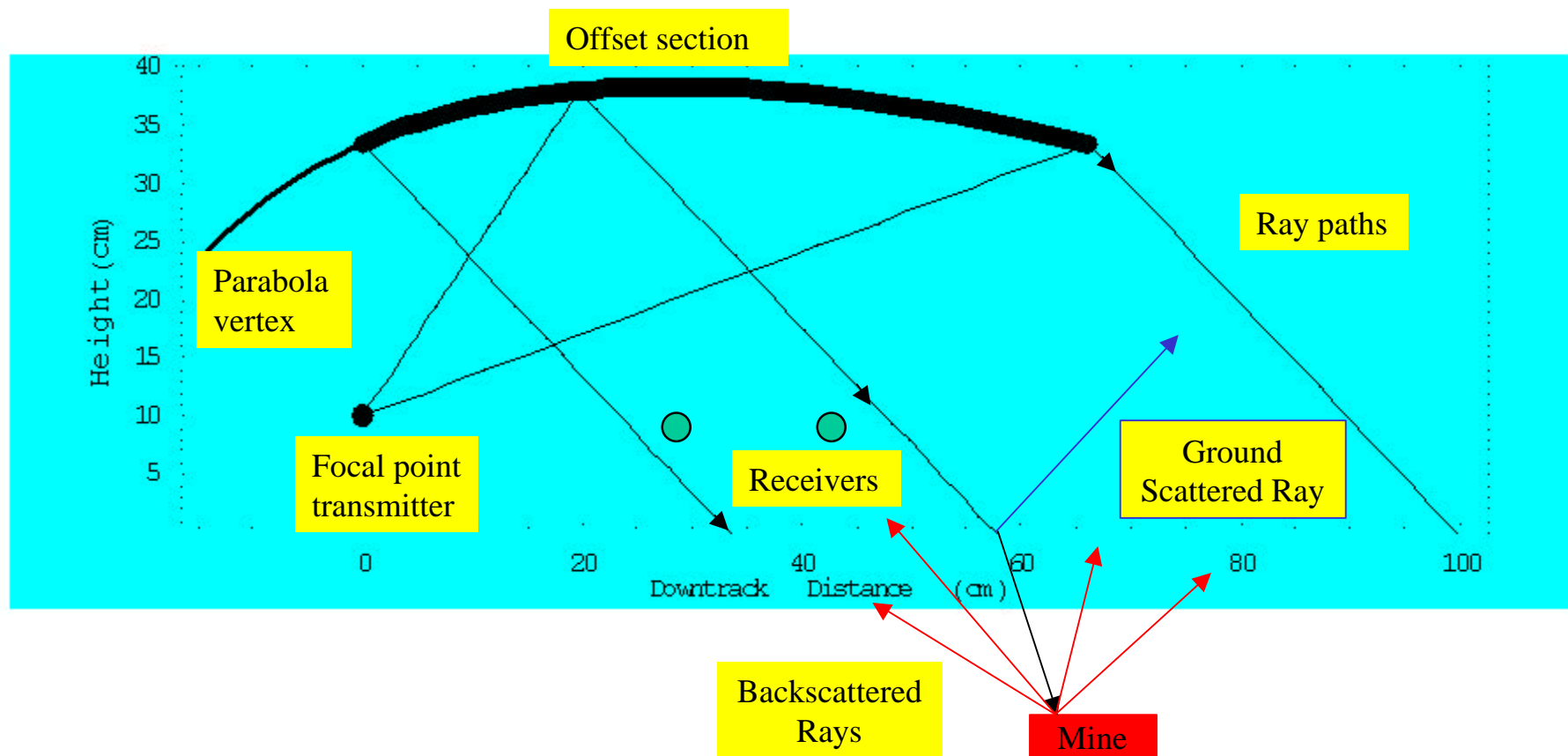
- Specular reflection from the ground depends only on incidence angle, frequency content, and wave polarization, not position or source.
- For flat ground with non-normal incidence, reflection scatters in forward direction.
- Transmitted waves obey Snell's law, the wavefronts remain parallel, and the signal arrives at a given target with the same delay and attenuation due to soil.
- Plane wave excitation generates scattered signals that are easier to distinguish from rough surface clutter.

Paraboloidal Reflector Transmitter

- Parabolas convert circular waves to plane waves.
- GeoCenters, Inc. TEMR wideband antenna element (700 MHz to 1.3 GHz) used to feed offset parabola.
- Collimated beam in the nearfield, is a (non-uniform) plane wave across its aperture.
- Trade-offs between reflector size and grazing incidence angle, -- and height above ground versus forward look -- must be made.

Novel Forward-Looking “Handheld” AP Mine Detector with Plane Incidence Wave Excitation

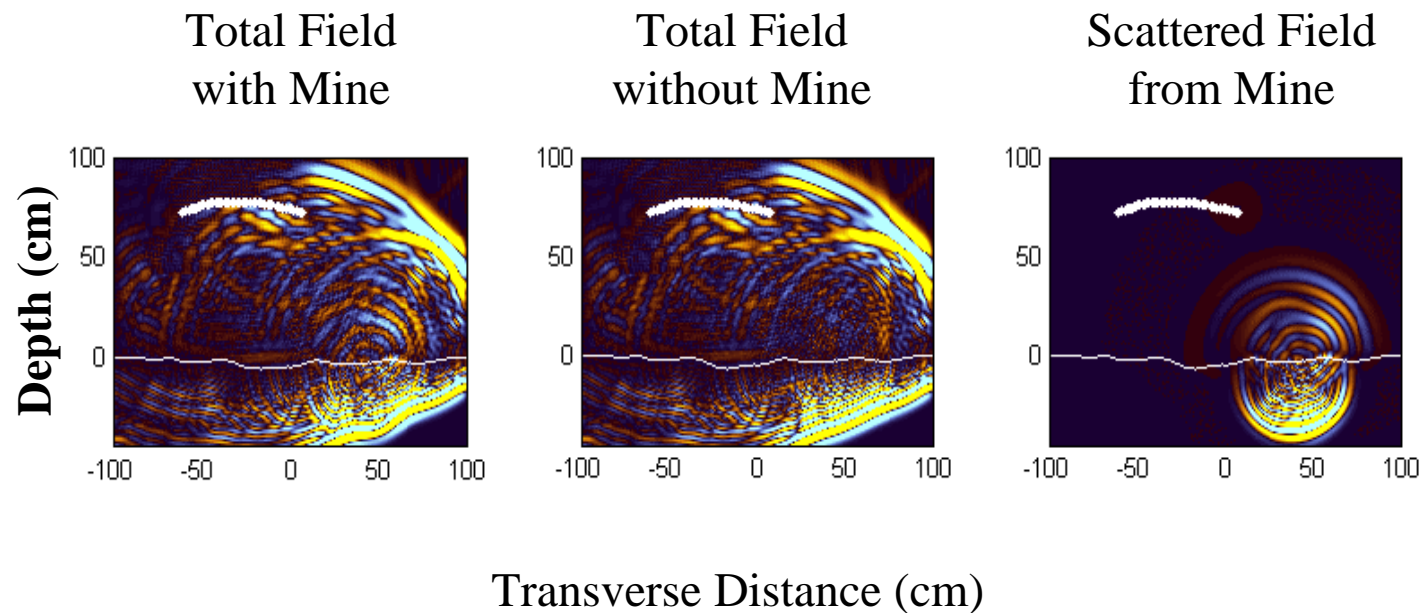
Offset Paraboloïd Ray Paths



Multistatic Receiving Array

- Collimating transmitter is coupled with a four element receiving array.
- 2 X 2 array gives four times the received power of a monostatic radar, as well as temporal focusing left to right and forward to back.
- By comparing time differences of received scattered signals, approach and localization information can be extracted.
- By measuring overall time delay, depth of burial information is available.

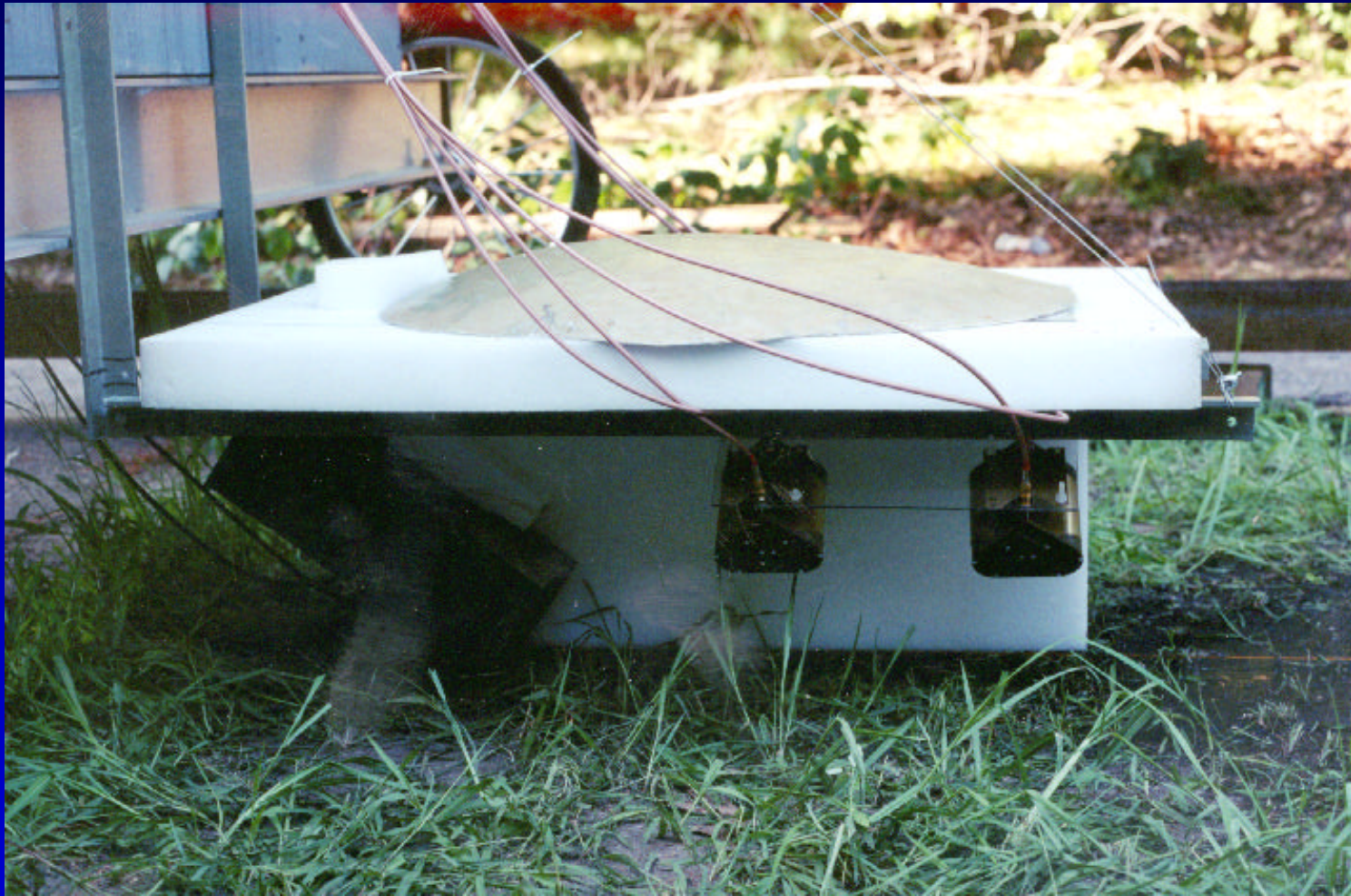
2-D FDTD Simulation of Parabolic Reflector Generated H-Pol. Planar Incident Wave: Non- Metallic Mine 10 cm Below Rough Surface



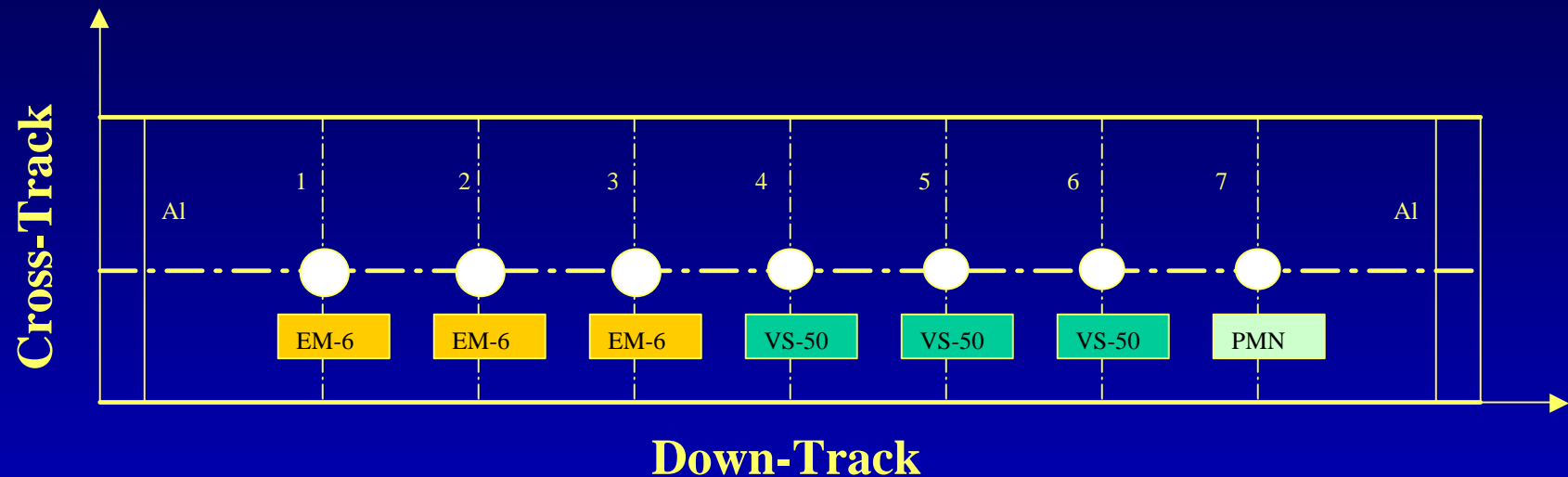
“Handheld” Offset Parabolic / Multistatic Array
Detector Using Geo-Centers TEMR GPR Elements:
Outdoor Data Collection at NEU Dedham Test Track



Outdoor Measurements at NEU Dedham Test Track with Dry Loam / Vegetation



Mine Target Placement Map of Experimental Test Track



EM-6: Cylindrical, 150mm diameter, 50 mm high.

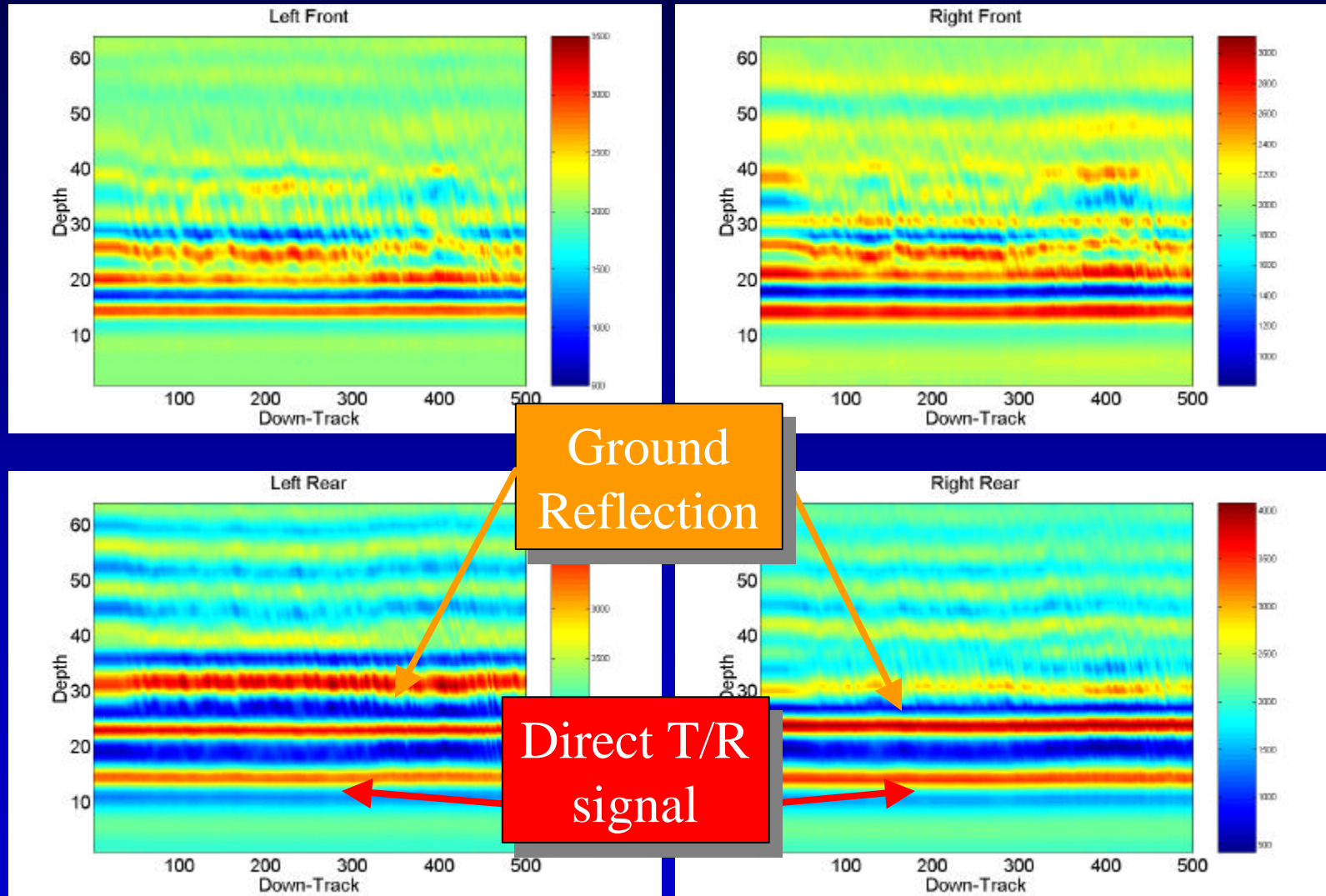
VS-50: Cylindrical, 90mm diameter, 45mm high, plastic cased.

PML: Cylindrical, 112mm diameter, 56mm high, plastic/bakelite/rubber.

Northeastern University Dedham Test Track
September 1999

Measured Signals at the Receivers

Time Step



Down Track Position (in.)

Data Analysis

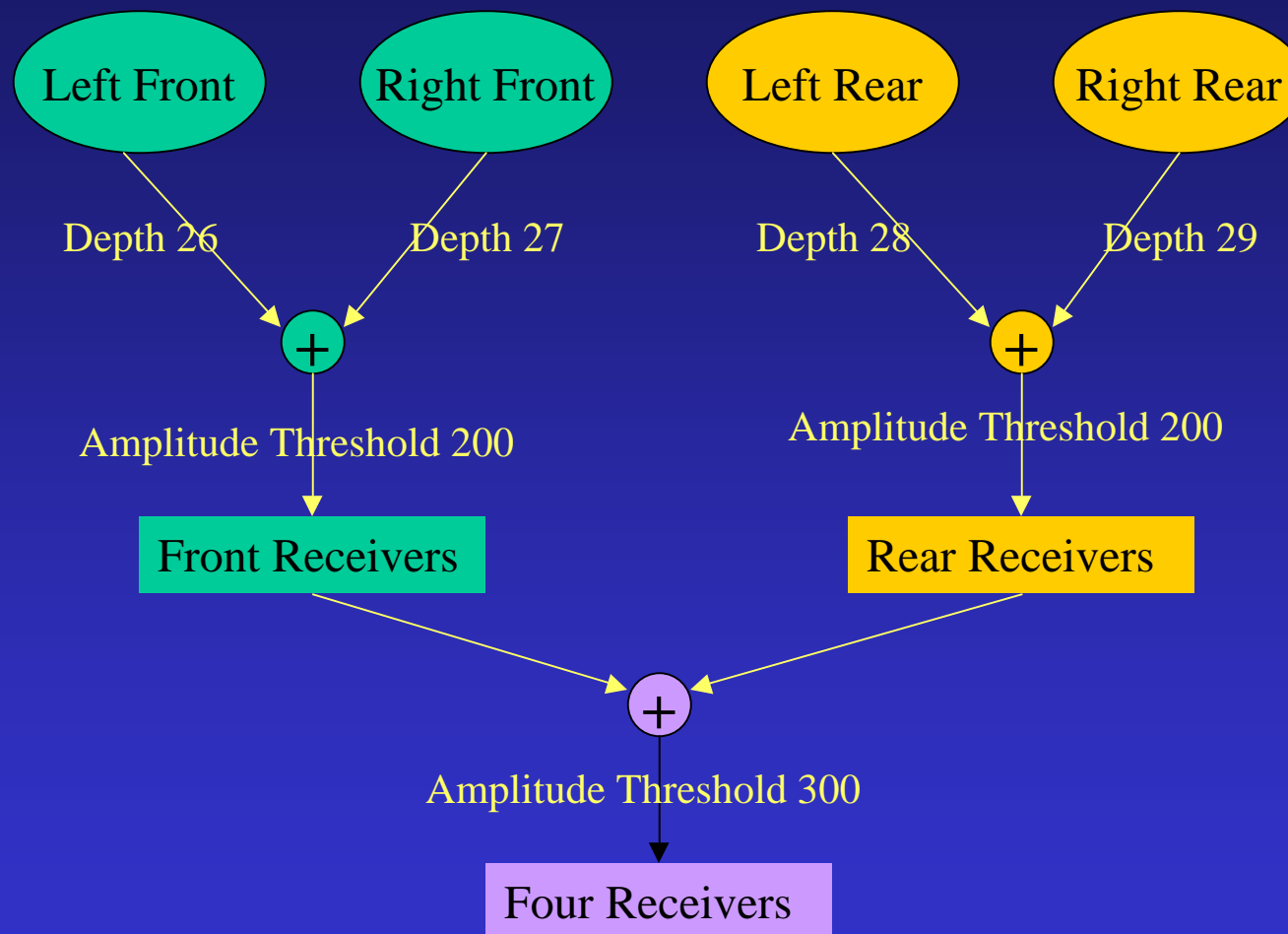
1. Remove soil clutter from the total field to obtain mine signals.

- Ideal ground signal
- Total average ground signal
- Moving average ground signal

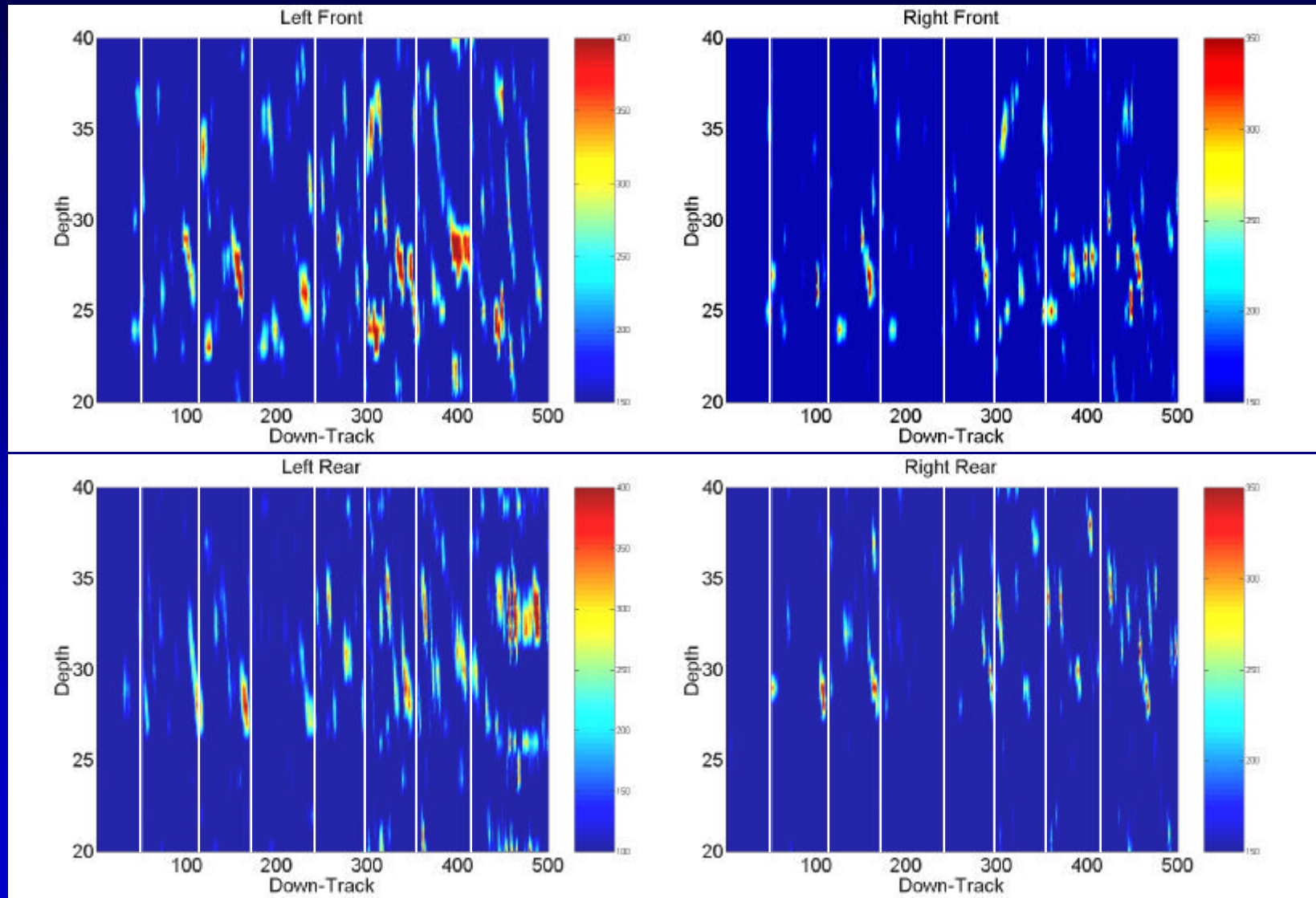
1. Evaluate the performance of different receivers.

2. Synchronize the mine signals at different receivers to obtain strong mine signals.

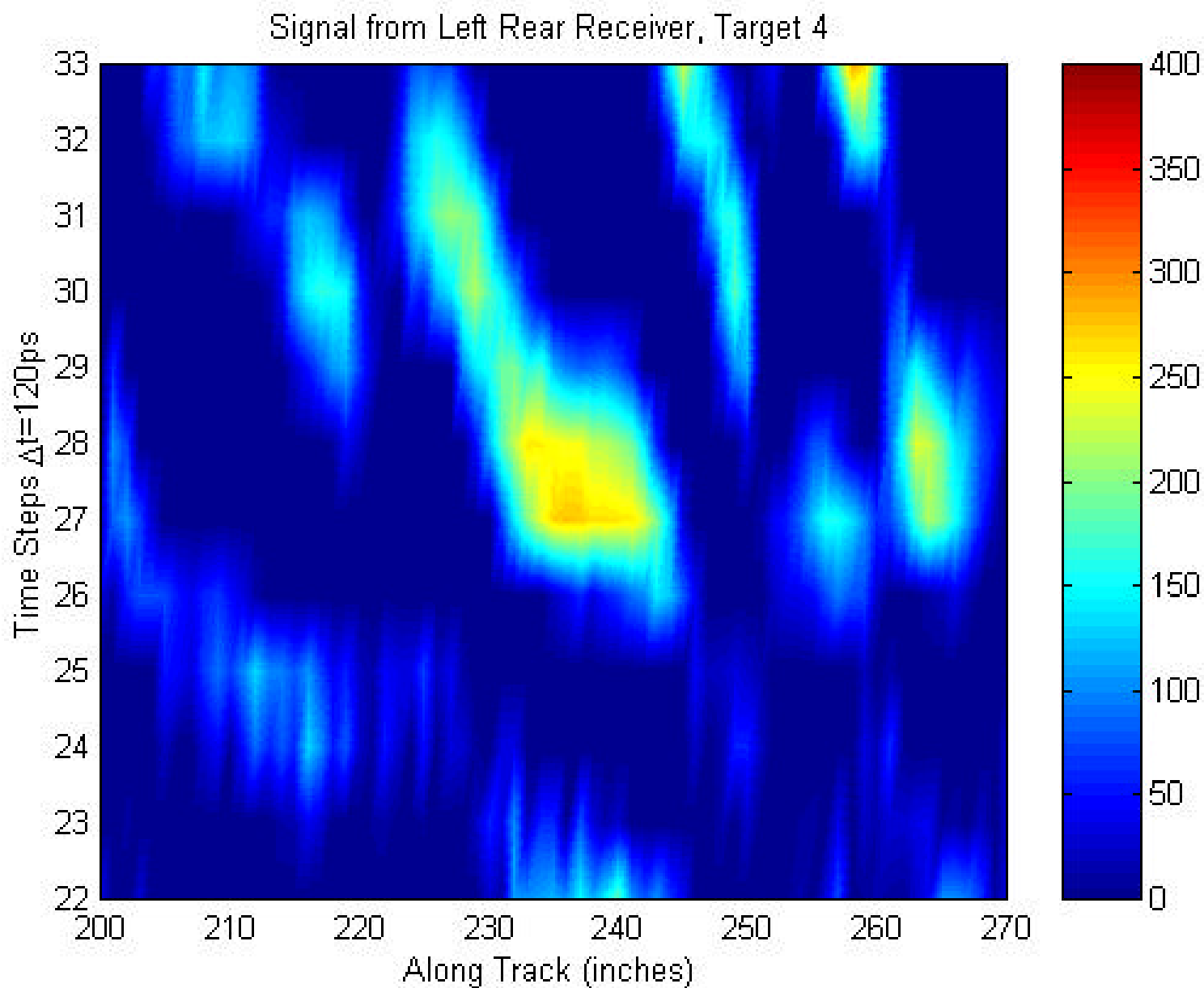
Mine Signals Synchronization and Summation



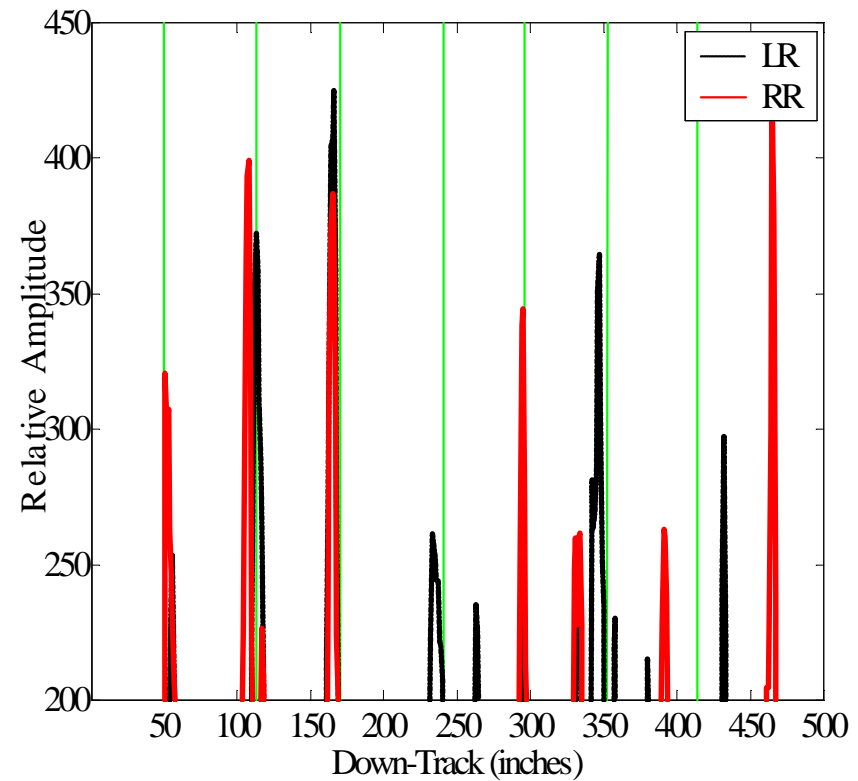
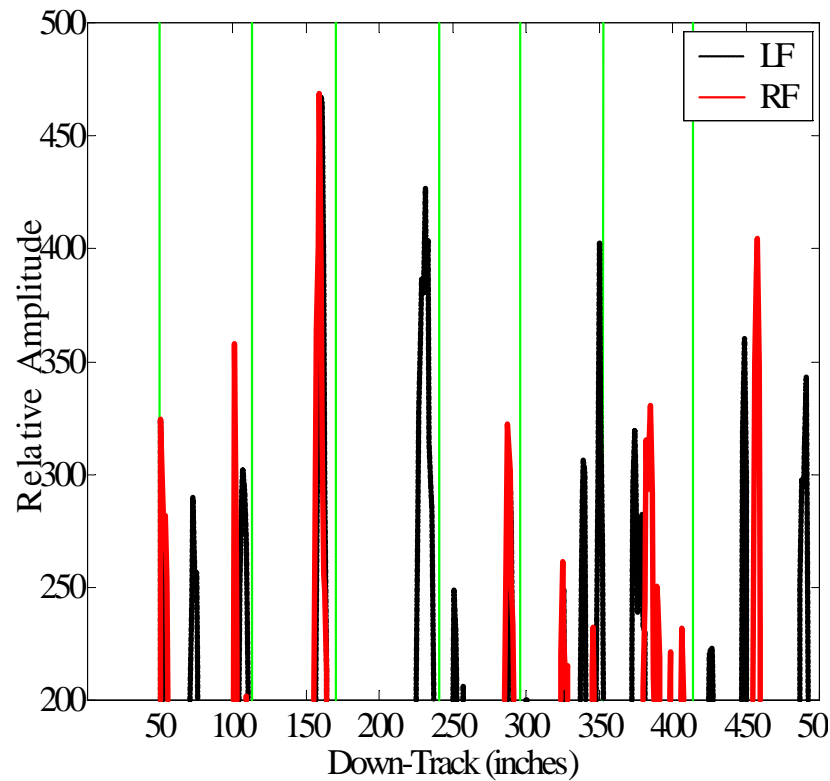
Mine Signals



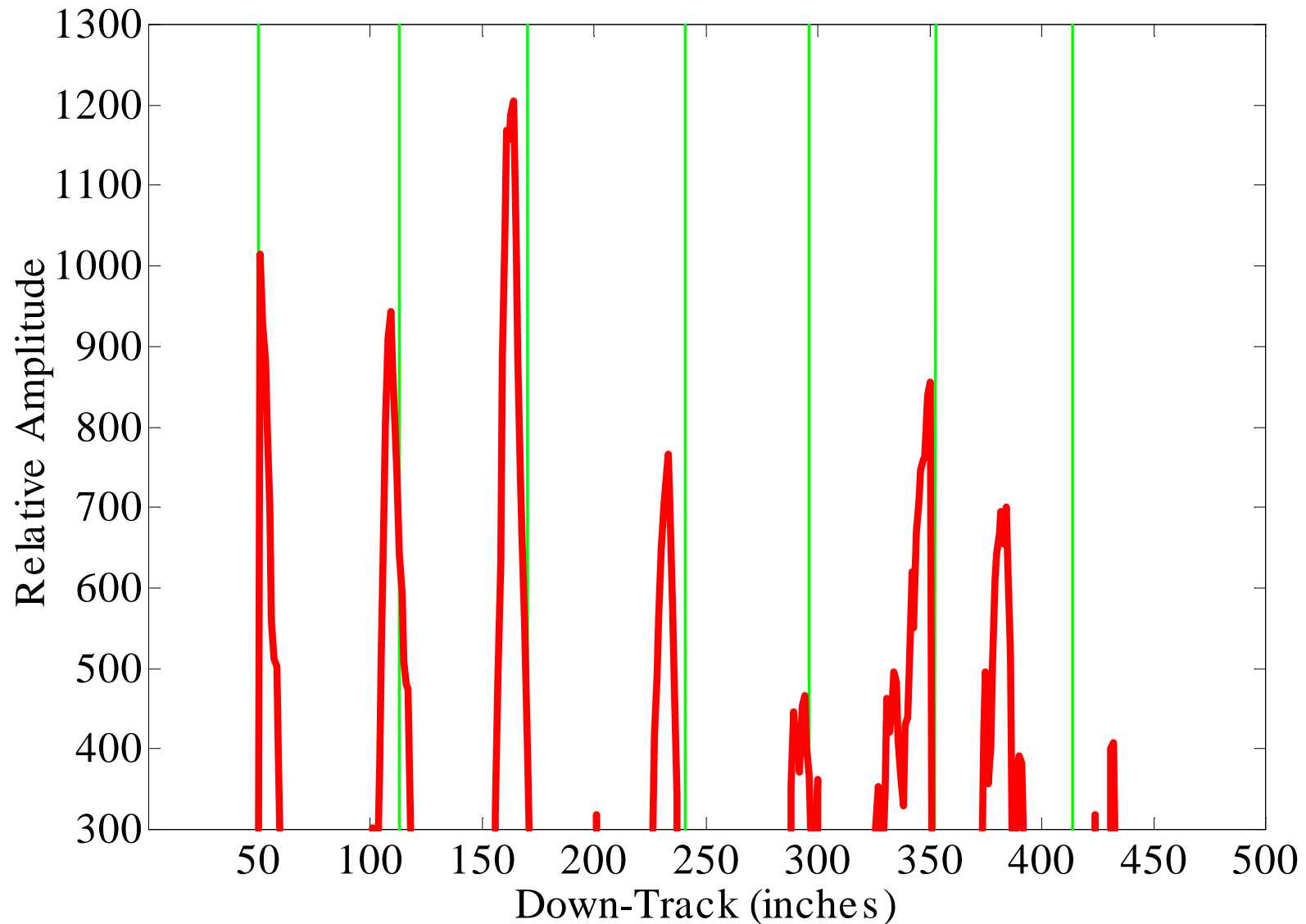
Low Intensity Feature with Distinctive Shape (to be exploited with future processing)



Mine Signals at Specific Depths



Summation of Four Receivers



Factors that Effect Performance

- Mines are of different types.
- Mine burial depth and orientation vary.
- Soil moisture and ground surface roughness change along the track.
- Radar pulse width and sensing time window are limited.

Conclusions

Initial experiments indicate that the parabolic antenna mine detector performs very well in detecting shallow buried nonmetallic mines in relatively dry soil.

The sensor gives strong mine signals at locations that agree with the locations of the buried targets with only a few false alarms and detection failures.

FDTD Simulation of a Parabolic Antenna GPR Transmitter



Dongping Jin and Carey Rappaport
ECE Dept.
Northeastern University
Boston MA

PIERS Cambridge MA, July 7, 2000

Outline

- Forward-looking, light-weight AP mine detector.
- Plane wave vs. point source impulse excitation for rough ground clutter minimization.
- Offset paraboloidal reflector as wideband collimator for handheld detector.
- FDTD modeling.
- Experimental results.

* Demining MURI Supported by ARO

Problem: Rough Ground Surface is Greatest Contributor to GPR Clutter

- For standoff GPR detection of small antipersonnel mines, the largest scattered signal comes from the ground surface.
- In the real world, the ground is never flat, the soil type is never known exactly, and the height of the antenna is never fixed.
- This clutter can be reduced by:
 - Time gating
 - Signal processing (matched filtering, deconvolving)
 - Spatially separating target/signal

Plane Wave Impulse Excitation

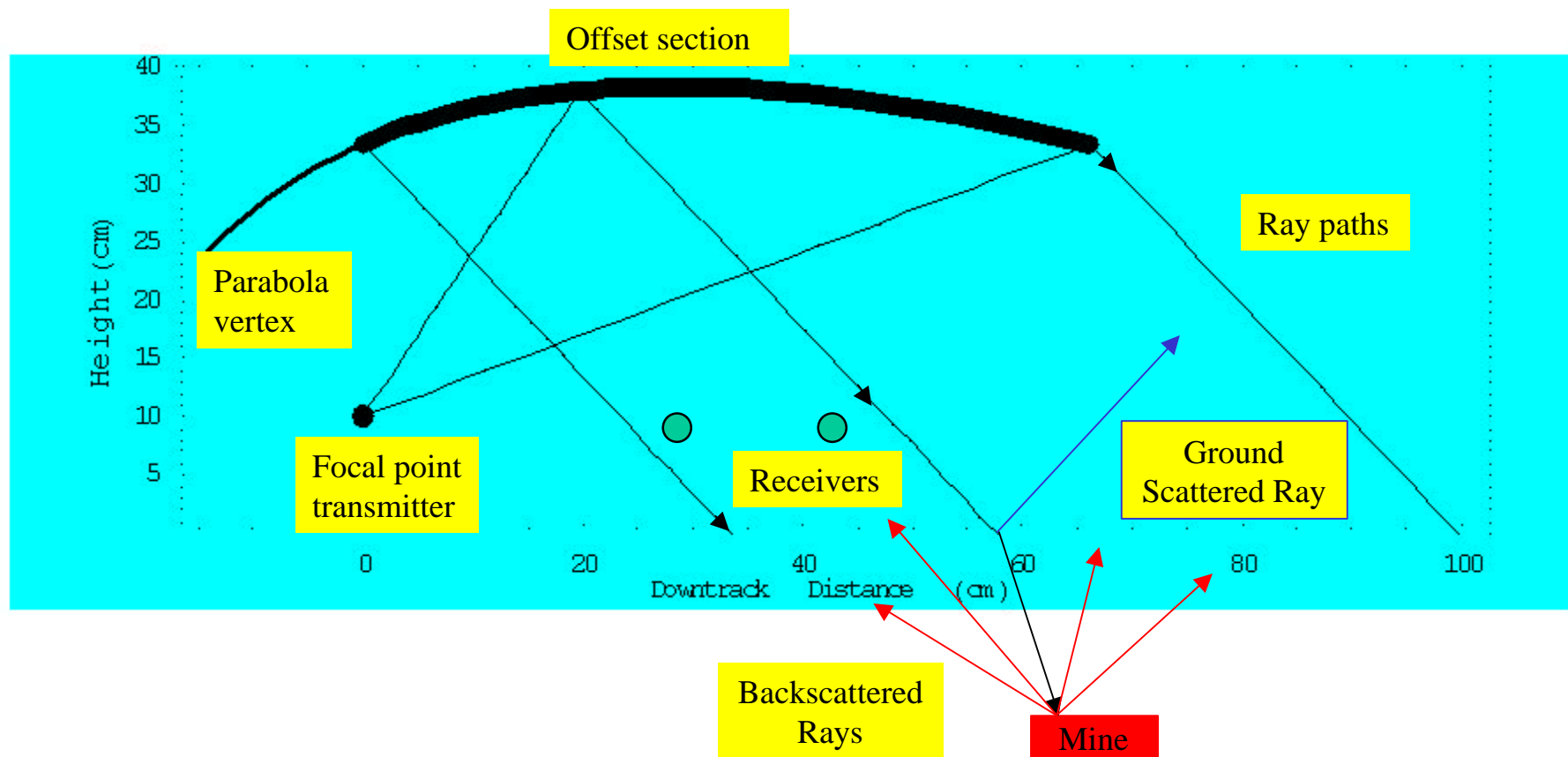
- Specular reflection from the ground depends only on incidence angle, frequency content, and wave polarization, not position or source.
- For flat ground with non-normal incidence, reflection scatters in forward direction.
- Transmitted waves obey Snell's law, the wavefronts remain parallel, and the signal arrives at a given target with the same delay and attenuation due to soil.
- Plane wave excitation generates scattered signals that are easier to distinguish from rough surface clutter.

Paraboloidal Reflector Transmitter

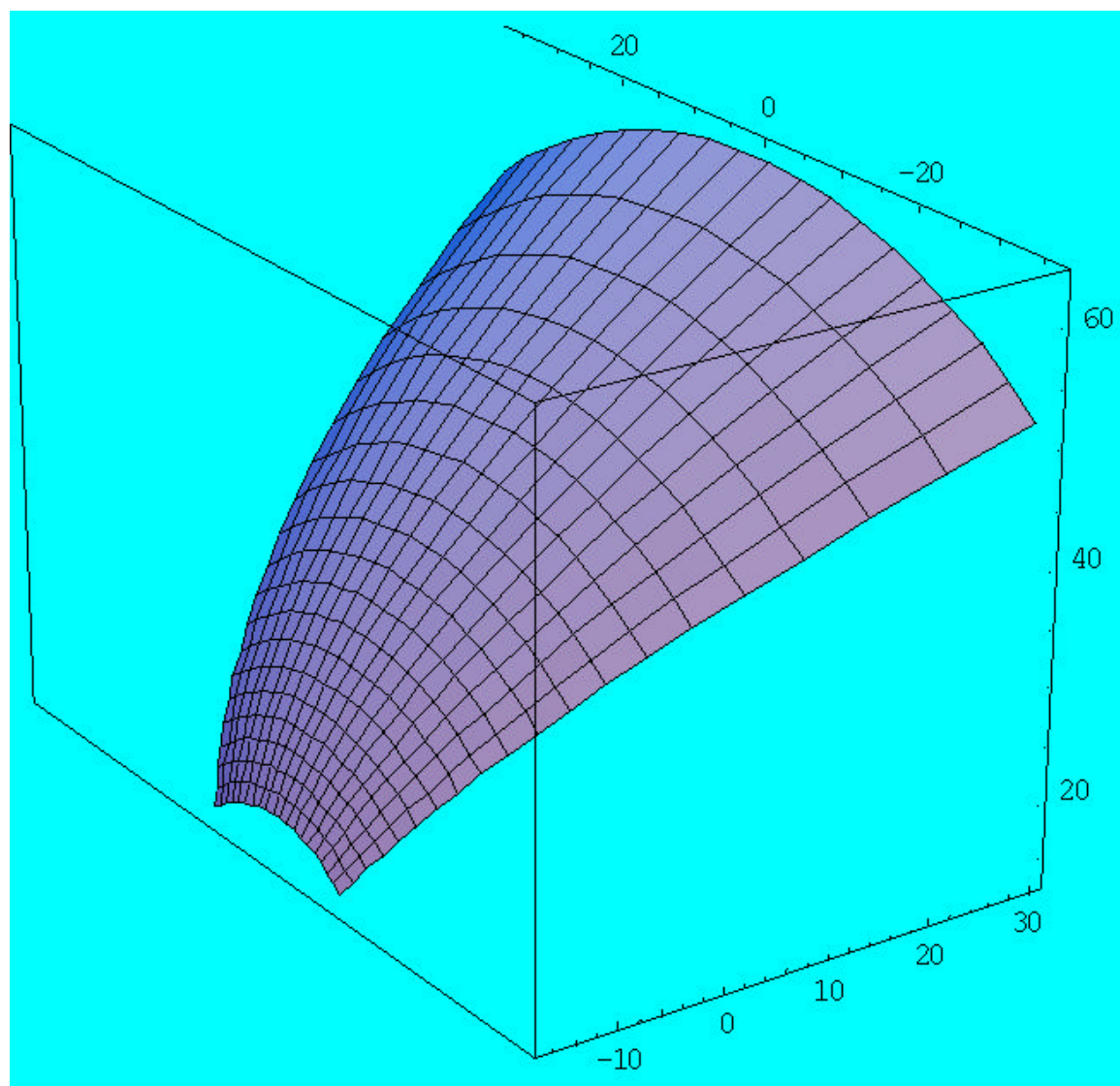
- Parabolas convert circular waves to plane waves.
- Collimation is based on parabolic geometry, independent of frequency.
- In the nearfield, a paraboloidal reflector generates a (non-uniform) plane wave across its aperture.
- An offset paraboloidal section must be used to avoid blockage and scattering by transmitting feed.
- Trade-offs between reflector size and grazing incidence angle, -- and height above ground versus forward look -- must be made.

Novel Forward-Looking “Handheld” AP Mine Detector with Plane Incidence Wave Excitation

Offset Paraboloid Ray Paths

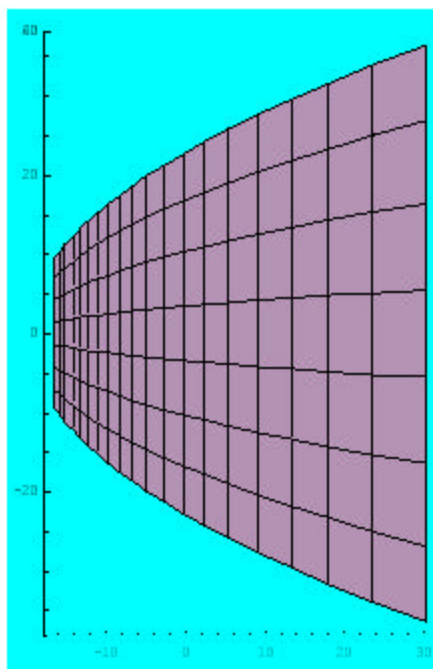


Portion of Paraboloid of Revolution

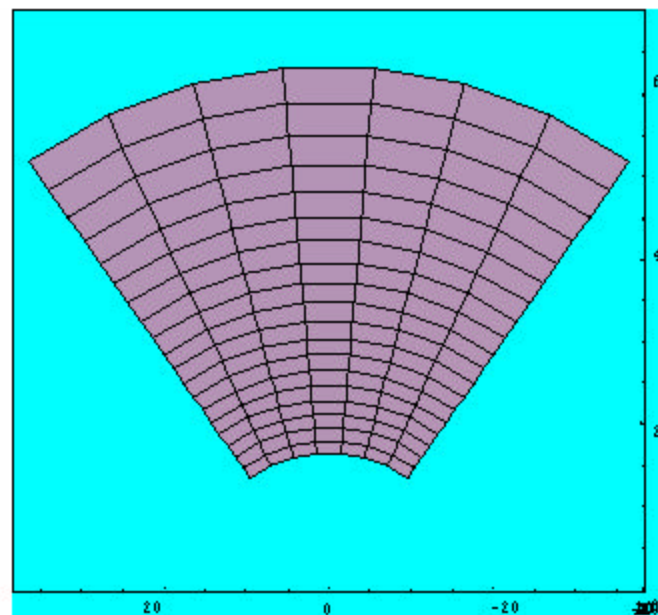


Handheld Offset Parabololoid

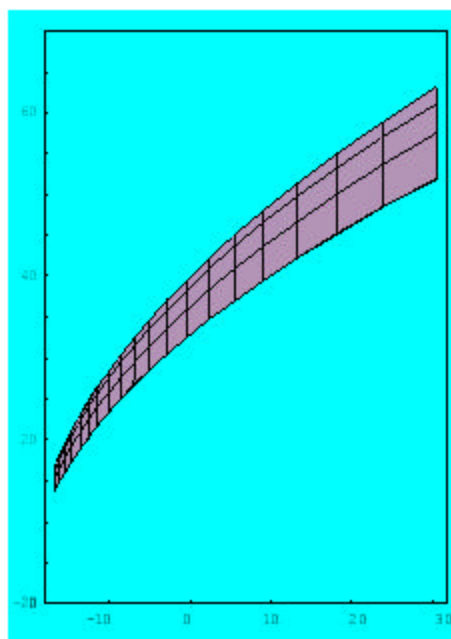
Top View



Front View



Side View



Propagation in Soil is Frequency Dependent

$$k(f) = b(f) - ja(f) = \frac{2\pi f}{c} \sqrt{\epsilon'(f) - j \frac{\sigma(f)}{2\pi f \epsilon_0}}$$

**Neglecting the frequency
dependence gives inaccurate results.**

Modeling Dispersion for Easy Transformation to Time Domain

Debye (Teixeira)

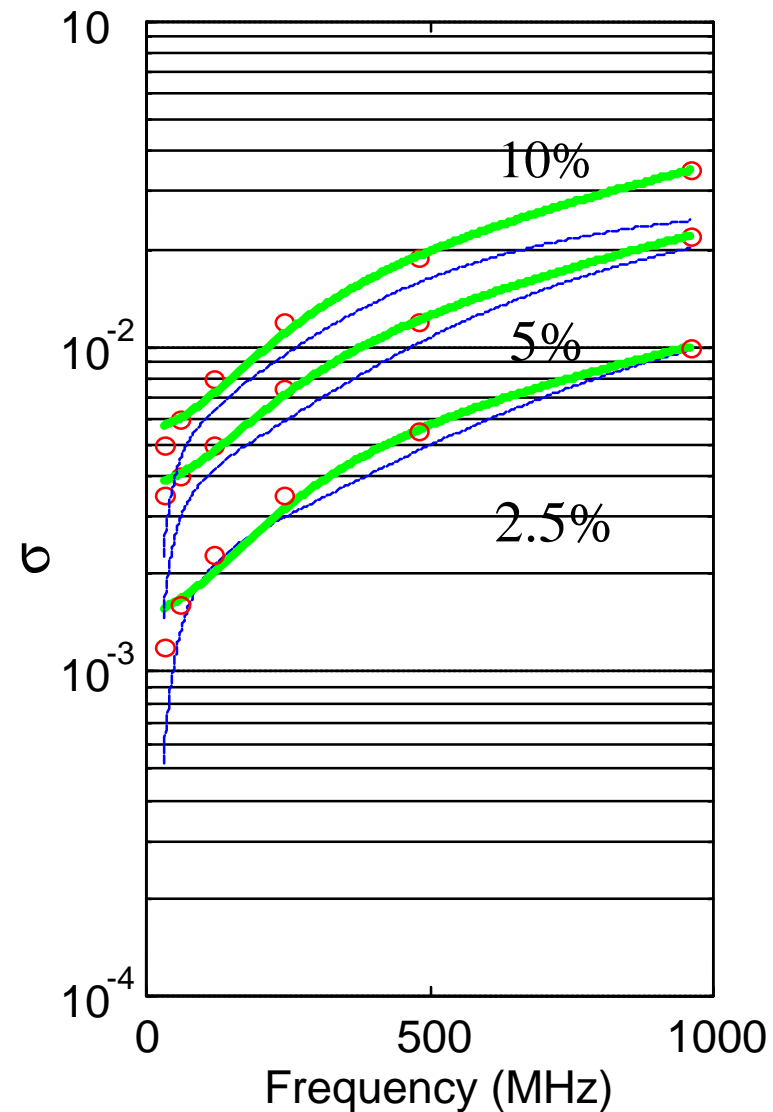
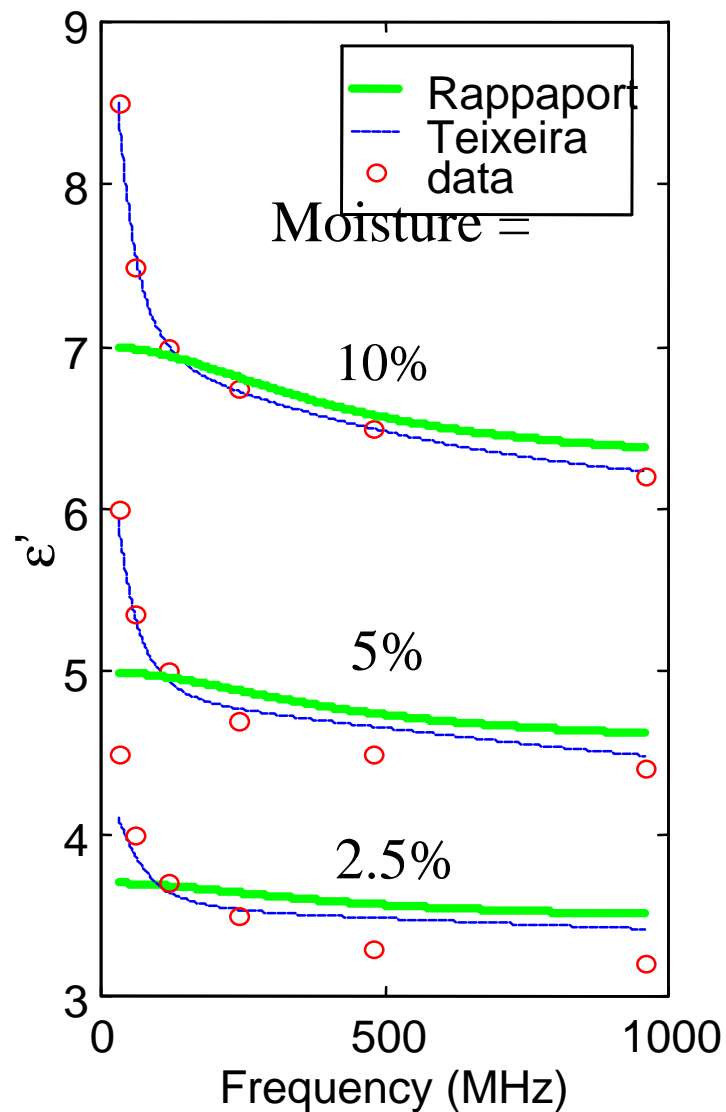
$$\epsilon = \epsilon_0 \epsilon_\infty + \epsilon_0 \sum_{p=1}^N \frac{A_p}{1 - i\omega\tau_p} \quad N=2$$

Pade' (Rappaport)

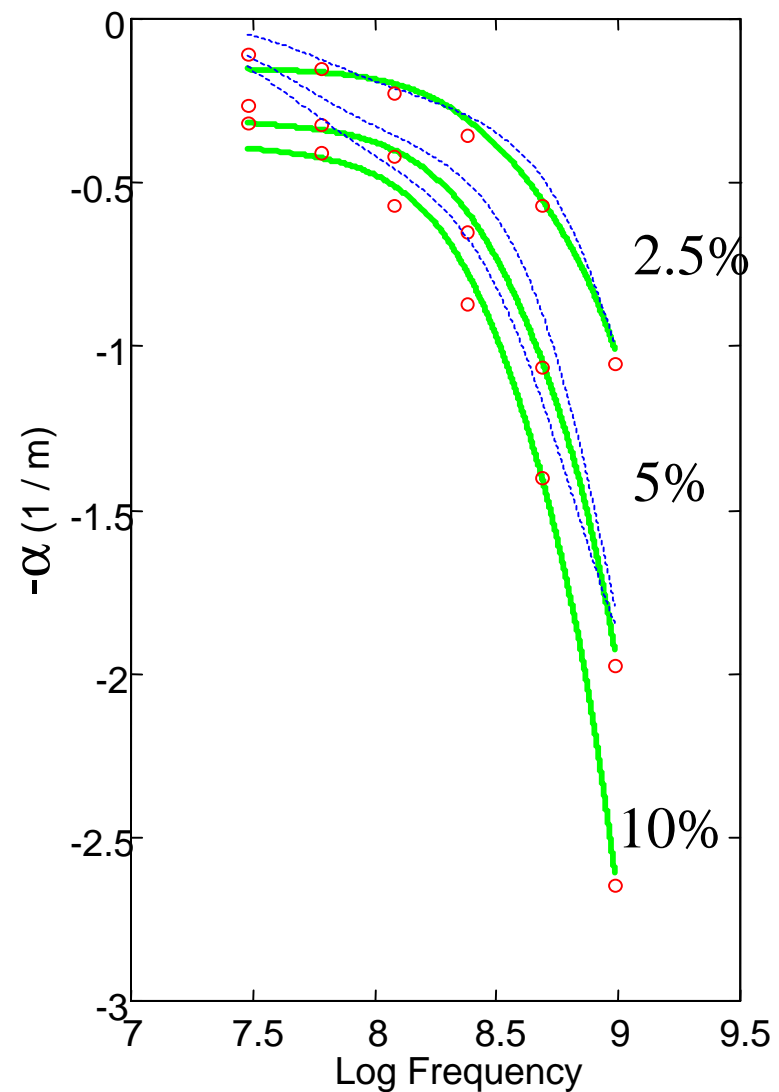
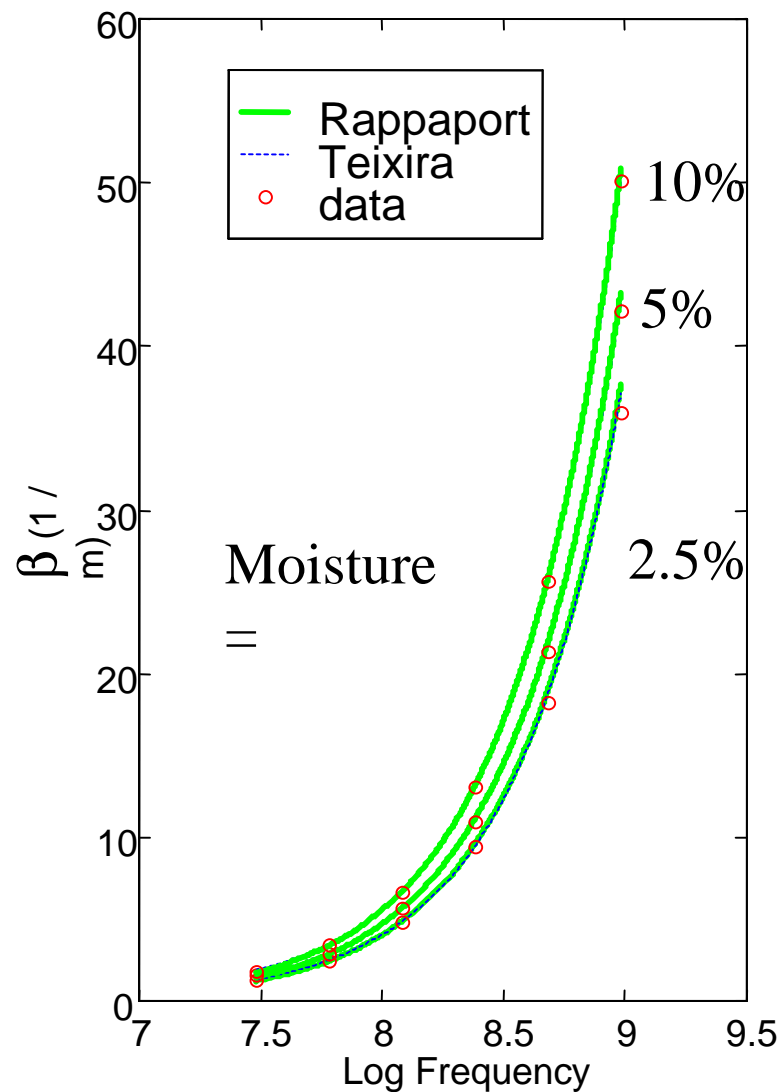
$$\epsilon' = \text{Constant}, \quad s = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2}}{1 + a_1 Z^{-1}}, \quad Z = e^{i\omega D t}$$

M o i s t u r e	2.5 %	5 %	10 %
b 0	. 5 8 1	1 . 3 6	2 . 1 5
b 1	- 1 . 1 2	- 2 . 6 4	- 4 . 1 8
b 2	. 5 4 6	1 . 2 8	2 . 0 3
a 1	- 0 . 9 5	- 0 . 9 5	- 0 . 9 5
ϵ'	2 . 1 8	1 . 5 2	1 . 4 6

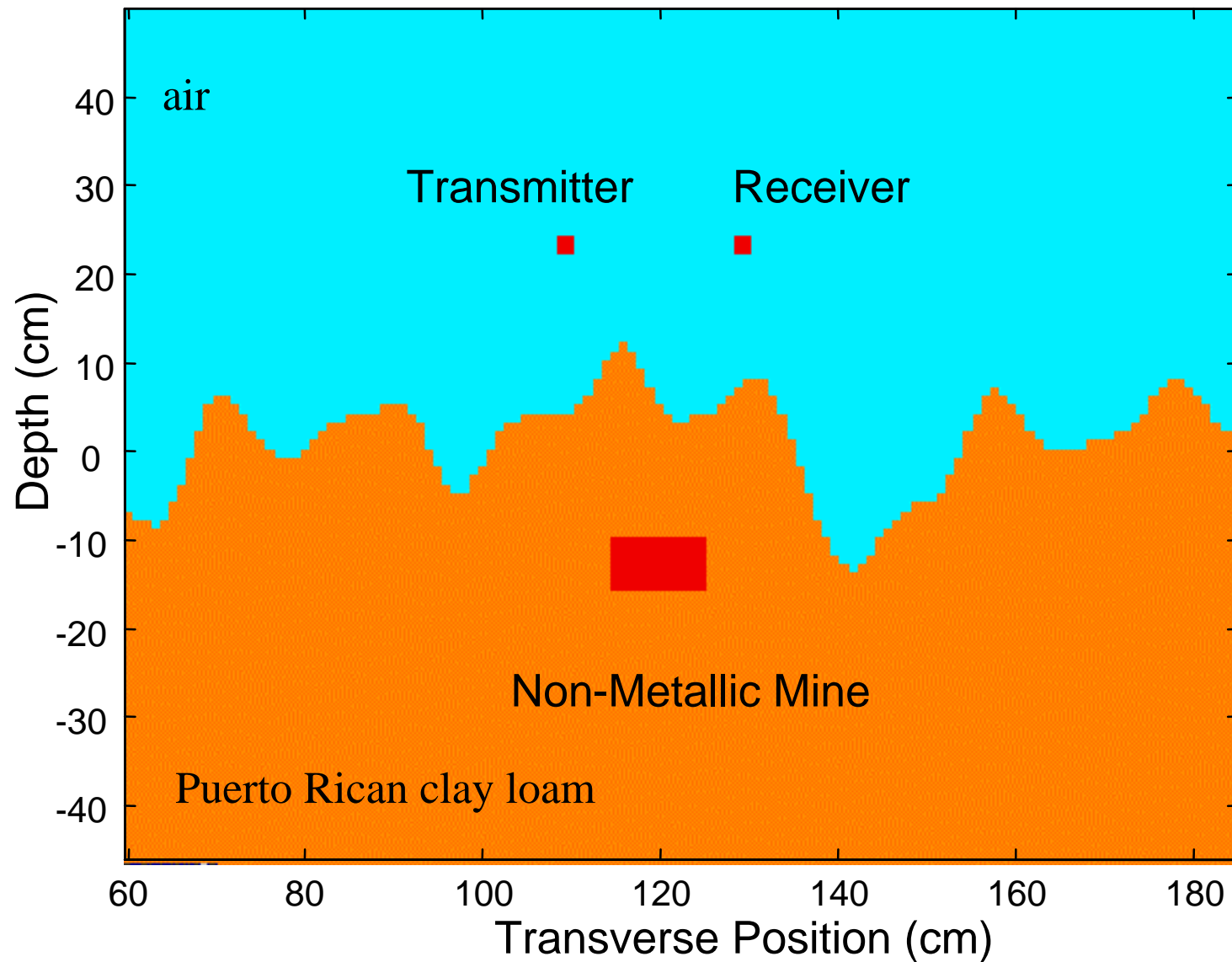
Dielectric Constant and Conductivity for Puerto Rican Clay Loam (1.2 g/cc)



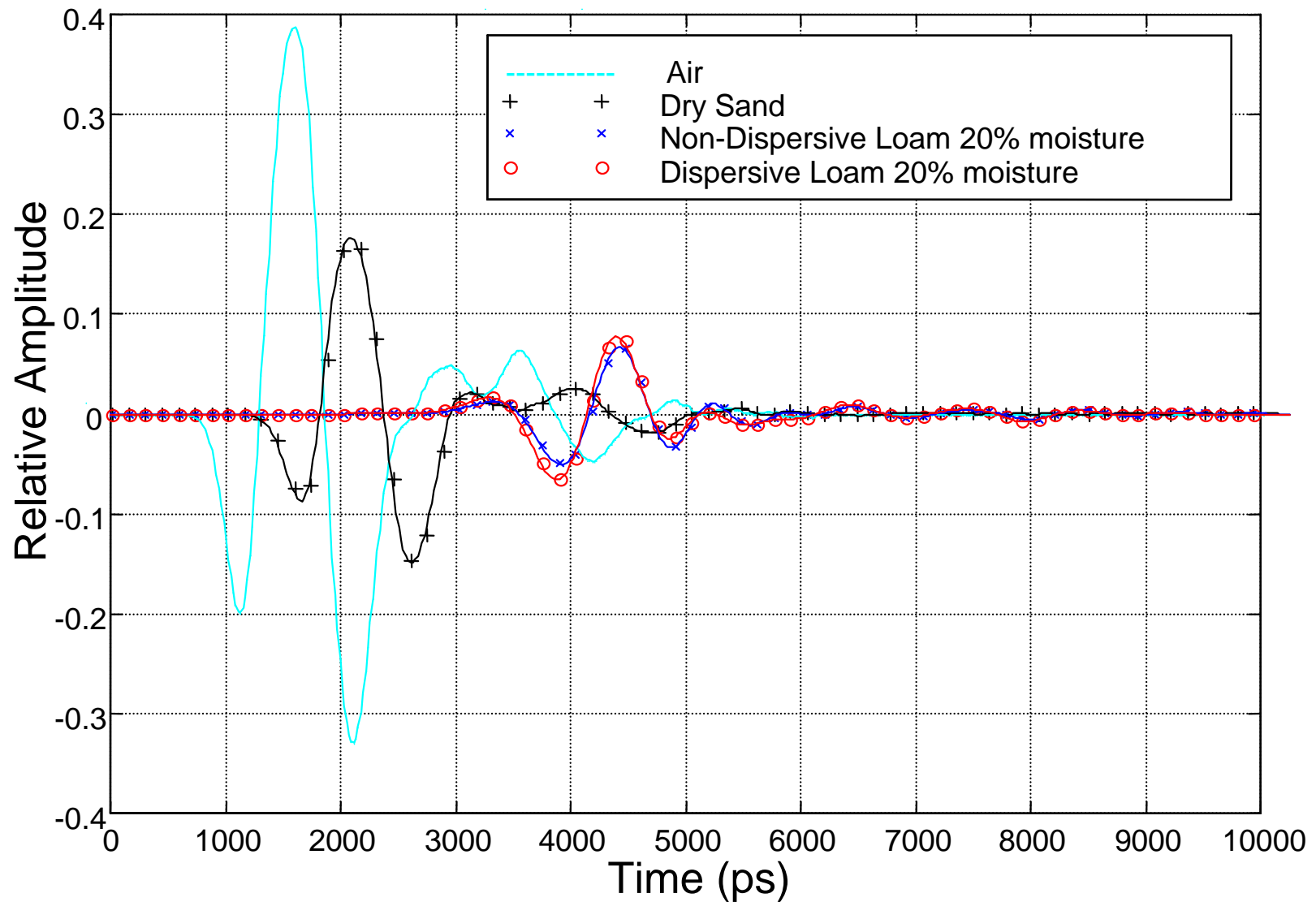
Real and Imaginary Wave Number for Puerto Rican Clay Loam (1.2g/cc)



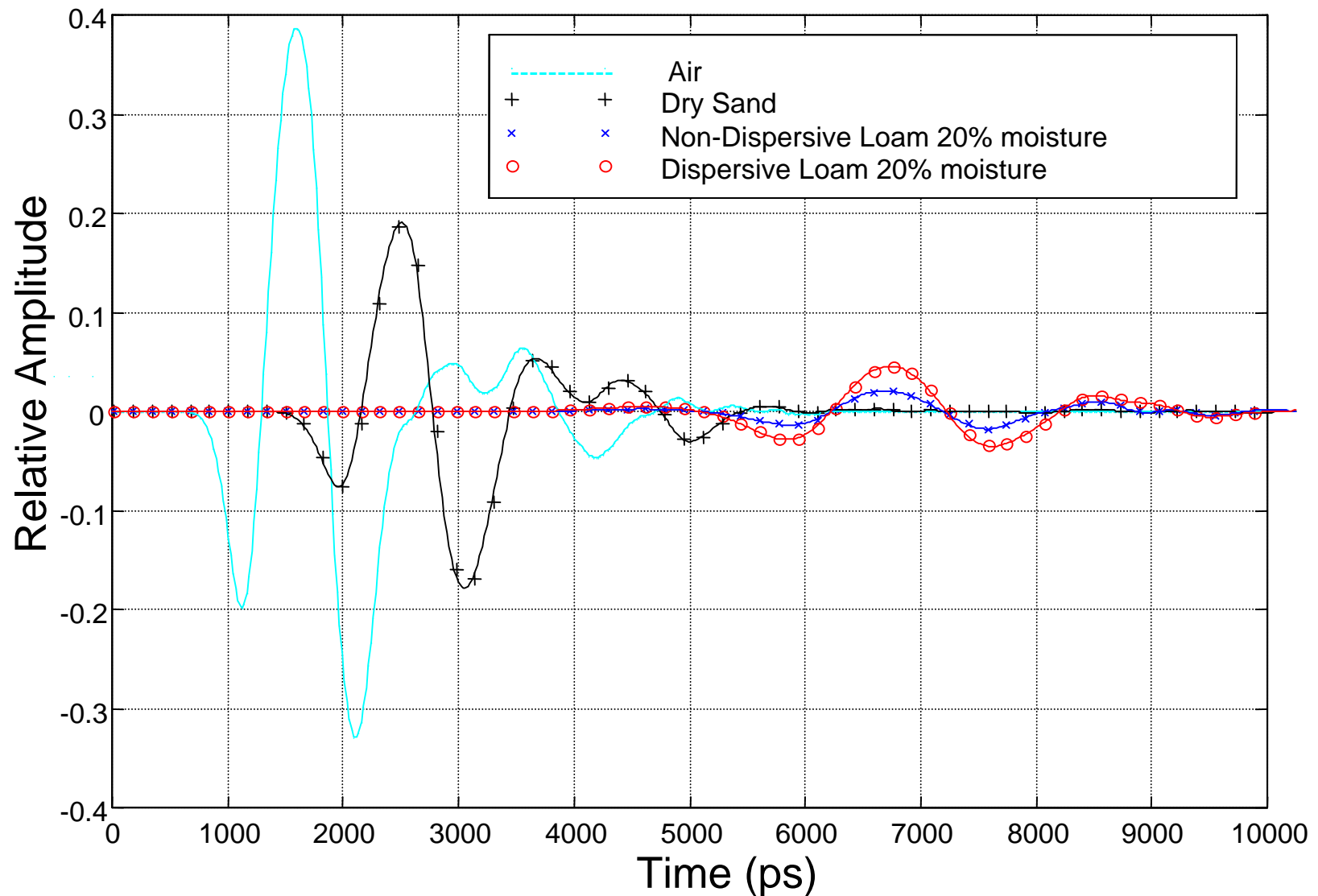
Dispersive Soil Example Geometry



Non-Metallic Mine Scattered Field (about 10 cm burial) - Smooth Surface

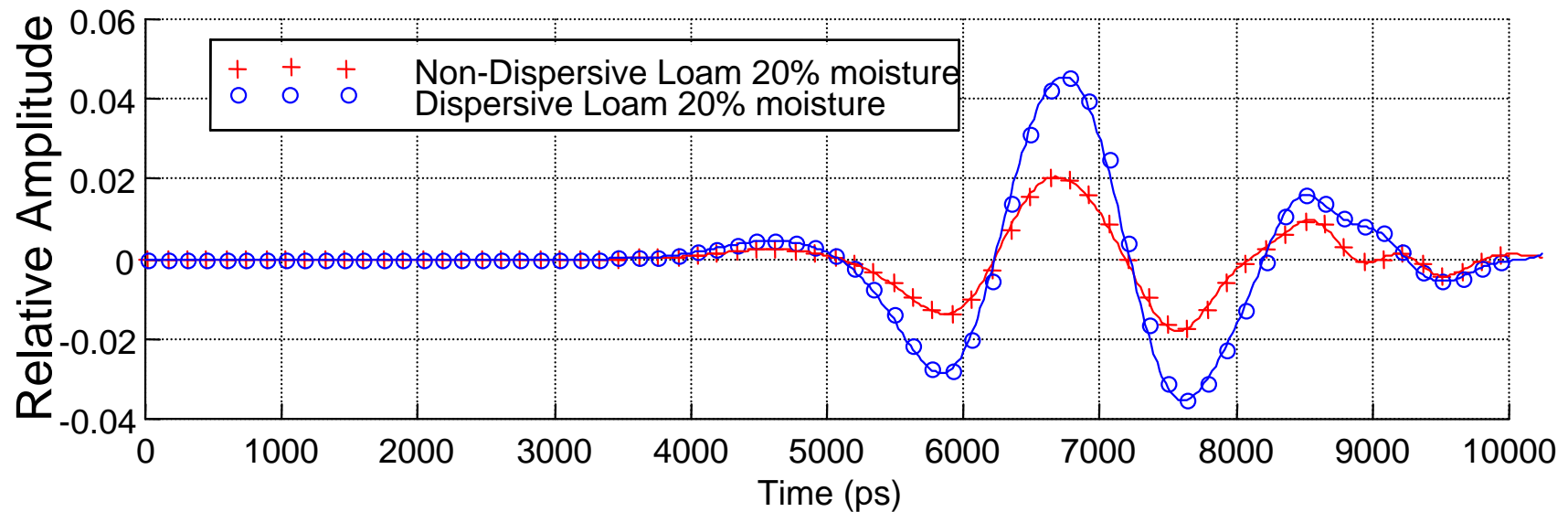
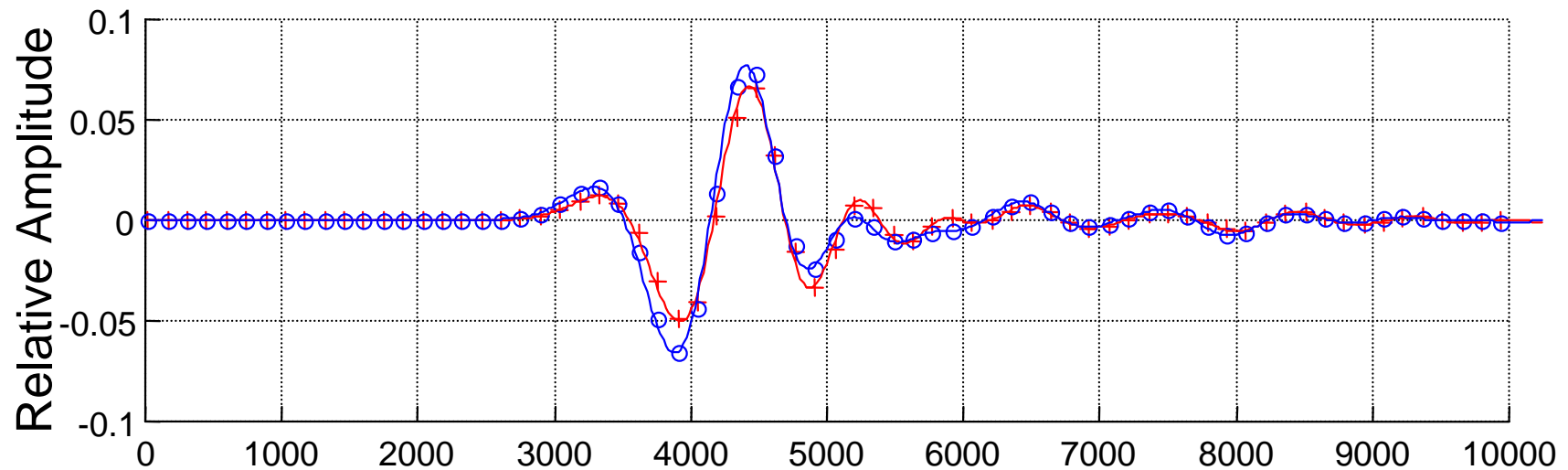


Non-Metallic Mine Scattered Field (about 10 cm burial) - Rough Surface

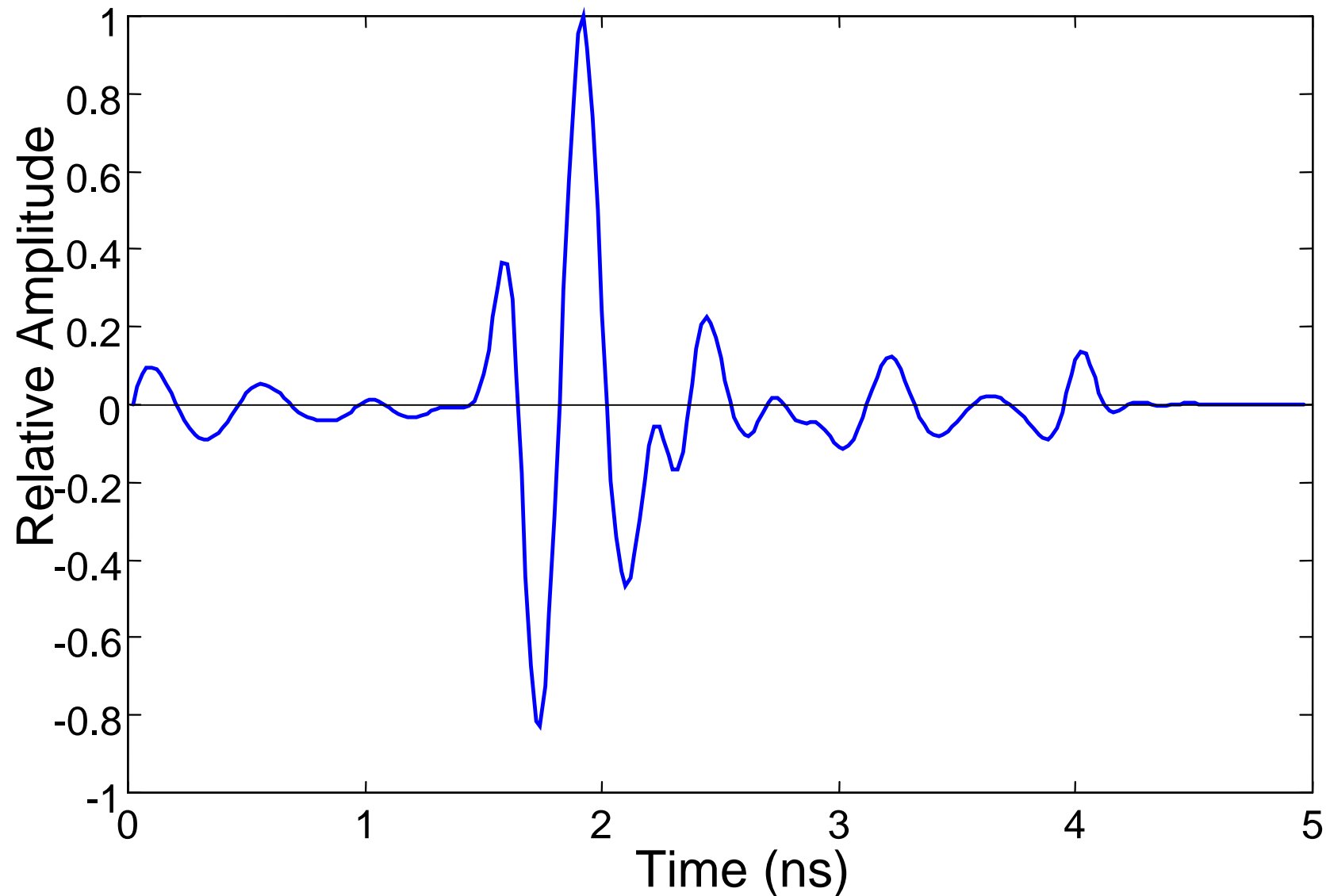


Non-Metallic Mine Scattered Field (10 cm depth)

a) Smooth Surface, b) Rough Surface

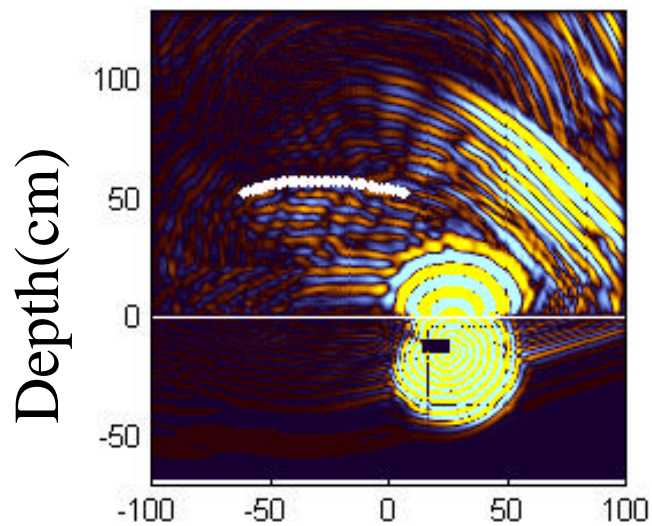


Numerically Modeling the Experiment: LLNL Micro-power Impulse Radar Transmitted Pulse

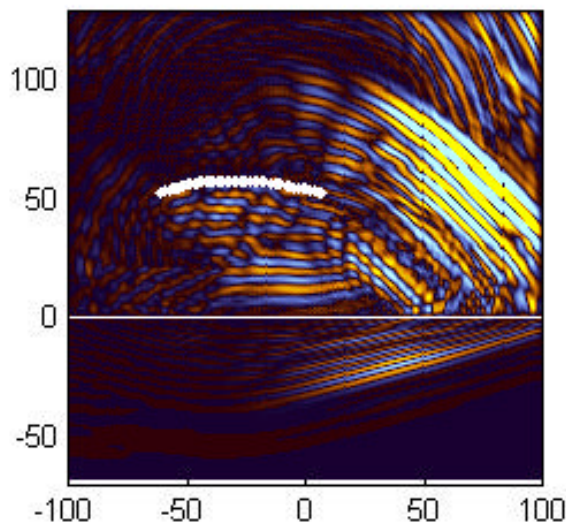


TM: with mine 10 cm below Smooth surface

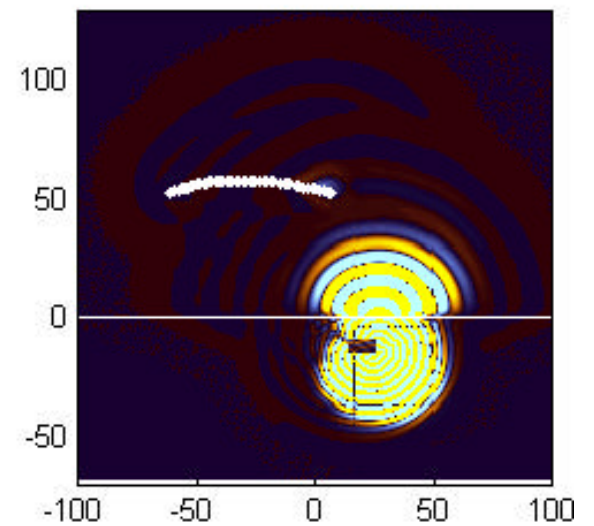
Total field with mine



Total field without mine



Scattered field from mine



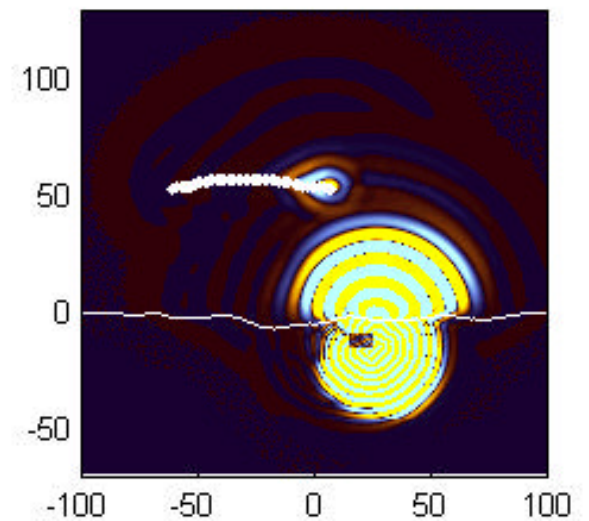
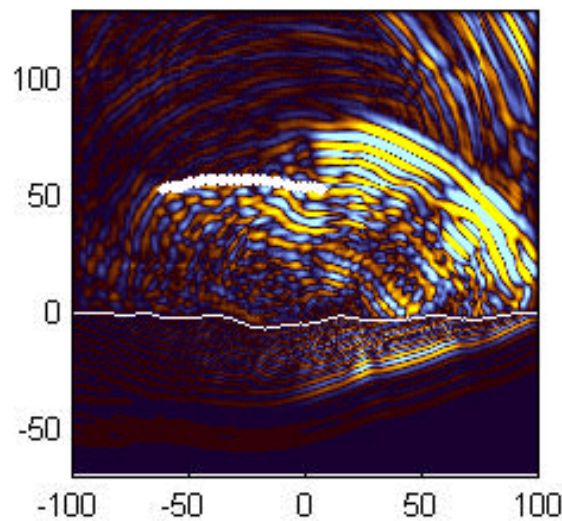
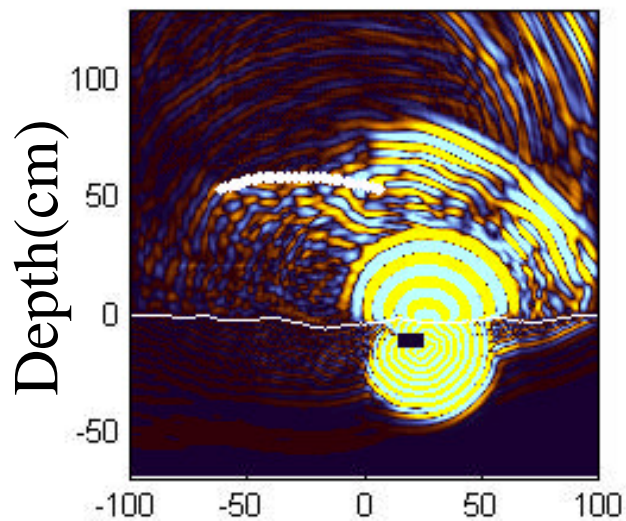
Transverse distance (cm)

TM: with mine 10 cm below **Rough** surface

Total field with mine

Total field without mine

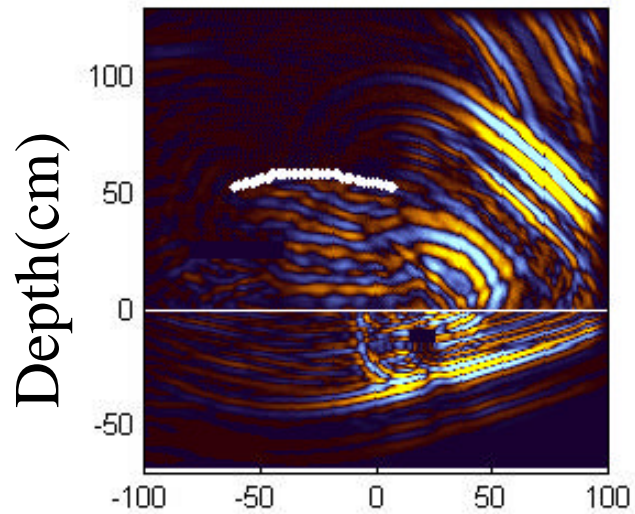
Scattered field from mine



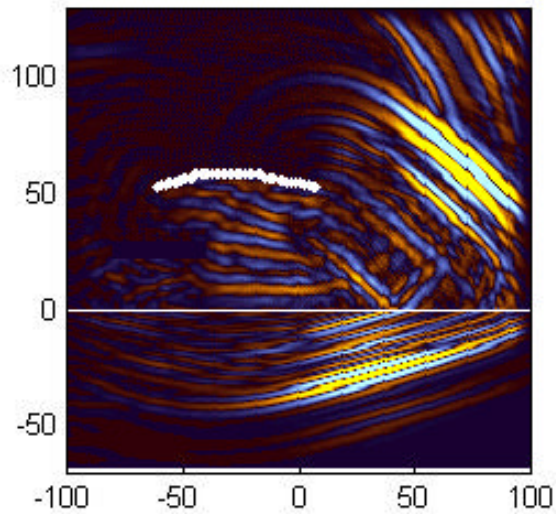
Transverse distance (cm)

TE: with mine 10 cm below Smooth surface

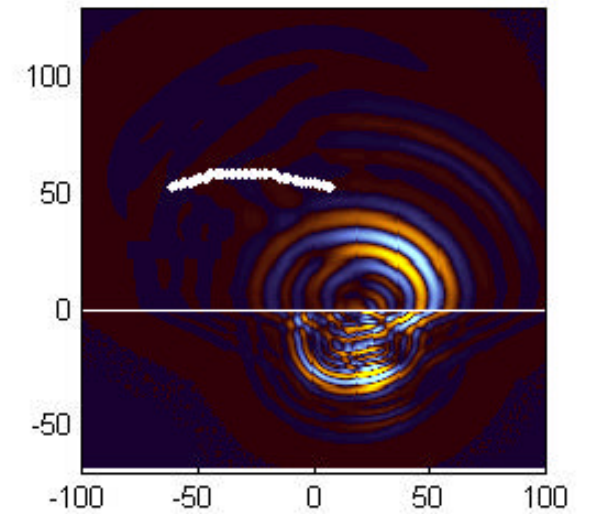
Total field with mine



Total field without mine



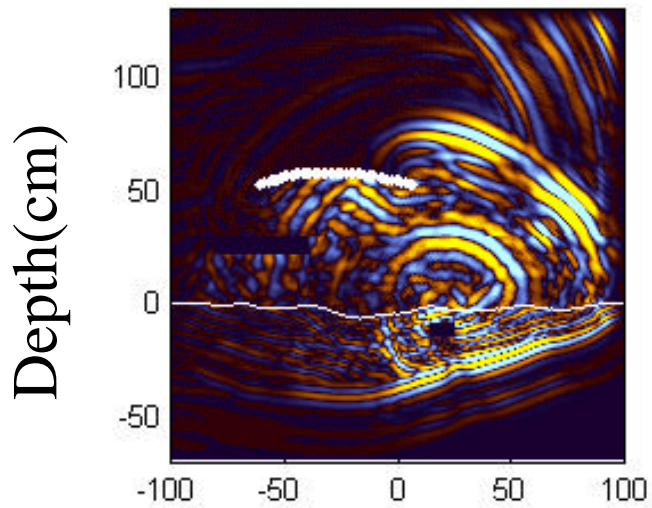
Scattered field from mine



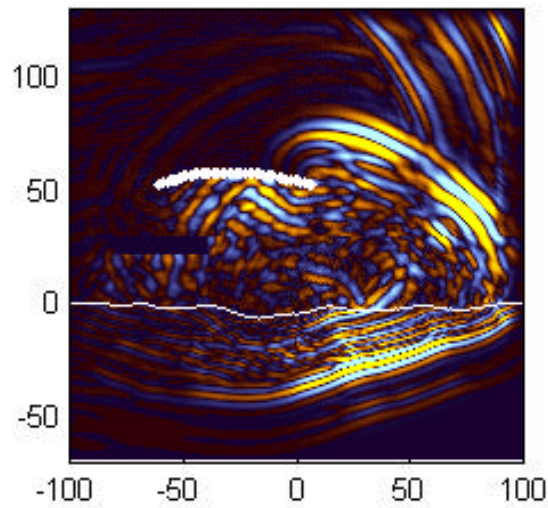
Transverse distance (cm)

TE: with mine 10 cm below **Rough** surface

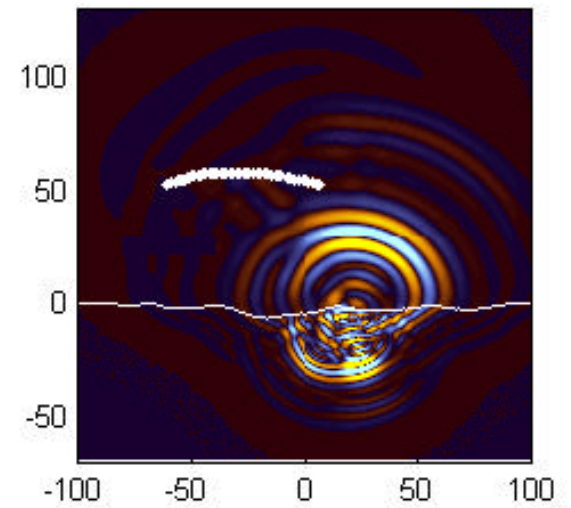
Total field with mine



Total field without mine

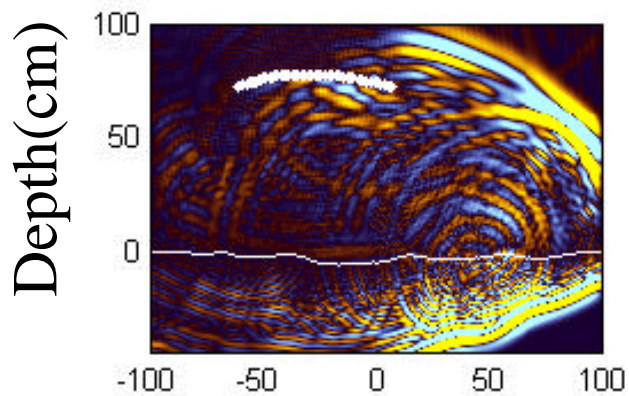


Scattered field from mine

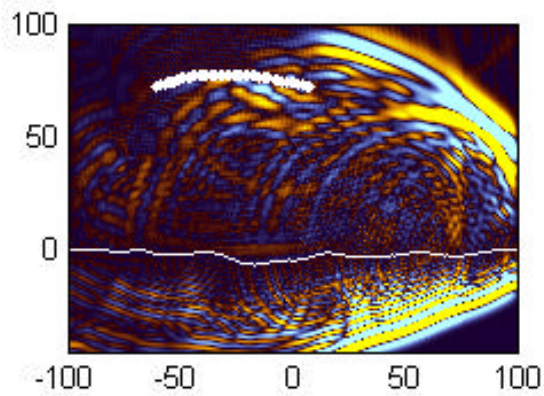


2-D FDTD Simulation of Parabolic Reflector Generated H-Pol. Planar Incident Wave: Non-Metallic Mine 10 cm Below **Rough** Surface

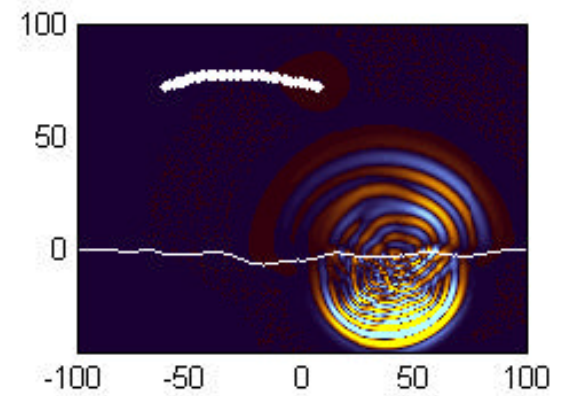
Total field with mine



Total field without mine



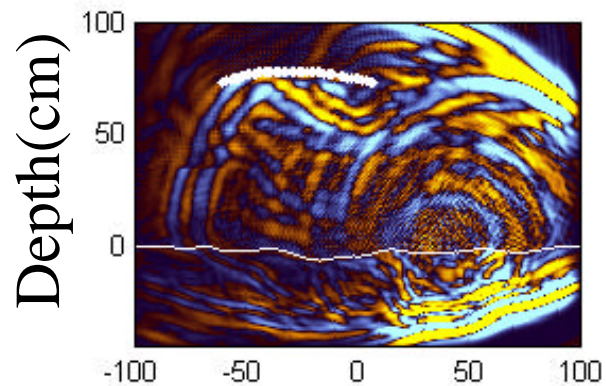
Scattered field from mine



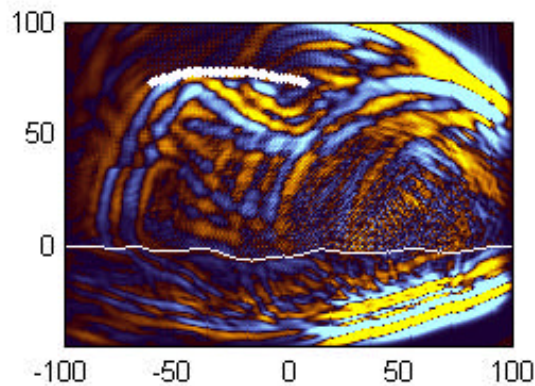
Transverse distance (cm)

2-D FDTD Simulation of Parabolic Reflector Generated V-Pol. Planar Incident Wave: Non-Metallic Mine 10 cm Below **Rough** Surface

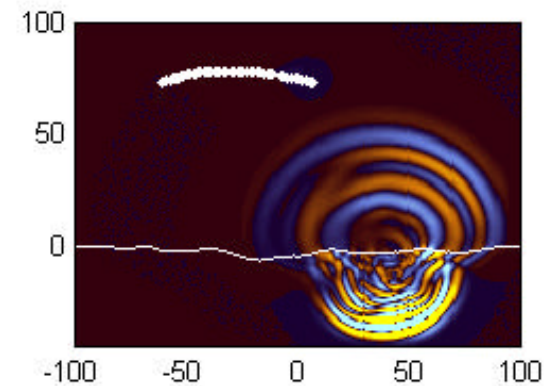
Total field with mine



Total field without mine



Scattered field from mine

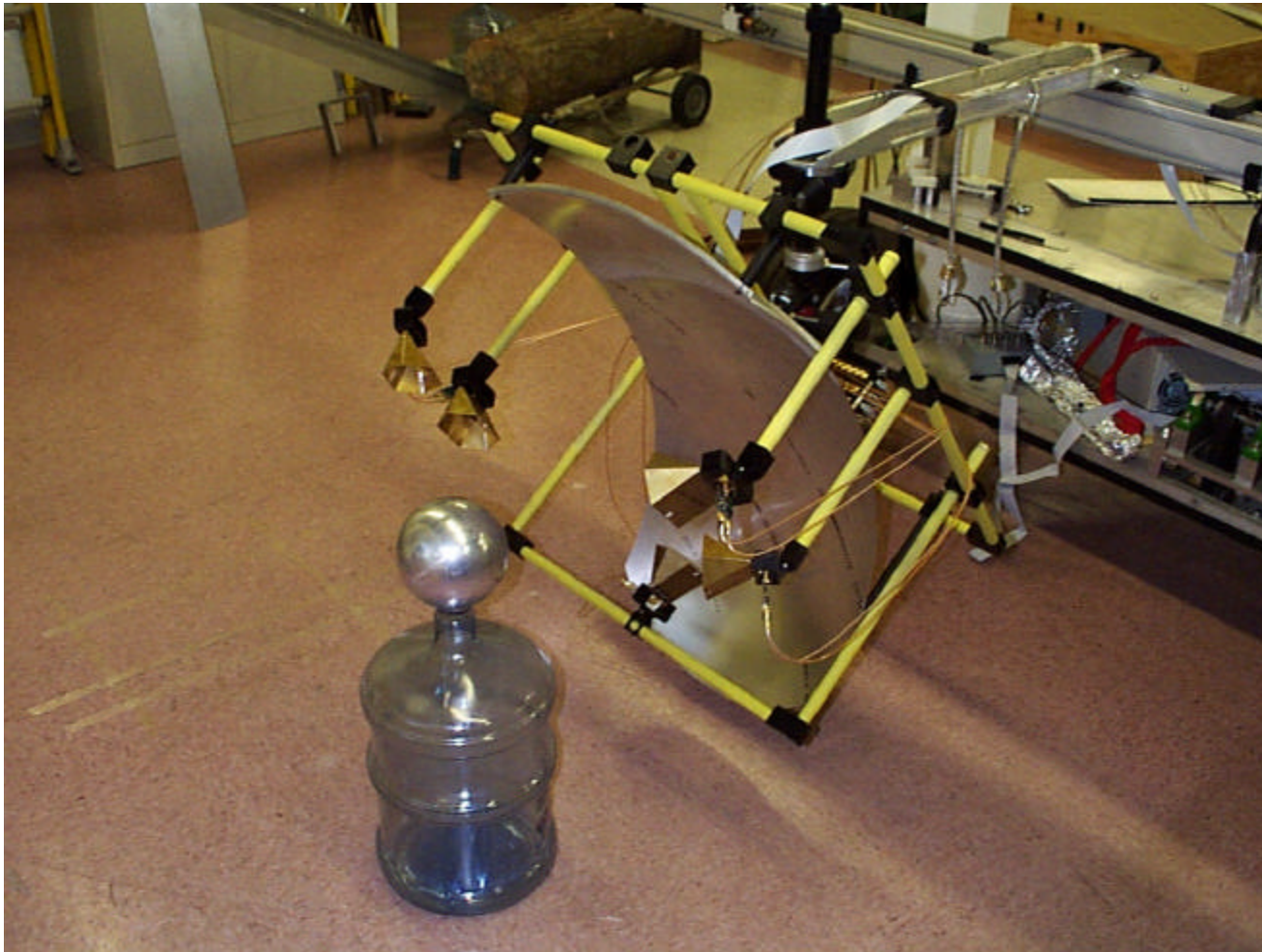


Transverse distance (cm)

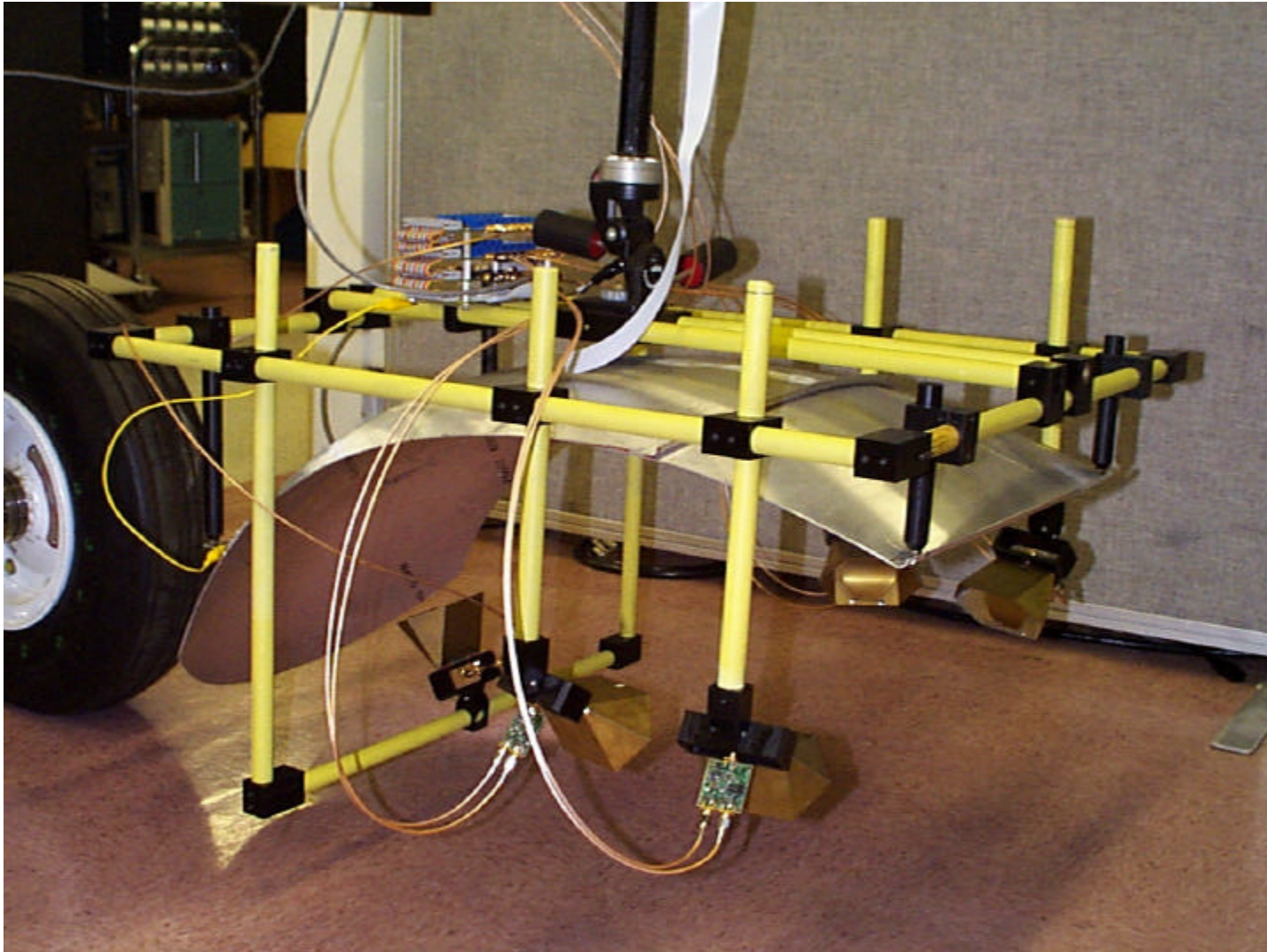
Multistatic Receiving Array

- Collimating transmitter is coupled with a four element receiving array.
- 2 X 2 array gives four times the received power of a monostatic radar, as well as temporal focusing left to right and forward to back.
- By comparing time differences of received scattered signals, approach and localization information can be extracted.
- By measuring overall time delay, depth of burial information is available.

Offset Parabolic Reflector Transmitter with Multistatic Array Receiver Using LLNL MIR Elements



Forward-Looking AP Array GPR: Operational Configuration



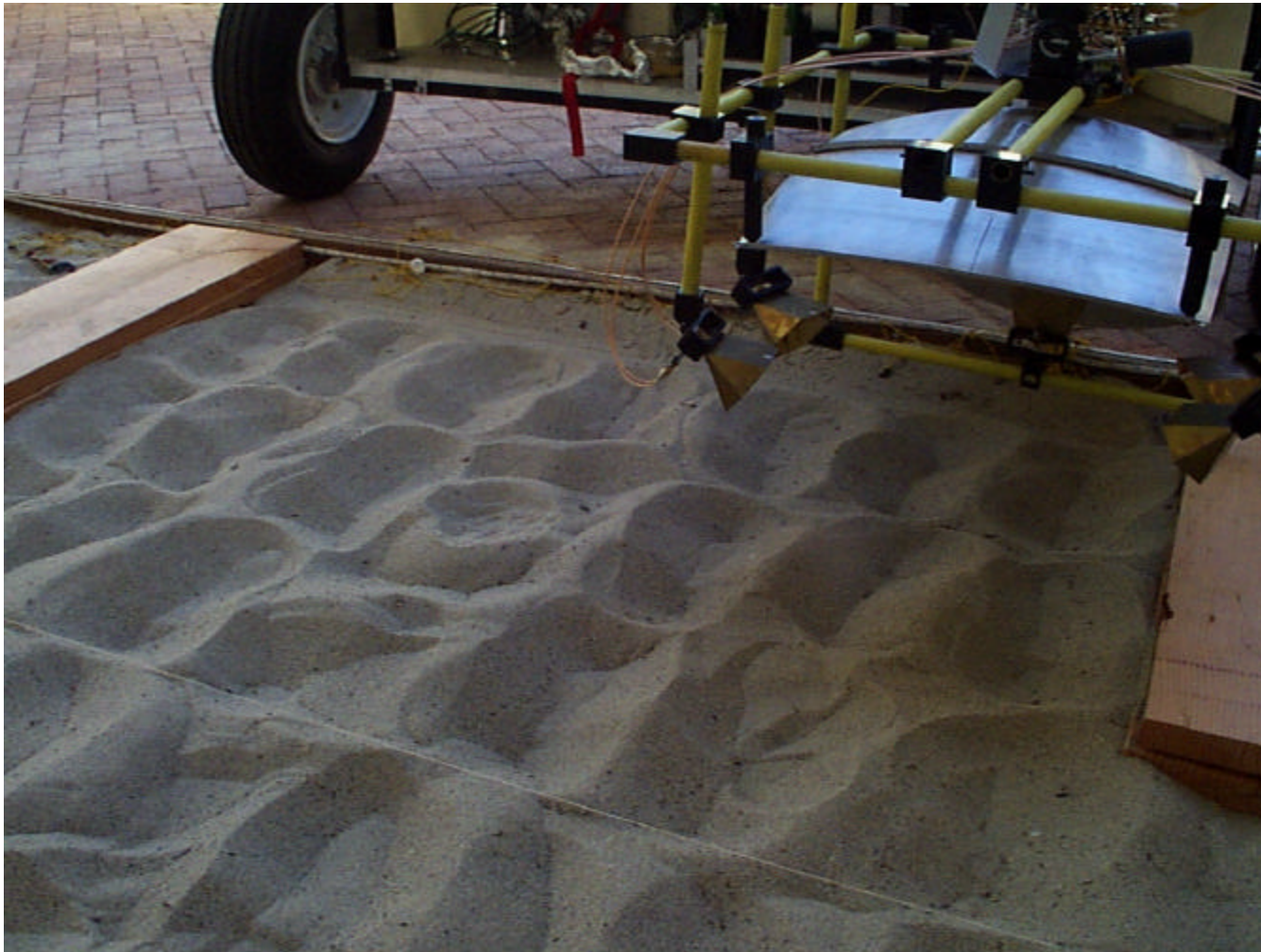
Forward-Looking AP Array GPR: Mounted for Outdoor Testing at LLNL Mine Test Facility



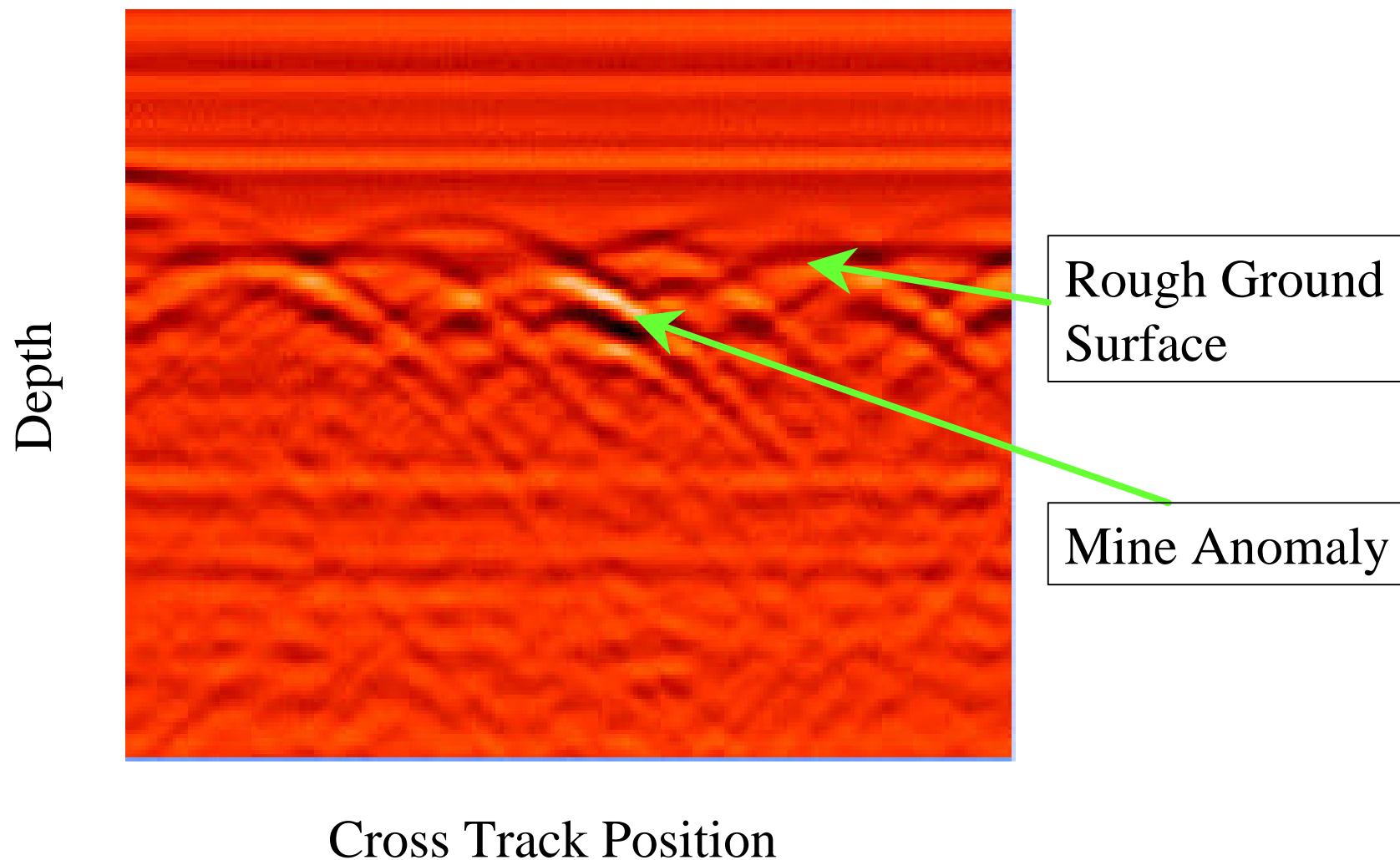
Anti-Personnel Nonmetallic Mine Simulant and Calibration Target



Rough Sand Surface Testing at LLNL



Processed Signal from Non-Metallic Antipersonnel Mine Buried 1 inch Deep in Dry Sand with Rough Ground Surface



Conclusions

- Plane wave excitation decreases scattering confusion.
- Forward looking radar takes advantage of specular ground reflection.
- Offset Paraboloid creates unblocked forward propagating plane wave.
- FDTD predicts distinct scattered and surface reflected signals, even for rough surfaces.
- Experimental results indicate that even small, nonmetallic AP mines are detectable in *sand*

An Efficient Mur-Type ABC for Lossy Scattering Media

John Talbot, Carey Rappaport, and Panos Kosmas
Center for Subsurface Sensing and Imaging Systems
Northeastern University
Boston MA

Progress in Electromagnetic Research Symposium,
July 5-14, 2000
Cambridge, MA USA

Outline:

- Motivation
- Contrasts of Mur and Mur-Type formulations
- Numerical Code Implementation
- Case Models
- Conclusions

Motivations:

- **Mur-Type ABC vs. PML:**
 - **PML:**
 - Multilayered PML loss coefficients must be recomputed at each sublayer for lossy media.
 - Code becomes increasingly complex at multiple boundary limits for 2-D or 3-D geometry.
 - **Mur-Type ABC:**
 - ABC computations are single-celled at all bounding grid points.
 - Only minor adjustments to the code are required as the dimensions of the computational space are selected.

Mur-Type ABC Contrast #1

- Ampere's Law has added conductivity term for lossy soil:

$$\nabla \times \vec{H} = \epsilon_o \frac{\partial}{\partial t} \vec{E} + \sigma \vec{E}$$

Propagation in Soil is Frequency Dependent

$$k(f) = b(f) - ja(f) = \frac{2\pi f}{c} \sqrt{\epsilon'(f) - j \frac{\sigma(f)}{2\pi f \epsilon_0}}$$

**Neglecting the frequency
dependence gives inaccurate results.**

Modeling Dispersion for Easy Transformation to Time Domain

Debye (Teixeira)

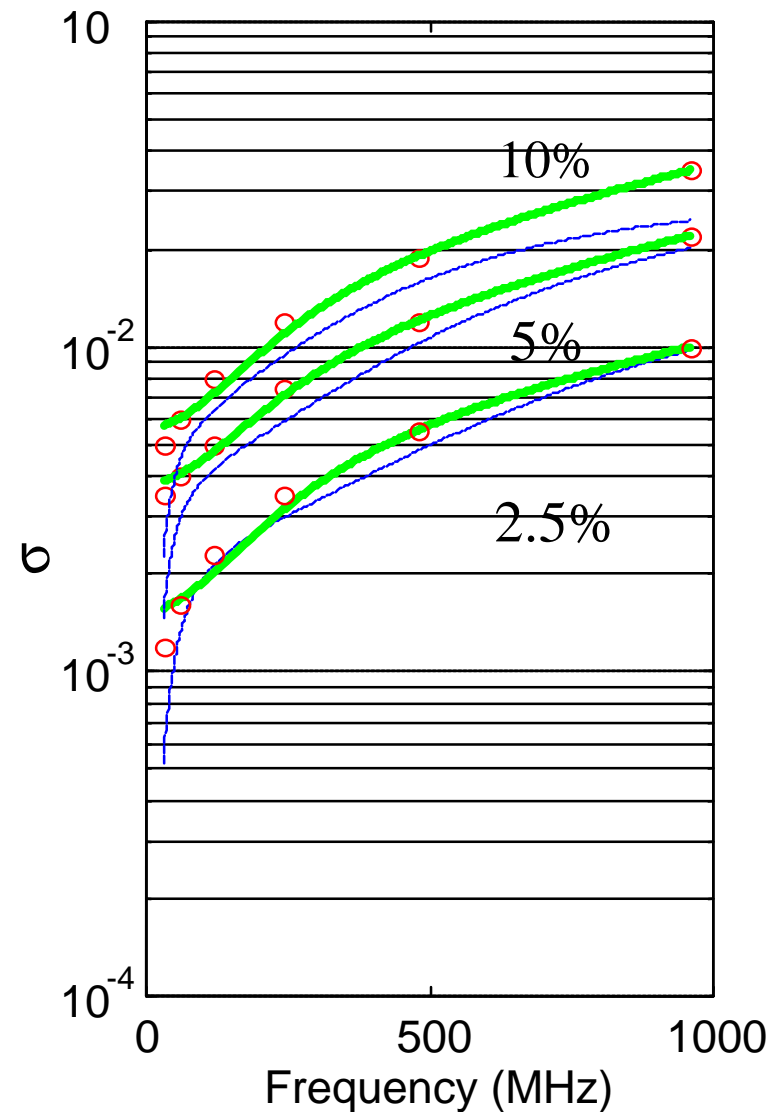
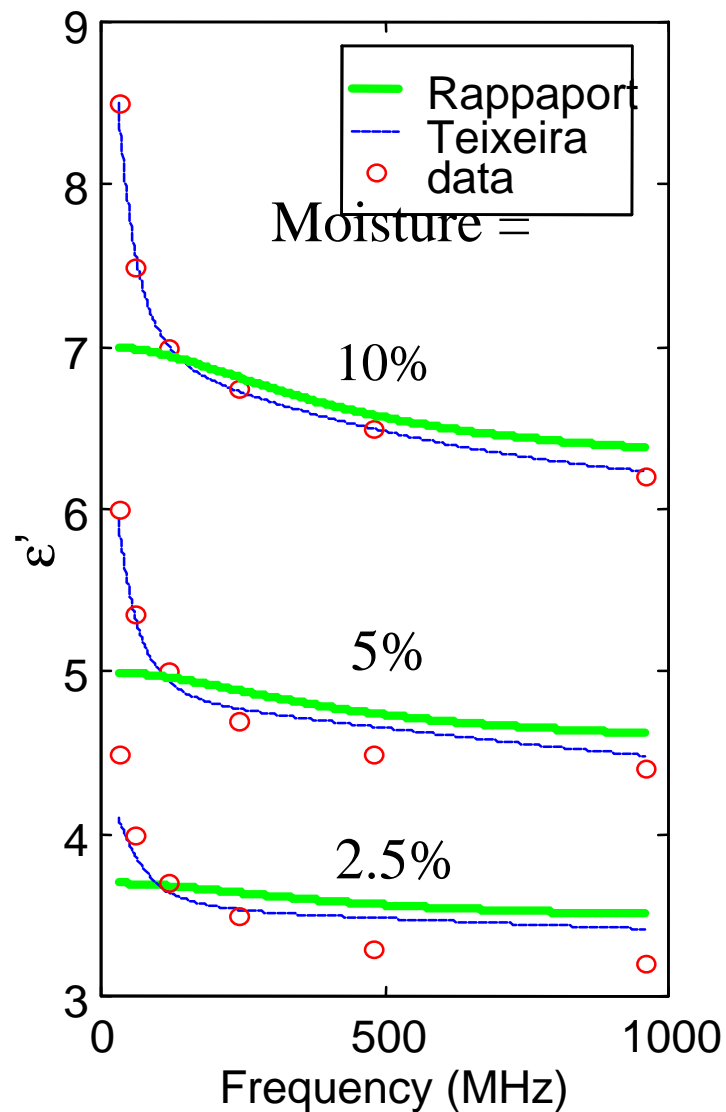
$$\epsilon = \epsilon_0 \epsilon_\infty + \epsilon_0 \sum_{p=1}^N \frac{A_p}{1 - i\omega\tau_p} \quad N=2$$

Pade' (Rappaport)

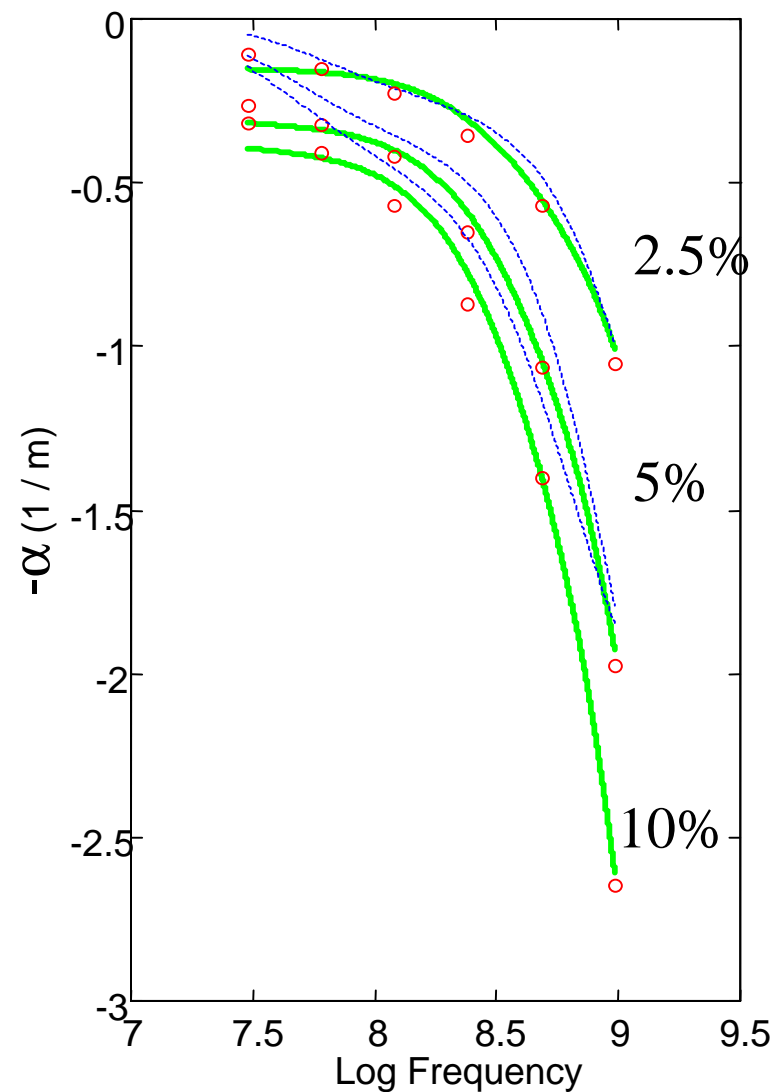
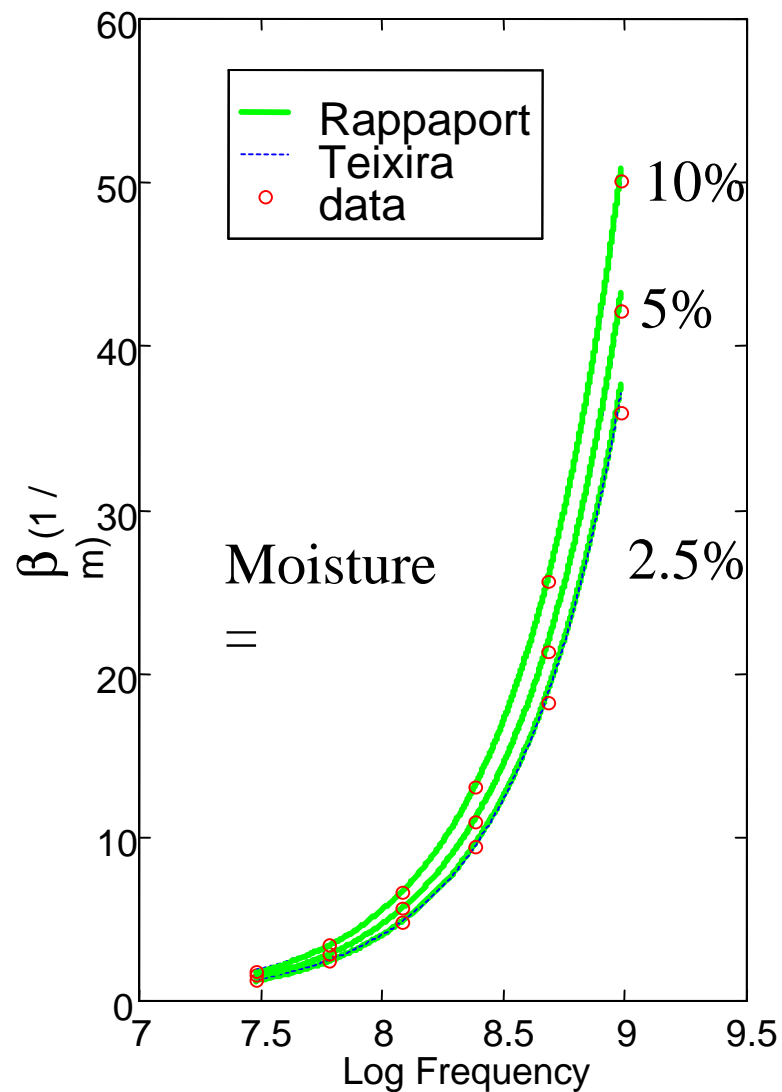
$$\epsilon' = \text{Constant}, \quad s = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2}}{1 + a_1 Z^{-1}}, \quad Z = e^{i\omega D t}$$

M o i s t u r e	2.5 %	5 %	10 %
b 0	. 5 8 1	1 . 3 6	2 . 1 5
b 1	- 1 . 1 2	- 2 . 6 4	- 4 . 1 8
b 2	. 5 4 6	1 . 2 8	2 . 0 3
a 1	- 0 . 9 5	- 0 . 9 5	- 0 . 9 5
ϵ'	2 . 1 8	1 . 5 2	1 . 4 6

Dielectric Constant and Conductivity for Puerto Rican Clay Loam (1.2 g/cc)



Real and Imaginary Wave Number for Puerto Rican Clay Loam (1.2g/cc)



Difficulty with PML ABC for Lossy, Dispersive Media in Time Domain

Must maintain
impedance match

Must gradually increase
conductivity

Free space	ϵ_0 μ_0	ϵ_0, S_1 $\mu_0, \mu_0 S_1$	$\epsilon_0, 4S_1$ $\mu_0, 4\mu_0 S_1$	$\epsilon_0, 9S_1$ $\mu_0, 9\mu_0 S_1$	$\epsilon_0, 16S_1$ $\mu_0, 16\mu_0 S_1$
Lossy soil	ϵ_0, S_s μ_0	$\epsilon_0, S_s + S_1$ $\mu_1, \mu_1 S_2$	$\epsilon_0, S_s + 4S_1$ $\mu_2, \mu_2 S_3$	$\epsilon_0, S_s + 9S_1$ $\mu_3, \mu_3 S_4$	$\epsilon_0, S_s + 16S_1$ $\mu_4, \mu_4 S_5$

μ_n and S_n are functions of
 $\epsilon_0, S_{\text{soil}}, S_1$ and frequency

Mur-Type ABC Contrast #2

- Wave equation for boundary condition has additional term for conductivity:

$$\left[\partial_x^2 + \partial_y^2 + \partial_z^2 + \omega^2 \mu_o \epsilon_o \left(\epsilon' - j \frac{\sigma}{\omega \epsilon_o} \right) \right] E(\omega)_{x=0} = 0$$

Mur-Type ABC Contrast #3

- Pre-approximation equation has conductivity term added to Mur approximation:

$$\text{Using: } \tilde{S}_y = \frac{\partial_y}{j\omega}, \quad \tilde{S}_z = \frac{\partial_z}{j\omega}, \quad \mu_o\epsilon_o = \frac{1}{c^2}$$

$$\partial_x = \left(\frac{j\omega\sqrt{\epsilon'}}{c} \right) \left[1 - j \frac{\sigma}{\omega\epsilon'\epsilon_o} - \frac{(c\tilde{S}_y)^2}{\epsilon'} - \frac{(c\tilde{S}_z)^2}{\epsilon'} \right]^{1/2}$$

Mur-Type ABC Contrast #4

- New approximation retains extra term to account for conductivity:

$$\left[\partial_x - \frac{\sqrt{\epsilon'}}{c} j\omega - \frac{\eta_s}{2} \sigma(\omega) \right] E(\omega)_{x=0} = 0 \quad \text{where : } \eta_s = \sqrt{\frac{\mu_o}{\epsilon' \epsilon_o}}$$

Mur-Type ABC Contrast #5

- Conversion to FDTD is accomplished using a Z-transform:

$$\left[\partial_x - \frac{\sqrt{\epsilon'}}{c} \left(\frac{1 - Z^{-1}}{\Delta t} \right) - \frac{\eta_s}{2} \sigma(Z) \right] E(Z)_{x=0} = 0$$

- Where the conductivity is represented by the Pade' approximant:

$$\sigma(Z) = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2}}{1 + a_1 Z^{-1}}$$

Mur-Type ABC Contrast #6

Second Order Approximation:

$$\frac{\partial_x}{j\omega} = \frac{\sqrt{\epsilon'}}{c} \left[1 - \frac{1}{2} \left(j \frac{\sigma}{\omega \epsilon' \epsilon_o} + \frac{(c\tilde{S}_y)^2}{\epsilon'} + \frac{(c\tilde{S}_z)^2}{\epsilon'} \right) \right]$$

$$\left[\frac{\sqrt{\epsilon'}}{c} j\omega \partial_x - \frac{\epsilon'}{c^2} (j\omega)^2 - \frac{1}{2c^2 \epsilon_o} j\omega \sigma - \frac{1}{2} (\partial_y^2 + \partial_z^2) \right] E = 0$$

For 2-D $E_z(x, y)$:

$$\left[\frac{\sqrt{\epsilon'}}{c} \partial_x - \frac{j\omega \epsilon'}{c} - \frac{\sigma \mu}{2} \right] E_z + \frac{\mu}{2} \partial_y H_x = 0$$

Numerical Code Implementation of Second-Order Approximation

The Z-Transform of the second-order Mur-Type approximation is incorporated into FDTD code with the following guidelines for stability:

- All spatial-derivatives (d/dx) are backward-averaged in time.
- All time-derivatives (d/dt) are forward-averaged spatially.
- All non-derivative terms are forward-averaged spatially.

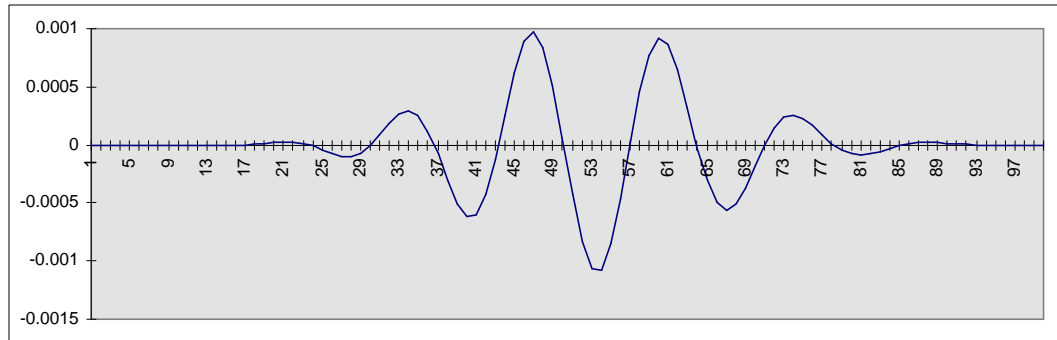
Numerical Implementation of Second-Order Approximation

$$E_{x+1}^{n+1} = -1/(1 + T + B_0 S) \left[E_{x+1}^n (-1 + T + b_0 S) \right. \\ + E_{x+1}^n [(1 + a_1) - T(1 - a_1) + S b_1] \\ + E_x^n [-(1 + a_1) - T(1 - a_1) + S b_1] \\ + E_{x+1}^{n-1} [a_1 - T a_1 + S b_2] \\ + E_x^{n-1} [-a_1 - T a_1 + S b_2] \\ \left. - \frac{S}{\Delta} \frac{\partial (H_{x+1}^n + a_1 H_{x+1}^{n-1} + H_x^n + a_1 H_x^{n-1})}{\partial y} \right]$$

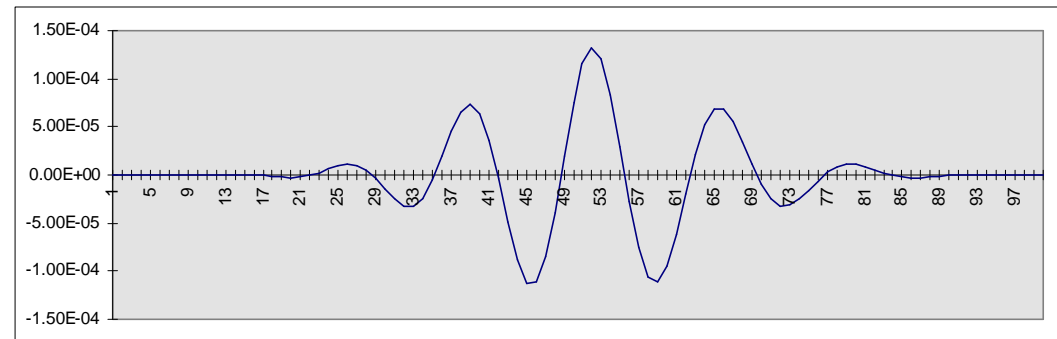
$$T = \frac{\delta \sqrt{\epsilon'}}{c \Delta t} = \frac{1}{R'} \text{ and } S = \frac{\eta_0 \Delta}{2 \sqrt{\epsilon'}}$$

1-D Case Models: Post Wave Reflection at X=0

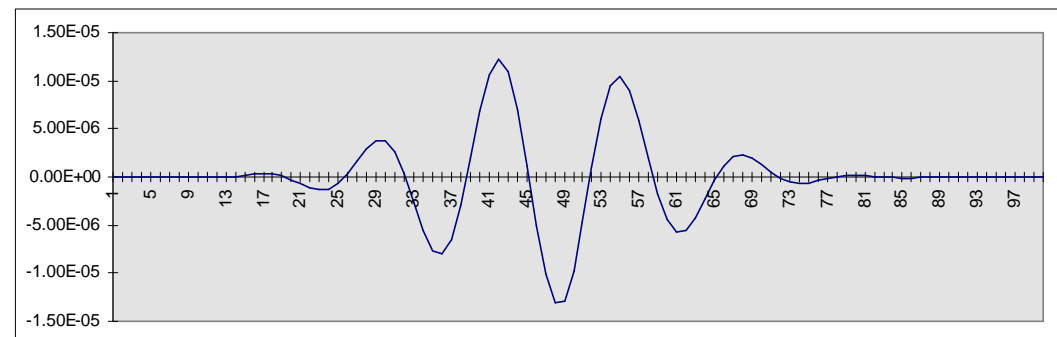
Perfect
Reflective
Boundary
(1.2E-3 Maxima):



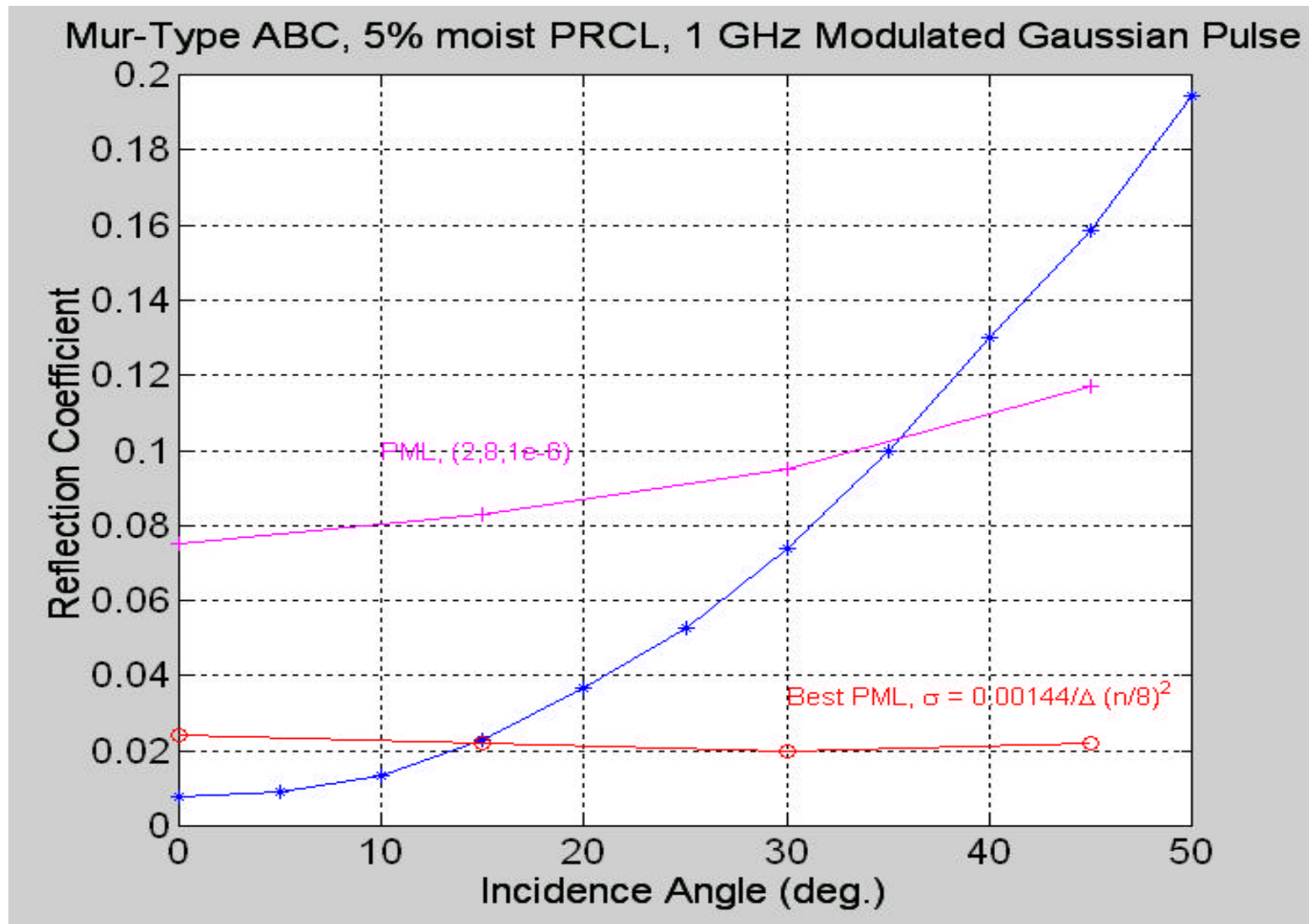
Mur Boundary
(1.3E-4 Maxima):



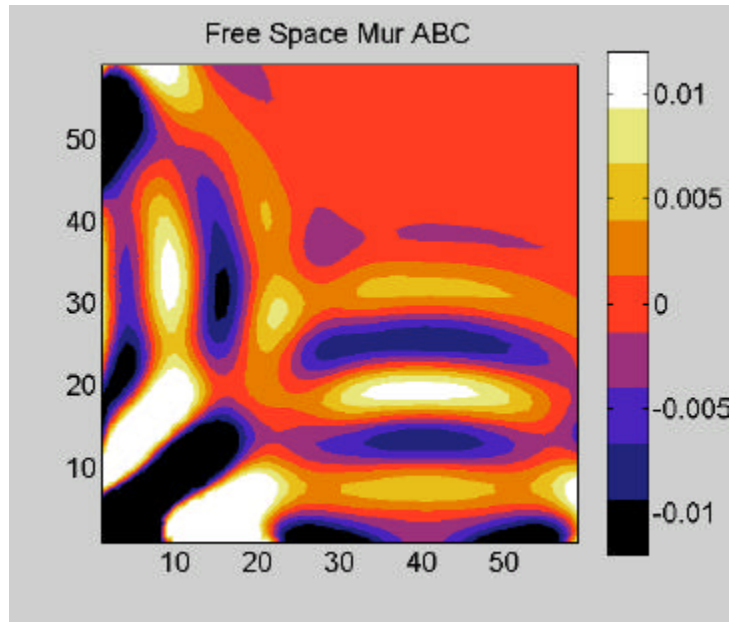
Mur-Type
Boundary
(1.3E-5 Maxima):



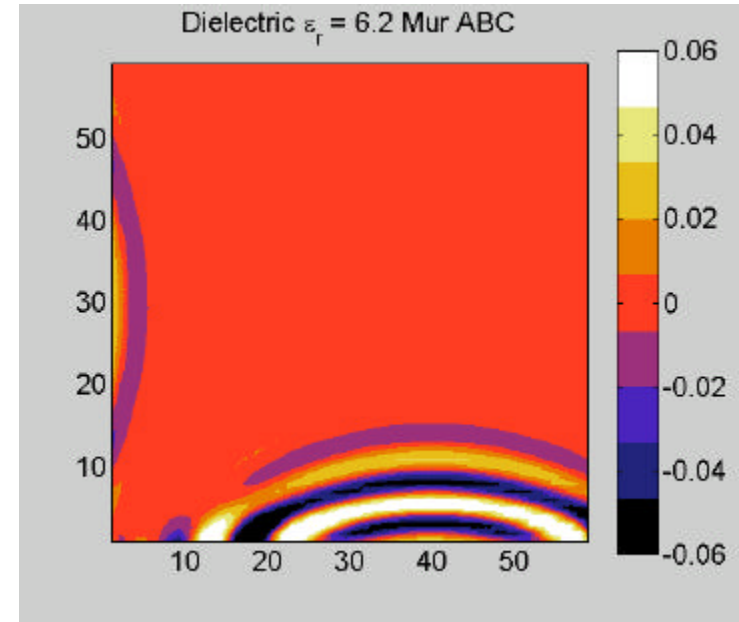
Reflection Coefficient vs. Incidence Angle



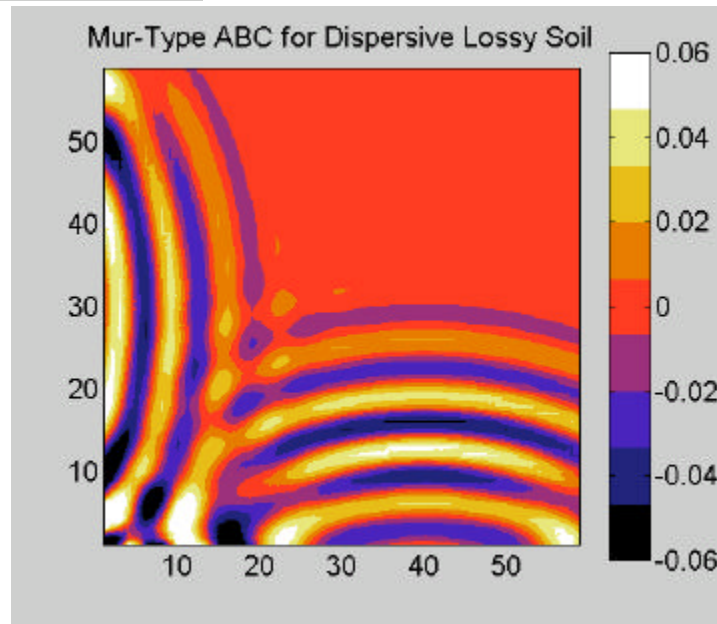
3-D FDTD Boundary Reflection Comparison



Free Space
1.4% max
reflection



Pure Diel.
8% max
reflection



Lossy Soil
6% max
reflection

Conclusions

- Additional conductivity considerations can satisfy the wave equation more appropriately for boundary conditions in lossy media.
- Improved performance by reducing back-scatter from ABC to less than 1% for normal incidence.
- Mur-Type calculations rely on a single-layer boundary thickness: simple, robust, fast.

Quantifying the Effects of Different Rough Surface Statistics for Mine Detection Using the FDTD Technique

Magda El-Shenawee and Carey Rappaport

Center for Electromagnetics Research
235 Forsyth Building
360 Huntington Avenue
Northeastern University
Boston, MA 02115

Email: magda@cer.neu.edu, rappaport@cer.neu.edu
Tel: (617)-373-2064, Fax: (617)-373-8627

ABSTRACT

The finite difference time domain technique, FDTD, is used to calculate the scattered field in the near zone from one-dimensional random rough surfaces. Different statistics for the random surface will be assumed in this work. First, the random rough surface will be characterized by one-scale roughness with Gaussian distribution for the heights and Gaussian auto-correlation function. In the second part, the surface will be assumed to have two-scale roughness (composite) with the same Gaussian statistics as before. The statistics of the scattered fields are calculated in this work using Monte Carlo simulations. Numerical results comparing scattered fields from one-scale roughness and two-scale roughness are shown. The results obtained indicate that the distortion in the scattered signals is primarily due to the small-scale roughness while the two-scale roughness (composite) causes more time delay in the scattered signals. Different rough surface parameters will be used to quantify their effect on the statistics of scattered signals.

Key words: rough surface statistics, composite rough surface, scatter, FDTD, mine detection.

1. INTRODUCTION

A realistic rough ground does not necessarily have a single scale of roughness. In fact, it is more accurate and practical to assume that the rough surface has multi-scale roughness. An investigation about the effect of roughness scale on the statistics of the scattered field is conducted in this work. The probability density function of the surface heights is assumed to be Gaussian for all scale-roughness. The surface profile spectrum (autocorrelation function) is also assumed to be Gaussian. For a composite rough surface (two-scale roughness), the small-scale random rough surface will be imposed on the large-scale rough surface given that each single-scale roughness is generated independently. For the large-scale roughness surface, the correlation length is assumed to be much larger than it in the small-scale roughness surface. This enforces the length of the one-dimensional surface to be sufficiently large, at least ten times the larger correlation length. The statistics of the scattered signals are computed in this work using Monte Carlo simulations. The effect of surface roughness parameters, root mean square

heights and slopes, and surface statistics assumptions on the scattered signals will be investigated as well. The geometry of the problem is shown in Fig. 1a, in which the transmitter and receiver antennas are located above the rough soil ground. Fig. 1b shows the magnitude of the measured incident pulse used in this work.

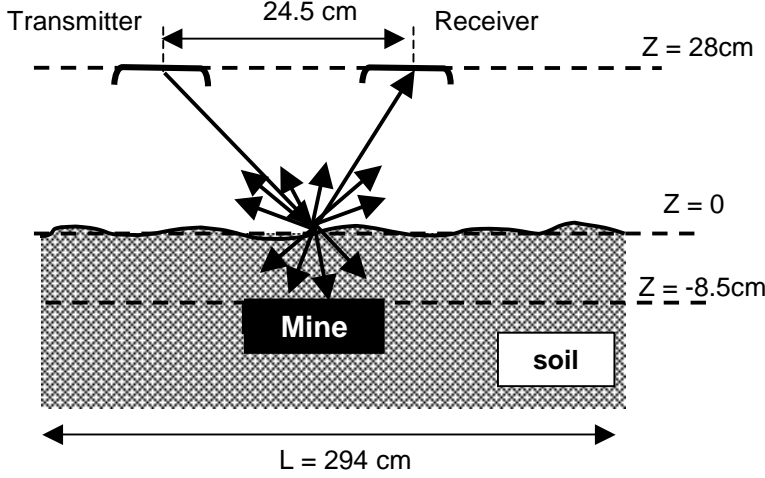


Fig.1a Rough surface sensing geometry.

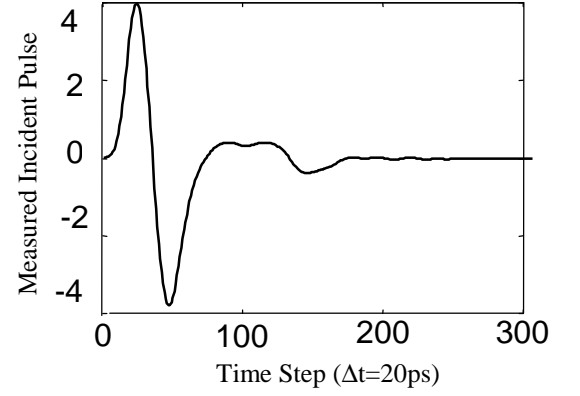


Fig.1b Measured incident pulse.

2. FORMULATIONS

2.1 Metric Measure for Signal Statistics

Three factors will be considered here as metric measure for the signal statistics. They are the average time shift, the average scaling, and the average distortion of the signal. The Monte Carlo simulations and the cross-correlation function will be used to calculate these statistics.

2.1.1 Cross-Correlation Function

The cross-correlation function, by definition, indicates the inter-dependence of the values of two different processes at two different times. The normalized cross-correlation functions between signal ' f ' and signal ' i ' is given by

$$C_{fi}(m) = \sum_{n=1}^{N-|m|} S_f(n) S_i(n+m) / \sqrt{\sum_{n=1}^N |S_f(n)|^2} \sqrt{\sum_{n=1}^N |S_i(n)|^2}, \quad \text{for } m > 0 \quad (1a)$$

and,

$$C_{fi}(m) = C_{if}(-m), \quad \text{for } m < 0 \quad (1b)$$

The relative scaling will be defined as

$$A_i = \sqrt{\sum_{n=1}^N |S_i(n)|^2} / \sqrt{\sum_{n=1}^N |S_f(n)|^2} \quad (1c)$$

where, $i = 1, 2, 3, \dots M$, is the rough surface realization index, M is the size of Monte Carlo Sample (total number of generated random realizations for the rough surface). And N is the total number of time steps in each signal. The above definitions are pictorially described in Fig.2. Numerical illustrative examples for these definitions will be shown in section 3.

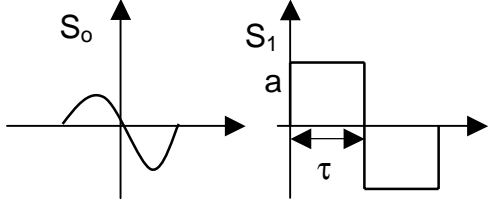


Fig.2a Signals S_0 and S_1 .

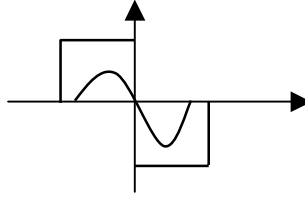


Fig.2b Signal S_1 shifted by τ .

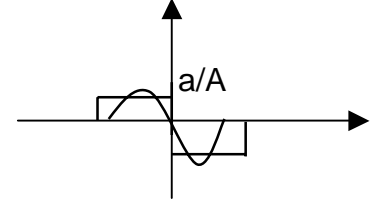


Fig.2c S_1 scaled by A .

2.2 Rough Surface Characterization

The one-dimensional rough surface is described here by $z = f(x)$, where z is the height of any point on the surface. Since the rough surface is assumed to be random, two functions are needed to fully characterize it. They are the probability density function of the surface random heights and the autocorrelation function (surface profile power spectral density function).

2.2.1 Gaussian Random Rough Surfaces

Assuming that the random height of the surface has a Gaussian distribution with zero mean and standard deviation equal to σ_h , the probability density function of the height is [1]

$$p(z) = \frac{1}{\sigma_h \sqrt{2\pi}} \exp\left(-z^2 / 2\sigma_h^2\right) \quad (2)$$

Note that the statistical distribution of the height provides no information about the distances between the hills and the valleys of the surface, i.e., about the density of the surface irregularities [1]. Thus the rough surface can not be described uniquely by the probability density function of the surface height. Another function is needed to complete the description of the rough surface. This function is the autocorrelation function or its Fourier Transform, the surface profile power spectral density function. The autocorrelation function $R(x_d)$ gives the correlation between the random heights at two different points on the surface, x_1 and x_2 . It is defined by [1], [2]

$$R(x_d) = \frac{\langle f(x_1)f(x_1 + x_d) \rangle}{\sigma_h^2} \quad (3a)$$

where,

$$x_d = x_2 - x_1 \quad (3b)$$

For full correlation, $\lim_{x_d \rightarrow 0} R(x_d) = 1$, and for independence, $\lim_{x_d \rightarrow \infty} R(x_d) = 0$. Moreover, if the surface profile spectral density function $W(K_x)$ is given, then the auto-correlation function $R(x_d)$ can be obtained by the inverse Fourier Transform as

$$R(x_d) = \frac{1}{\sigma_h^2} \int_{-\infty}^{\infty} W(K_x) \exp(iK_x x_d) dK_x \quad (4)$$

Thus, the Gaussian joint probability density function for two heights on the surface, z_1 and z_2 , with zero mean and standard deviation σ_h is given by [1],[3]

$$p(z_1, z_2) = \frac{1}{2\pi\sigma_h^2 \sqrt{1 - R^2(x_d)}} \exp\left(-\frac{z_1^2 - 2R(x_d)z_1z_2 + z_2^2}{2\sigma_h^2(1 - R^2(x_d))}\right) \quad (5)$$

If the surface spectral density $W(K_x)$ is assumed to be Gaussian as [2]

$$W(K_x) = \frac{l_c \sigma_h^2}{2\sqrt{\pi}} \exp\left(-\frac{K_x^2 l_c^2}{4}\right) \quad (6a)$$

in which K_x is the surface profile wave number. Thus from (3), the auto-correlation function will also be Gaussian given by

$$R(x_d) = \exp\left(-\frac{x_d^2}{l_c^2}\right) \quad (6b)$$

Where l_c is the correlation distance for which $R(x_d)$ will drop to the value e^{-1} .

2.2.2 Composite Random Rough Surfaces (two-scale roughness)

It is very rare if not impossible to characterize a realistic rough ground soil by only one scale of roughness. In practice, the rough soil is composed of several roughness scales. In this work we will characterize the soil surface by two-scale roughness (composite surface). This two-scale roughness surface is composed of two single-scale independently generated surfaces. The first surface has small roughness parameters, root mean square height and correlation length, and the second surface has large-scale roughness parameters [4]. In other words, the correlation length in the large-scale surface should be at least ten times larger than the correlation length in the small-scale surface [4] and [5]. This requires the length of the 1-D rough surface to be sufficiently large, several correlation lengths. Examples of small-scale, large-scale, and composite rough surfaces are shown in Fig.3. The three random rough surfaces in Fig.3 are assumed to have zero-mean as shown in the small-roughness surface (upper trace). Both the large-roughness (middle trace) and composite surfaces (bottom trace) are shifted in this figure for the purpose of clarification.

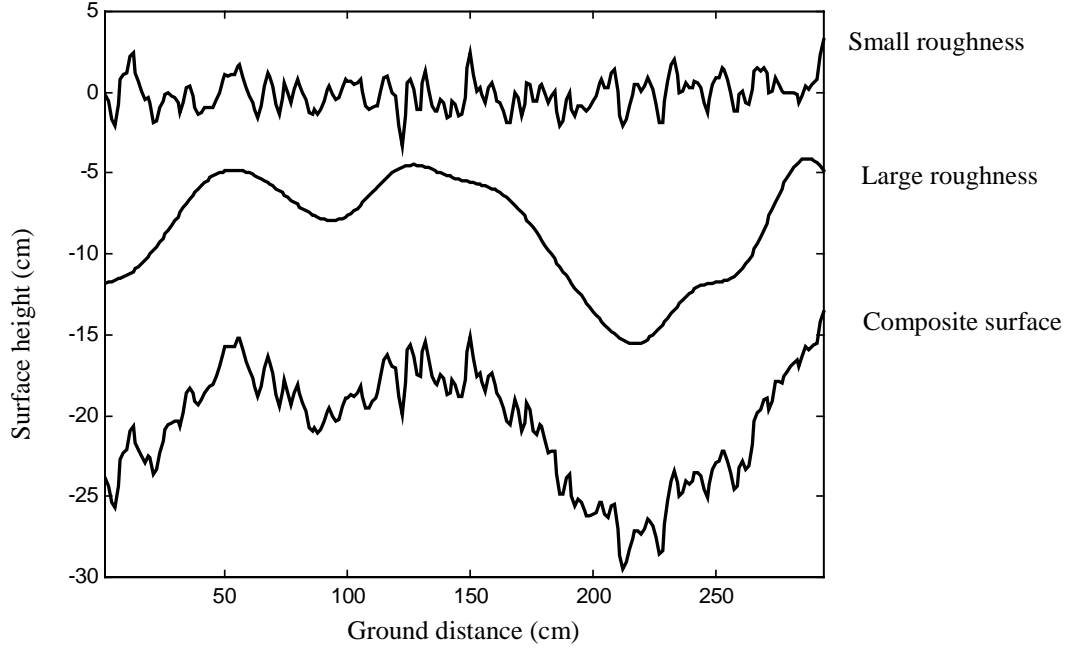


Fig.3 Small ($\sigma_h=1\text{cm}$, $l_c=3\text{cm}$), large ($\sigma_h=3\text{cm}$ and $l_c=30\text{cm}$) and composite roughness surfaces

2.4 Monte Carlo Simulations

The Monte Carlo simulation is the process of taking the ensemble average of several independent random variables. The random variables here are the near fields scattered from several independently generated random rough surfaces. There are several methods to conduct Monte Carlo simulations for the scattered signals, in this work it will be defined as

$$\langle S \rangle = \frac{1}{M} \sum_{i=1}^M S_i \quad (8)$$

where S_i is the amplitude of the scattered near field signal from rough surface realization ' i '. Ideally M should be equal to infinity to assure the application of the central limit theorem. Thus, M will be chosen large enough and each rough surface realization will be generated independently using the computer random number generator.

3. NUMERICAL RESULTS

In all presented results in this work, the ground is assumed to be dispersive soil [6] and the scattered electric field is in the normal direction to the plane of incidence (TM polarization for 2D FDTD calculations). Moreover, the signal that propagates directly from transmitter to receiver has been removed from scattered signals. The nominal frequency of the incident signal is 1GHz (wavelength=30cm). The FDTD time step is $\Delta t=20\text{ps}$ and the space step is $\Delta=1.219\text{cm}$. Numerical results for the definitions of the metric measure for signal statistics, given in section 2.1, are shown in Figs.4-6. The amplitudes of the scattered signal from flat ground and from one rough surface realization

are shown in Figs.4a and 5a, respectively. The roughness parameters in these figures are $\sigma_h=3\text{cm}$ and $l_c=10\text{cm}$. Figs.4b and 5b show the normalized signals of Figs.4a and 5a according to the scaling definition of Eq. 1c. The cross correlation functions, given in Eq.1, for this one rough surface realization with the flat surface signal and for the flat surface with itself are shown in Fig.6a. Notice that the cross correlation function of the flat surface with itself is symmetric around the zero, which is not the case for the rough surface as shown in Fig.6a. Moreover, that there is a time shift τ_i between the flat and rough surface cross correlation functions. These parameters, scaling A_i and time shifting τ_i , will be obtained for each rough surface realization 'i' in the Monte Carlo sample and their average values will be calculated, as will be shown later in this section. Fig.6b shows the scaled and time shifted scattered signals of Figs.4a and 5a.

In Figs.7-10, different statistics for the scattered fields from small-, large-, and composite roughness surfaces, Fig. 3, will be shown. The roughness parameters for the small-scale surface are $\sigma_h=1\text{cm}$ and $l_c=3\text{cm}$, while they are $\sigma_h=3\text{cm}$ and $l_c=30\text{cm}$ for the large-scale surface. The composite surface is obtained by algebraically adding the small- and large- scale surfaces as explained in section 2.2.2. The obtained statistics for the scattered fields are based on the Monte Carlo simulations as given by Eq. 8. In this work, the size of each Monte Carlo sample is chosen to be 500 realizations. In Figs. 7, 8, 9 the histograms for the scaling and time shifting parameters are shown for the small-, large- and composite rough surfaces, respectively. In Fig.10a-c, the scattered signals from the flat surface are compared with the mean value of the scattered signals from these three rough surfaces. The signal scattered from the small-scale roughness surface, Fig10a, is more distorted than the one scattered from the large-scale roughness surface, Fig.10b. Almost no difference is observed in Fig.10b between the two signals. As a result the scattered signal from the composite surface is more distorted than the one scattered from the large-scale roughness surface, Fig.10c. Table 1 shows the comparison between the average values of the time shift and scaling values for these surfaces. As might be expected, large-scale roughness does not appreciably affect the average scaling, where small-scale roughness distorts and spreads the signal in time as well as space. The scaling for the composite surface is dominated by the small-scale roughness value. Interestingly, the similar average time shift for the two single scale roughness add for the composite surface.

The effect of different surface roughness parameters on the statistics of scaling and time shifting values is shown in Table 2. These results are for single-scale roughness surfaces. The first column pair of the table shows the surface roughness parameters, root mean square height and correlation length. The second column pair shows the average scaling and time shifting values of scattered signals. The results show that the time delay and scaling average values increase with the increase of the surface roughness. All these results are obtained using Monte Carlo simulations of 500-sample size each.

4. CONCLUSIONS

Signals scattered from the rough ground are received at the receiver with amount of distortion and time delay compared with those scattered from the flat ground. This amount of the signal distortion and the time delay depend primarily on the surface roughness. Moreover they depend on the roughness scale of the surface as it has been shown here that the surface small-scale roughness affects the amount of distortion while the surface composite surface affects the amount of time delay. The significance of this work is to fully understand the effect of different rough surface statistics on the scattered signals and consequently from the buried mines under these surfaces.

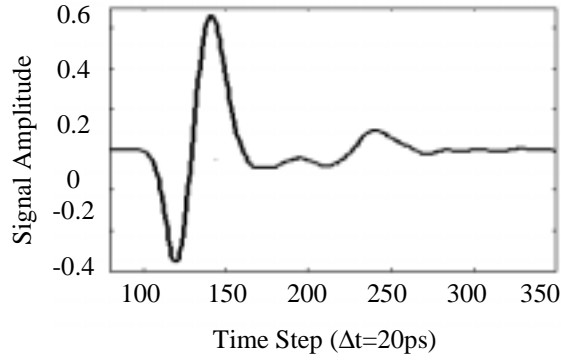


Fig.4a. Amplitude of scattered signal from flat surface.

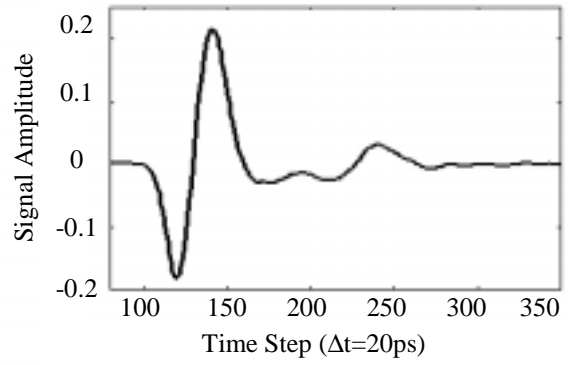


Fig.4b. Normalized Signal for Fig. 4a.

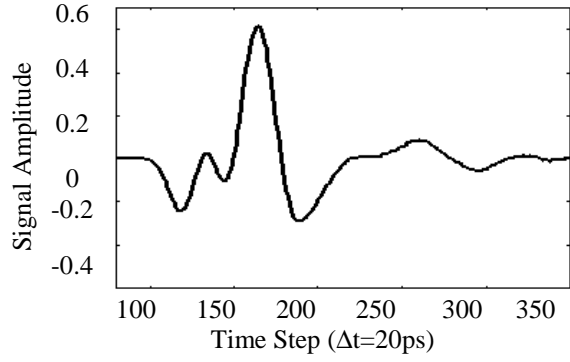


Fig.5a. Amplitude of signal scattered from rough surface.

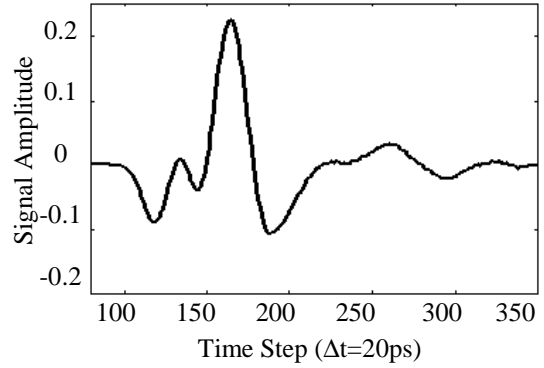


Fig.5b. Normalized Signal for Fig. 5a.

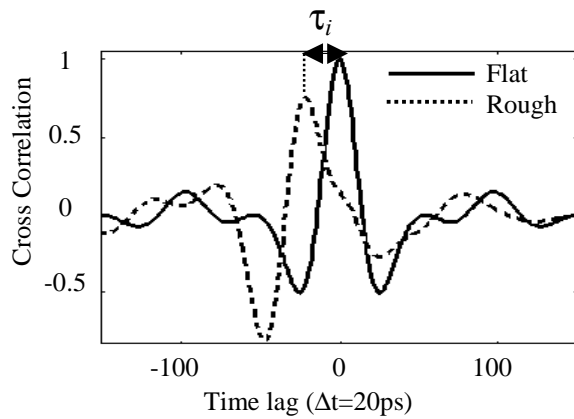


Fig.6a. Cross correlation functions.

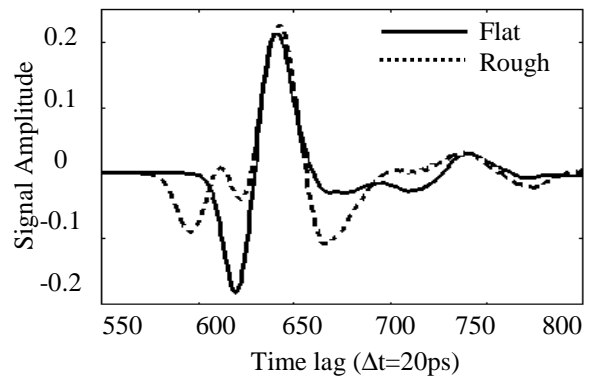


Fig.6b. Signals of Figs.4b and 5b shifted by τ_i .

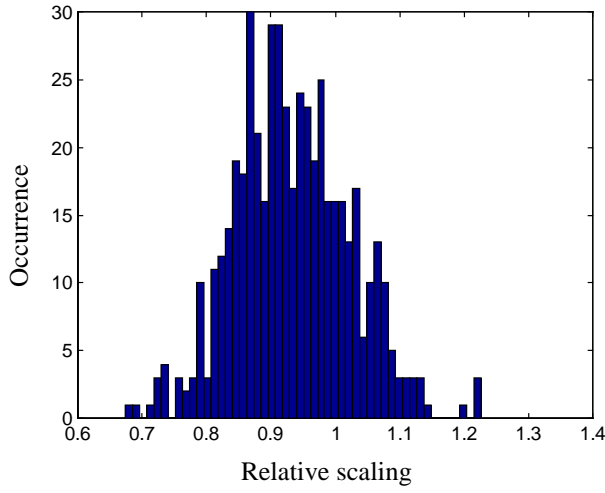


Fig.7a. Scaling for small roughness surface.

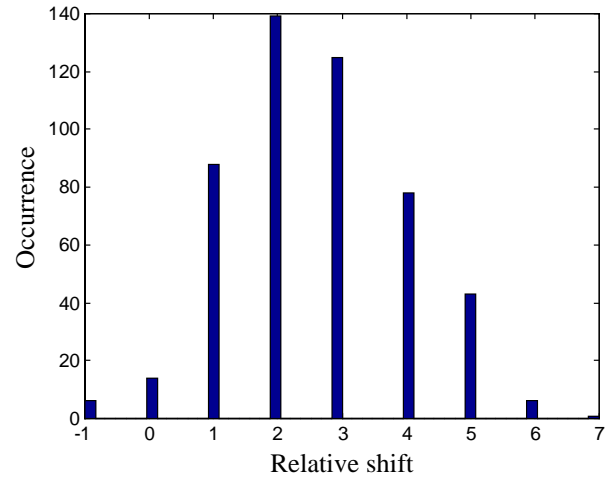


Fig.7b. Time shift for small roughness surface

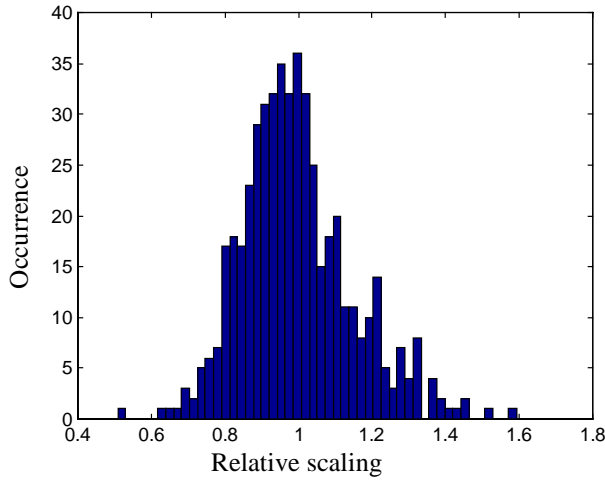


Fig.8a. Scaling for large roughness surface.

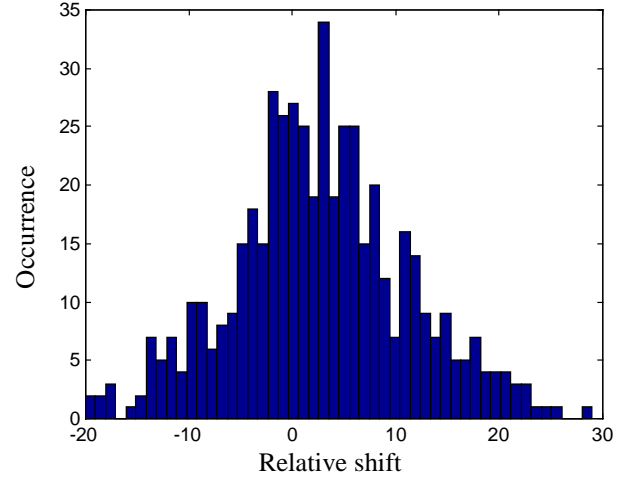


Fig.8b. Time shift for large roughness surface

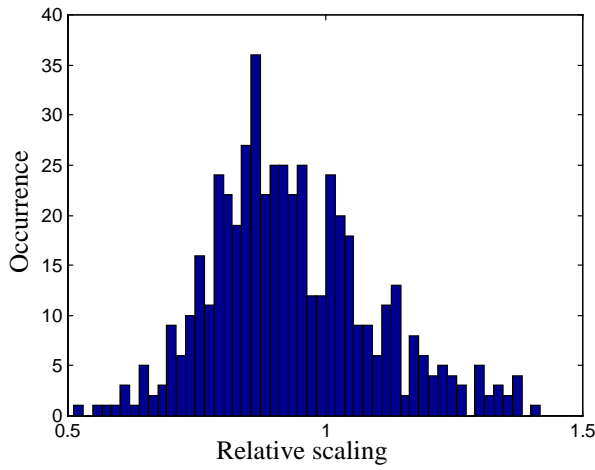


Fig.9a. Scaling for composite surface.

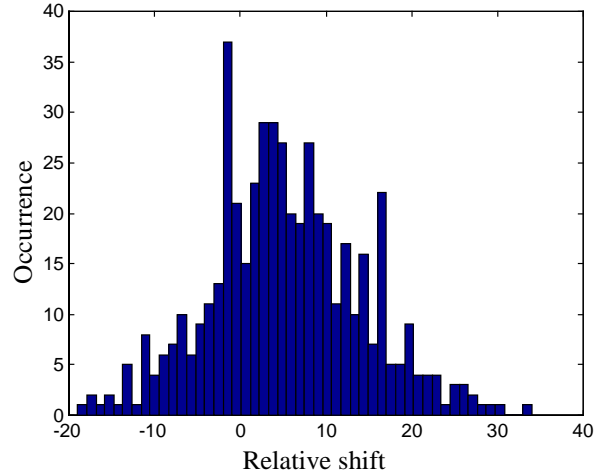


Fig.9b. Time shift for composite surface.

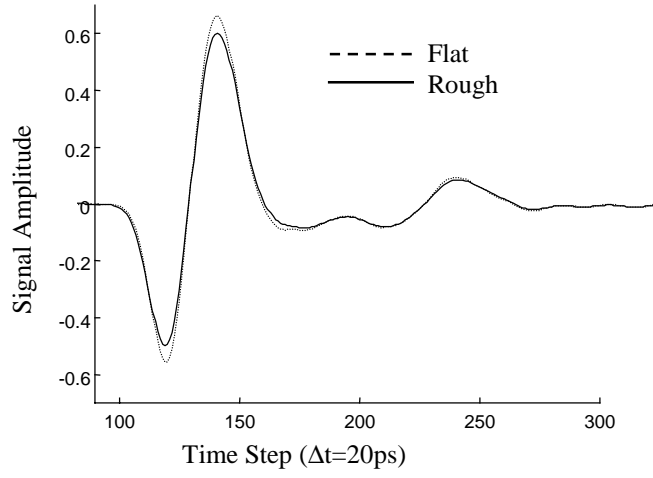


Fig.10a. Average signal for small roughness surface.

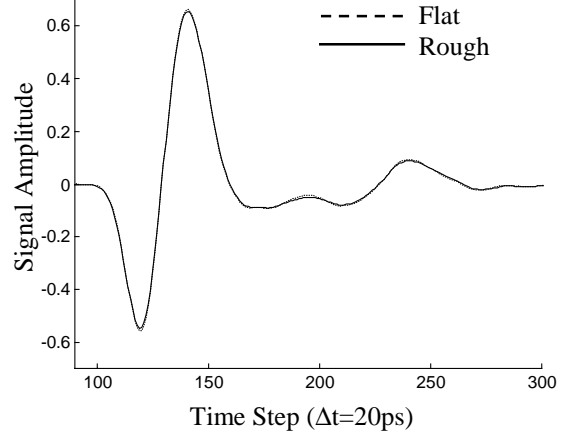


Fig.10b. Average signal for large roughness surface.

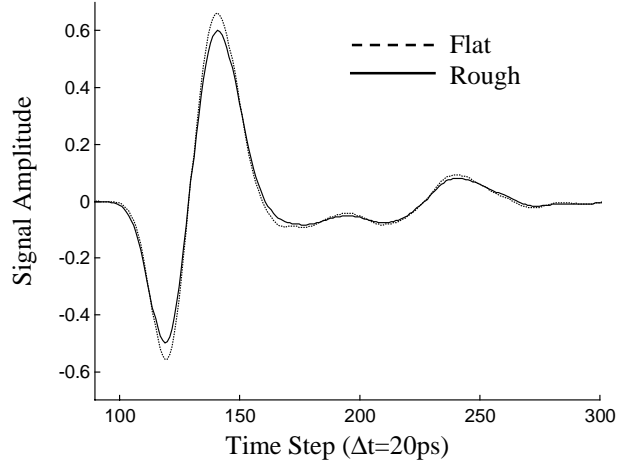


Fig.10c. Mean signal for composite surface.

Table 1 Shifting and scaling statistics for different scale roughness surfaces

Statistics	Small roughness	Large roughness	Composite roughness
Average shifting τ (ps)	52.2	53.28	105.84
Average scaling A	0.9327	0.996	0.9348

Table 2 Effect of surface roughness parameters on signal statistics

Rough surface parameters		Signal statistics	
σ_h (cm)	l_c (cm)	τ (ps)	A
2	10	-67.2	0.944
2	9	-61.4	0.942
2	8	-66.8	0.935
2	7	-81.4	0.933
2	6	-105.8	0.937
2	5	-95.2	0.93
2	4	-109.6	0.94
2	3	-104.6	0.99
3	10	-70.6	0.9
3	9	-105.4	0.901
3	8	-106.6	0.897
3	7	-146.8	0.92
3	6	-139.0	0.89
3	5	-185.6	0.93
3	4	-190.8	0.94
3	3	-198.4	0.99

ACKNOWLEDGEMENTS

This work is supported by the Army Research office grant No. DAAG55-97-0013. All computations were conducted using the High Performance Computer Center at Northeastern University, Boston.

REFERENCES

1. P. Beckmann and A. Spizzichino, The Scattering of Electromagnetic Waves from Rough Surfaces, New York, 1963.
2. E. Thorsos, "The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum," J. Acoust. Soc. Am. 83 (1), 1988.
3. A. M. Mood, F. A. Graybill, and D. C. Boes, Introduction to the Theory of Statistics, McGraw-Hill, 1974.
4. L. Tsang, C. H. Chan, K. Pak, H. Sangani, A. Ishimaru and P. Phu, "Monte Carlo simulations of large-scale composite random rough-surface scattering based on the banded-matrix iterative approach," J. Optical Society of America, Vol. 11, No. 2, pp.691-696, 1994.
5. E. Bahar, "Scattering cross sections for composite rough surfaces: full wave analysis," Radio Science, Vol. 16, No. 6, pp.1327-1335, 1981.
6. W. H. Weedon and C. M. Rappaport, "A general method for FDTD modeling of wave propagation in arbitrary frequency-dispersive media," IEEE Trans. on Antenn. and Prop., Vol. 45, No. 3, pp. 401-410, March 1997.

Analysis of Three Dimensional Scattering from Random Rough Surfaces with Buried Penetrable Objects for Mine Detection Applications

M. El-Shenawee*, E. Miller and C. Rappaport
Center for Electromagnetics Research
Department of Electrical and Computer Engineering
Northeastern University
360 Huntington Ave., Room 235 Forsyth Bld
Boston, MA 02115
magda@cer.neu.edu

The analysis of scattering and transmission of electromagnetic waves in the presence of a random rough dielectric interface and in the nearfield of the sensing systems is a crucial step for subsurface object detection problems in general and landmine remediation applications in particular. Generally, this fully three dimensional problem must be treated numerically, however the calculation of the required fields using conventional techniques (e.g. moment method, finite elements, or finite differences) is a computationally intensive undertaking especially for large dielectric constants. The complexity of the problem dramatically increases upon inserting objects under the rough interface especially when these objects are penetrable. Therefore a fast and accurate computational technique is needed for such applications.

The integral equation-based Fast Multipole Steepest Descent Method (SDFMM), that was originally developed at the University of Illinois (UIUC), will be modified and expanded here to analyze this intensive scattering problem. The rough surface is assumed a random one characterized with Gaussian statistics for the height with zero mean. A single penetrable object is buried at less than one wavelength beneath the mean plane of the rough interface. The incident wave, which is located above the surface, is assumed to be a Gaussian beam that is carefully tapered to minimize surface edge excitations. The PMCHW (Poggio, Miller, Chang, Harrington, and Wu) integral equations are implemented in this work for three regions; air, soil, and buried object. Upon applying the appropriate boundary conditions of the electric and magnetic fields on the air-soil interface, four integral equations are obtained. Both the rough surface and the buried object are discretized into triangular patches. The moment method surface currents are approximated using the RWG (Rao, Wilton, and Glisson) vector basis functions. The interactions between the rough surface and the buried object are fully taken into account here. Thorough investigation will be conducted to test several approximations to ignore some of these interactions aiming to simplify the involved intensive calculations.

Results for the near field complex vectors will be shown. Monte Carlo simulations will be conducted to obtain the statistics of both scattered and transmitted near fields as functions of receiver position, frequency, and incident angle. This work can be extended to include several buried objects of different dielectric constants and/or perfectly conducting ones. The ultimate objective of this research is to analyze, understand and consequently to be able to differentiate between scattering from buried objects and scattering from rough surfaces, clutter in this case.

Analysis of Low Grazing Angle Scattering from Composite Random Rough Surfaces Using the Steepest Descent Fast Multipole Method

Magda El-Shenawee¹, Vikram Jandhyala², Eric Michielssen³ and Weng C. Chew³

¹Northeastern University
Center for Electromagnetics Research
360 Huntington Ave, Boston, MA 02115
Tel: (617)-373-2064, Fax: (617)-373-8627
E.mail: magda@cer.neu.edu

²Ansoft Corporation
Four Station Square
Pittsburgh, PA 15219

³University of Illinois at Urbana-Champaign
Center for Computational Electromagnetics
1406 W. Green St., Urbana IL 61801

The Steepest Descent Fast Multipole Method (SDFMM), a fast multipole inspired integral equation solver for quasi-planar structures, has been used successfully to analyze low grazing angles (LGA) scattering from perfectly conducting rough surfaces. As demonstrated in our previous work, at LGA the length of the finite simulated surface should be at least ten times larger than the width of the surface to adequately project the incident beam onto the surface. As a result, the number of the moment method surface current unknowns becomes excessively large. The SDFMM has been successfully and efficiently used to solve for more than a million surface current unknowns. In our previous work, the random rough surface was assumed to have a single scale roughness with Gaussian statistics. In reality, many rough surfaces exhibit multi-scale roughness, e.g. ocean like surfaces. In this work, we assume that the rough surface is characterized by two-scale roughness, i.e., two root mean square heights and correlation lengths characterize the surface. The rough surface is still assumed to adhere to Gaussian statistics for both the random heights and the auto-correlation functions. The small- and large-scale random surfaces are independently generated using the random Gaussian generator. The composite two-scale surface is obtained by imposing the small-scale surface on the large-scale one. The incident wave is assumed to be a Gaussian beam with incident angle equal to 80 degrees from the normal to the surface and the rough surface is assumed to be perfectly conducting. The goal here is to investigate the effect of the small-scale roughness versus the large-scale roughness on the RCS at low grazing angles. Monte Carlo simulations are used to calculate both the coherent as well as the incoherent RCS as function of the scatter angle. As shown before, for single-scale rough surfaces, the RCS has a peak in the specular direction at LGA. There are still several studies related to the LGA scattering phenomena that need an efficient, fast and accurate algorithm such as the SDFMM to be solved.

Modeling Clutter from Random Rough Ground for GPR Subsurface Sensing Applications

Magda El-Shenawee and Carey Rappaport
Center for Electromagnetics Research
Forsyth BLD, Room # 235
360 Huntington Avenue
Northeastern University
Boston, MA 02115

Email: magda@cer.neu.edu, rappaport@cer.neu.edu
Tel: (617)-373-2064, Fax: (617)-373-8627

Scattering of electromagnetic waves from two-dimensional penetrable rough surfaces (3-D scattering) with different characteristics will be presented here. The integral equation-based fast algorithm, Steepest Descent Fast Multipole Method, will be used to analyze this intensive computational problem. The objective is to investigate the scattering phenomena of electromagnetic waves from typical Bosnian, Puerto Rican, and A.P. Hill soils. Soil moisture level, frequency band and surface roughness parameters are considered major factors that affect the scattered fields. For example, the range of volume moisture level in a typical Bosnian soil will be chosen from 3.8 to 25.3% for a frequency range 600Mhz-2Ghz. Thus the relative dielectric constant will range approximately from 3 to 10 and from 0.04 to 1.3 for the real and imaginary parts, respectively. As well known, for larger soil water content, the relative dielectric constant becomes larger, and as a result the problem becomes more computationally expensive. Different soil samples from Puerto Rican clay loam and A.P. Hill, Firing Point will be used in this work as well. The rough surface is characterized with Gaussian statistics for the random heights and also for the auto-correlation function. The roughness parameters (root mean square height and correlation length) are chosen to be in the moderate roughness range for the current application. For example, the root mean square height will range from 0.1 to 0.2wavelength and the correlation length will range from 0.5 to 1.5wavelength. The scattered near field E-patterns of an incident Gaussain beam will be calculated at different locations above the dielectric interface. The receiver locations are chosen to simulate GPR measurement protocols. Moreover, the transmitted E-patterns will be computed at locations below the rough dielectric interface.

The goal of this paper is to fully understand the effect of different soil properties on the scattering phenomena as an *a priori* phase of investigating scattering from buried mines under the same types of soil.

Statistics of Electromagnetic Near Fields Scattered from 3-D AP Mines Buried under Random Rough Surface Calculated Using the Steepest Descent Fast Multipole Method (SDFMM)

Magda El-Shenawee¹ and Carey Rappaport²

¹Department of Electrical Engineering

University of Arkansas

Fayetteville, Arkansas 72701

Tel: (501)-575-6582, Fax: (501)-575-7967

magda@uark.edu

²Center for Subsurface Sensing and Imaging Systems (CenSSIS)

Northeastern University, 235 Forsyth Bld

Boston, MA 02115

Tel: (617)-373-2064, Fax: (617)-373-8627

rappaport@neu.edu

Abstract

The Steepest Descent Fast Multilevel Multipole Method (SDFMM) is used to analyze the distorting effects of random rough ground surfaces on scattered electromagnetic waves from buried *TNT* mines. The SDFMM method is an integral equation-based fast algorithm that is well suited for two-dimensional penetrable rough surfaces (3-D scattering) in the frequency domain, and it is used to calculate the unknown surface currents on both the rough ground and the buried target as well. In this study all interactions between the rough interface and the buried target are taken into account. The scattered near field E-patterns of an incident Gaussian beam are calculated at different locations above the mean plane of the dielectric rough interface. The receiver locations are chosen to simulate GPR measurement protocols. The dimensions and burial depth of the *TNT* mine are smaller than the free space wavelength with material slightly different from the surrounding soil. The average and the standard deviation of the scattered fields for just the target are calculated and results showed that the presence of the rough interface tremendously distorts the target signal even for the small roughness parameters. Moreover, results showed the degradation of signal as the *TNT* mine is located away from incident beam. This knowledge can significantly contribute to inventing better sensing systems for less false alarm detection strategies.

1. Introduction

The integral equation-based Steepest Descent Fast Multipole Method (SDFMM), that was originally developed at UIUC [1], [2], has been modified to handle the AP-mine detection application as shown in Fig. 1. The significant potential of the SDFMM code is that it calculates the unknown moment method surface electric and magnetic currents on the scatterer in a dramatically fast, efficient and accurate manner. The random rough ground surface was characterized with Gaussian statistics for surface height and for surface autocorrelation function [3]. New formulations of integral equations were obtained and presented in [4]. These new formulations had taken into account all interactions between the rough surface interface and the buried *TNT* mine.

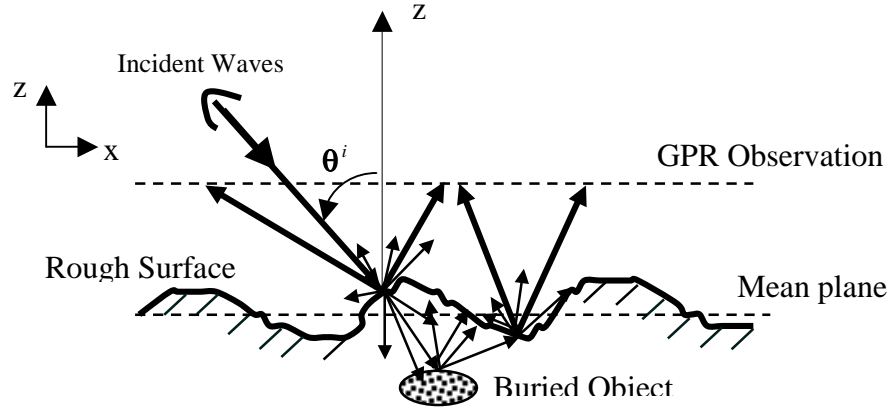


Figure1. Cross section of 2-D rough surface ground with 3-D object buried under the interface.

As an *a priori* phase, the SDFMM was used to analyze the distorting effects of random rough ground surface alone (no buried targets) on the scattered and transmitted electromagnetic waves as presented in [5]. Two well-measured loamy soils [6]: Bosnian and Puerto Rican clay loam were investigated, each with a variety of surface roughness. This study is important in understanding the effects of different soil properties before investigating scattering from buried targets. A specialized fast algorithm was necessary, since multiple calculations of this realistic, but computationally intensive, model were needed for many different realizations of surface roughness. The scattered and transmitted nearfield E-patterns of an incident Gaussian beam [7] were calculated at different locations above and below the mean plane of the dielectric rough interface. The receiver locations above were chosen to simulate GPR measurement protocols. The obtained numerical results showed that the scattered field undergoes more distortion than the transmitted field from both soil types. Moreover, the transmitted fields into the higher dielectric constant Puerto Rican soil experienced more distortion than those transmitted into Bosnian soil [5].

Due to the observed distorting effects of the random rough surface in [5], Monte Carlo simulations are needed to estimate the statistics, average and standard deviation, of the AP-mine signature. The SDFMM has been modified to calculate the fields scattered from a shallow *TNT* mine, buried under the two-dimensional random rough ground. The

buried target is assumed to have size and burial depth comparable to the free space wavelength λ_0 . The scattered electric fields from the rough ground with the buried target are calculated in the near zone and their complex vector average over many rough surface realizations are computed. The target signature was obtained by subtracting the electric fields scattered from the rough ground only from those scattered from the ground with the buried *TNT* mine. The average and standard deviation of the electric fields are calculated as

$$Average = \left| \frac{1}{M} \sum_{i=1}^M \bar{E}_i \right| \quad (1)$$

$$STD = \sqrt{\frac{1}{M} \sum_{i=1}^M |\bar{E}_i|^2 - \left| \frac{1}{M} \sum_{i=1}^M \bar{E}_i \right|^2} \quad (2)$$

where, \bar{E}_i represents the complex vector electric field scattered from the i^{th} rough surface realization where $i=1, 2, \dots, M$, where the size of Monte Carlo set (M) is assumed to be 65 in all results here.

2. Numerical Results

Excellent agreements are obtained upon comparing the modified SDFMM computer code with other published techniques. In our previous work [4], the SDFMM code has been validated with the SMCG method (sparse canonical conjugate gradient method) where the scatterer is a PEC sphere buried under a random rough surface and very good agreement is presented. More validations of the SDFMM code are presented in this section; first, the SDFMM code is used to compute the bistatic scattering cross section of penetrable sphere with dielectric constant ($\epsilon_r = 1.75 - i0.3$) and radius ($a = 0.2\lambda_0$) due to an incident plane wave on the sphere as shown in Fig. 2. The scatterer is just the penetrable sphere immersed in air (no rough surface) and only the MoM part of the SDFMM code is used. The comparison shows very good agreement between the MoM part in the SDFMM code and both the Mie and MoM solutions published in Medgyesi-Mitschang et al [8]. Moreover, Excellent agreement between the SDFMM and the MoM is presented in Fig. 3, where the bistatic scattering cross section σ of the buried spheroid under an individual rough surface realization is shown. To be able to calculate the surface currents using the MOM, the ground dimensions are assumed (in this example) to be $2.96\lambda_0 \times 2.96\lambda_0$. Results in Fig. 3 are obtained at incident angle $\vartheta^i = 0^\circ$ and for the VV-polarization. For these results, the total CPU time (filling the impedance matrix, convergence f iterative solver, and calculating near fields and radar cross sections) is 59 minutes for the MoM with 511MB required computer memory, while they are only 31 minutes and 140MB, respectively, for the SDFMM.

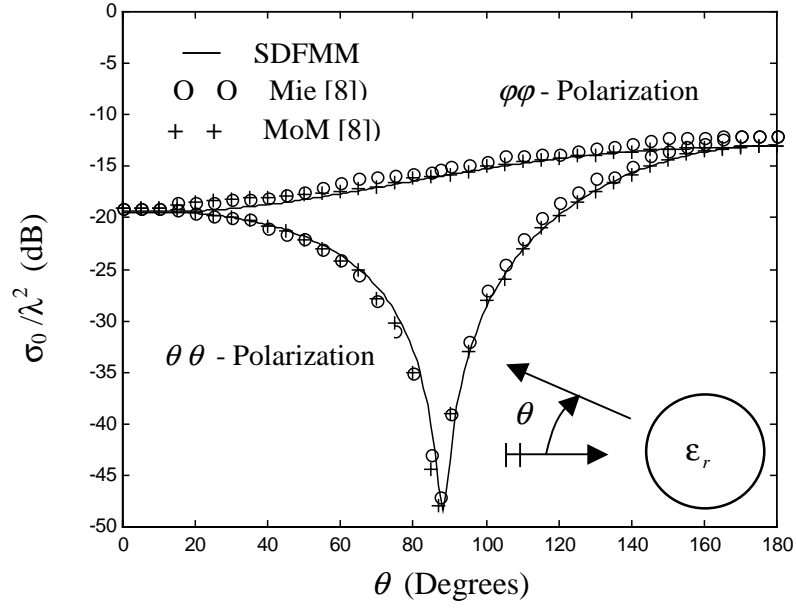


Figure 2. Validation of the SDFMM code (target only) for bistatic scattering cross section σ_0 of penetrable sphere ($\epsilon_r = 1.75 - i0.3$, and $a = 0.2\lambda_0$). Comparison between current results and both Mie and MoM solutions as published in [8].

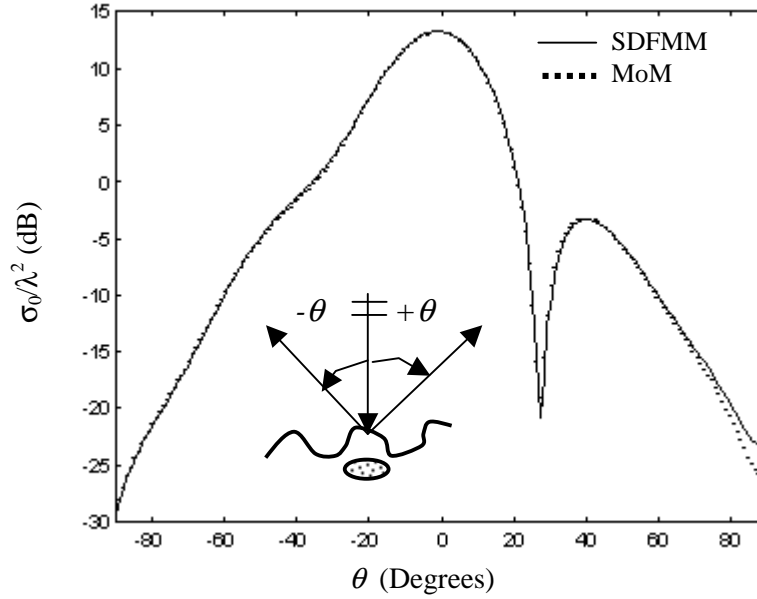


Figure 3. Validation of the SDFMM with the MoM for bistatic scattering cross section σ_0 of penetrable oblate spheroid ($\epsilon_r = 2.9 - i0.072$, $a = 0.3\lambda_0$, $b = 0.15\lambda_0$) buried at depth $d = 0.3\lambda_0$ under rough ground of dielectric constant $\epsilon_r = 2.5 - i0.18$, rms height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$ (one realization) at incident angle $\vartheta^i = 0^\circ$ for VV - polarization.

The dimensions of the ground are assumed $8\lambda_0 \times 8\lambda_0$ with roughness parameters as rms height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$. The ground is a typical Bosnian soil with 3.8% moisture and dielectric constant $\epsilon_r = 2.5 - i0.18$ at 1GHz [6]. The dielectric constant of the *TNT* mine is $\epsilon_r = 2.9 - i0.072$ which is modeled as an oblate spheroid with $a=2b=0.3\lambda_0$ (top view is a circle). The *TNT* mine is buried at $d = 0.3\lambda_0$ from its center to the mean plane of the rough ground. The incident Gaussian beam is pointing normally at the center of the ground. The difference in the near electric fields scattered from the rough ground without buried target and those scattered from the same ground but with buried target, are the scattered fields due just to the buried target. We repeated this subtraction process for each of the 65 rough surface realizations, and took the statistical average of the all scattered fields using Eq. 1. Thus the average signature of the target is obtained and plotted in Fig. 4a. The results show that the *TNT* mine has signature almost equal to 5% of the total scattered fields [5]. Similar results to Fig. 4a are obtained at different locations of the buried target as $x = y = 5.0\lambda_0$ and shown in Fig. 4b and $x = 6.5\lambda_0, y = 4.0\lambda_0$ as shown in Fig. 4c. These results show the degradation of object signature as function of its closeness to the center of the incident Gaussian beam. Using Eq. 2, the standard deviation (*STD*) of the 65 scattered electric fields due just to the target is calculated and plotted in Fig. 5. The results show that the *STD* of the target signal is almost 30% of the target signature shown in Fig. 4a.

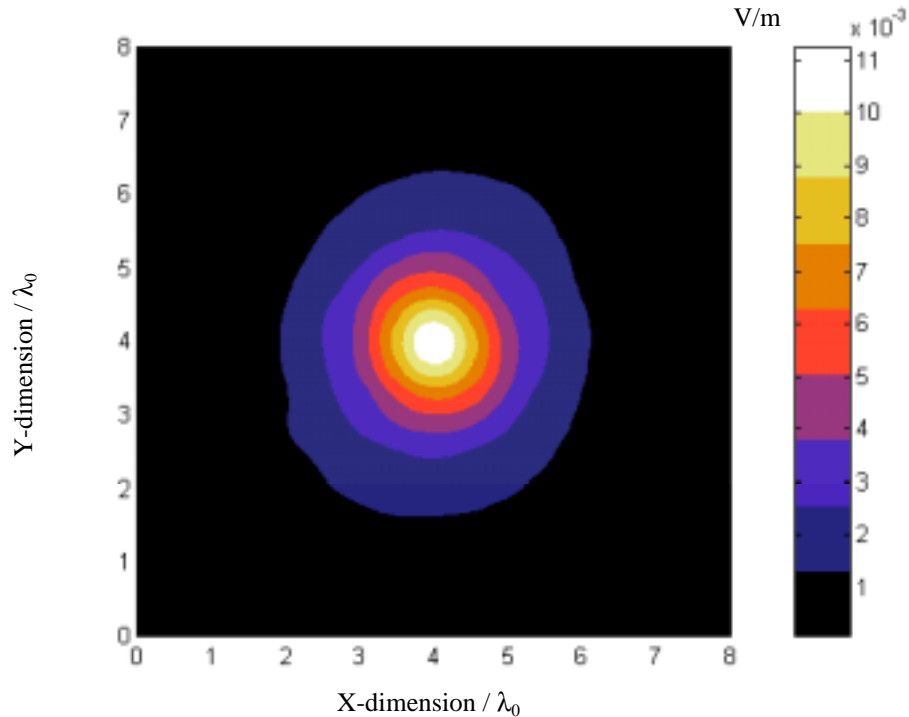


Figure 4a. Average near electric field scattered due to only buried object obtained by subtracting scattered from rough ground without buried target from those scattered from rough ground with buried target, then take the statistical average. Incident angle $\vartheta^i = 0^\circ$ for horizontal polarization. The target is buried at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

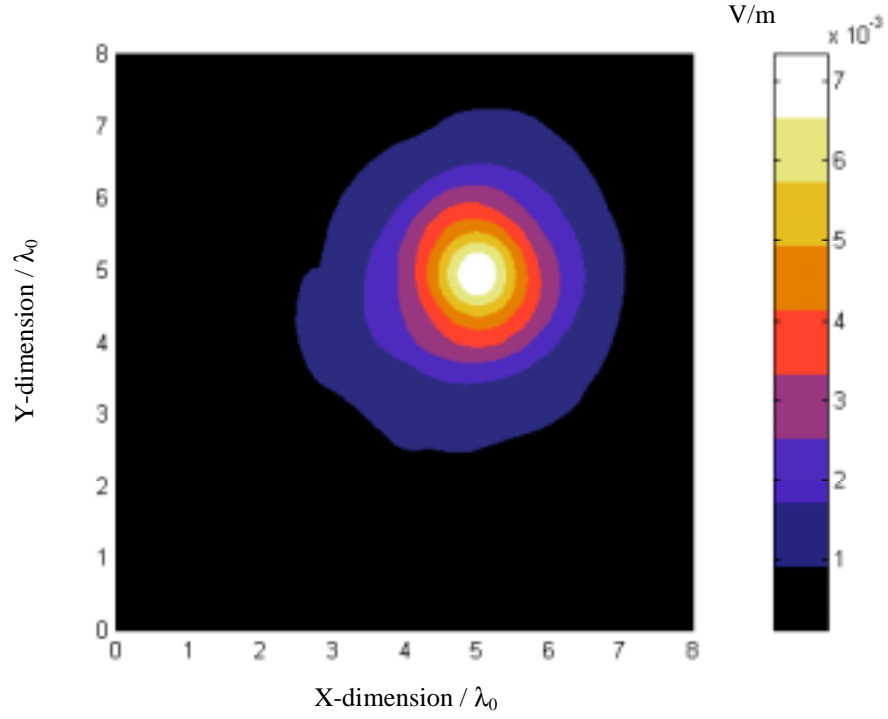


Figure 4b. Average near electric field scattered due just to the object for the same data in Fig. 4a except for the target is buried at $x = y = 5.0\lambda_0, z = -0.3\lambda_0$.

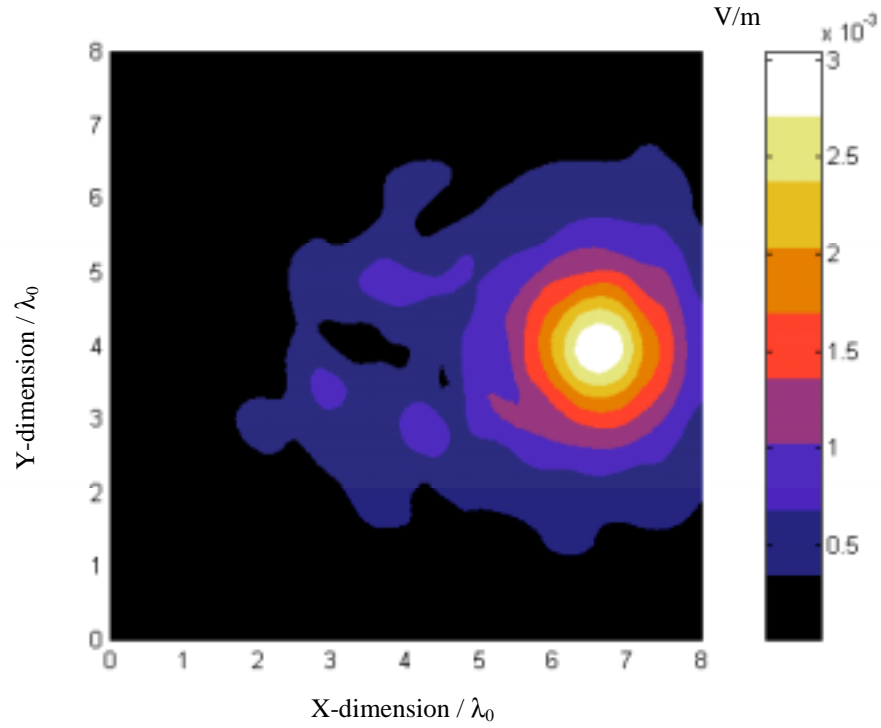


Figure 4c. Average near electric field scattered due just to the buried object for the same data as in Fig. 4a except for the target is buried at $x = 6.5\lambda_0, y = 4.0\lambda_0, z = -0.3\lambda_0$.

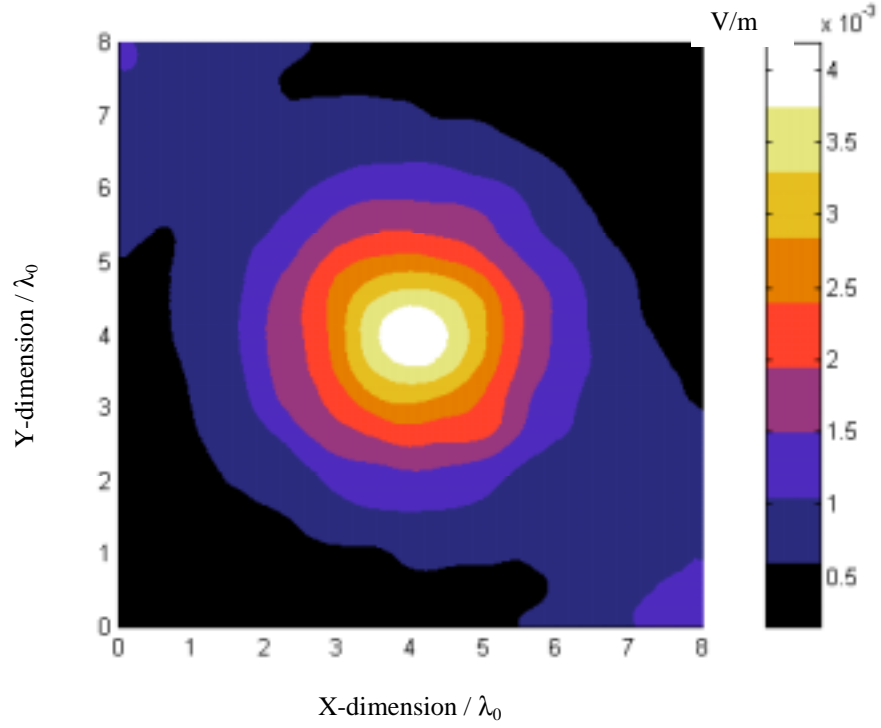


Figure 5. Standard deviation (*STD*) of near electric fields scattered due to just buried spheroid for the same data in Fig. 4a.

3. Conclusions

In this work, the statistics of the scattered fields due just to the buried *TNT* mine are obtained using the modified fast code SDFMM. Monte Carlo simulations are conducted to obtain the scattered fields from the two-dimensional random rough ground once with the target and once without the target. The average and standard deviation of the scattered fields due to just the target are calculated by subtraction. The results showed the significant distorting effect of the rough ground. Moreover, this study showed the degradation of the *TNT* mine signal if it is located away from incident beam.

ACKNOWLEDGEMENTS

This work is supported by the Army Research office grant No. DAAG55-97-0013. All computations were conducted using the High Performance Computer Center at Northeastern University, Boston.

References

1. V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
2. M. El-Shenawee, V. Jandhyala, E. Michielssen and W. C. Chew, "The steepest descent fast multipole method (SDFMM) for solving combined field integral equation pertinent to rough surface scattering," *Proc. of the IEEE APS/URSI '99 conf.*, Orlando, Florida, pp. 534-537, July 1999.
3. N. Garcia and E. Stoll, "Monte Carlo calculation for electromagnetic-wave scattering from random rough surfaces," *Phy. Rev. Let.*, vol. 52, no. 20, pp. 1798-1801, May 1984.
4. M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, "3-D Subsurface Analysis of Electromagnetic Scattering from Penetrable/PEC Objects Buried Under Rough Surfaces: Use of the Steepest Descent Fast Multipole Method (SDFMM)," *IEEE Trans. Geosci. Remote Sensing*, to be published in June 2000.
5. M. El-Shenawee and C. Rappaport, "Modeling Clutter from Bosnian and Puerto Rican Rough Ground Surfaces for GPR Subsurface Sensing Applications Using the SDFMM Technique," submitted to *J. Subsurface Sensing Technologies and Applications*.
6. J. Curtis, "Dielectric properties of soils; various sites in Bosnia," *US Army Corp. of Eng., Waterways Experim. Station Data Rep.*, 1996.
7. P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: Enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.
8. L. Medgyesi-Mitschang, J. Putnam, and M. Gedera, "Generalized method of moments for three-dimensional penetrable scatterers," *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1383-1398, April 1994.

Electromagnetic Wave Scattering from Two Nearby Objects Buried Under Random Rough Surface Using the SDFMM: Subsurface Sensing Applications

*Magda El-Shenawee¹ and Carey Rappaport²

¹Department of Electrical Engineering
3217 Bell Engineering Center
University of Arkansas
Fayetteville, Arkansas 72701
Tel: (501)-575-6582, Fax: (501)-575-7967
magda@uark.edu

²Center for Subsurface Sensing and Imaging Systems
Department of Electrical and Computer Engineering
Northeastern University, 235 Forsyth BLD
Boston, MA 02115
Tel: (617)-373-2043, Fax: (617)-373-8627
rappaport@ece.neu.edu

In anti-personnel mine detection applications, explosive objects buried under ground with rough soil surface could be surrounded by different clutter objects such as rocks, roots, sticks, metallic nails, vegetation, etc. Target discrimination between the AP mine and any sort of object clutter is necessary to minimize false alarms.

In this work, we present the analysis of the electromagnetic wave scattered from two objects buried under a 2-D random rough surface. These two buried objects could both be perfect electric conductors (PEC), one plastic object and one PEC, or two plastic objects. The random height variation of the rough ground is assumed to have Gaussian probability density function and the surface autocorrelation function is assumed Gaussian as well. A carefully tapered Gaussian beam is assumed for incident waves.

The integral equation-based fast algorithm, the Steepest Descent Fast Multipole Method (SDFMM), will be implemented in this work to calculate the electric and magnetic surface currents on the soil/ground interface and the multiple objects. The great advantage of the SDFMM lies in its $O(N)$ computational complexity versus the $O(N^2)$ for the Method of Moment (MOM) to solve N linear system of equations using an iterative solver. The SDFMM was originally developed at UIUC.

In our previous work, we used the SDFMM to analyze the near electric field scattered from single plastic object buried under the rough ground. The rough ground surface was the only source of clutter in the surrounding environment. Here, we are adding another source of clutter that is the proximity of an un-explosive object to the AP mine. The effect of the proximity and orientation of clutter object on target signature will be investigated here. Our objective is to understand the physics behind the mechanism of scattering from these two buried objects and hence to be able to discriminate between the explosive target and the clutter object.

Electromagnetics Computations Using the MPI Parallel Implementation of the Steepest Descent Fast Multipole Method (SDFMM).

M. El-Shenawee¹, C. Rappaport², D. Jiang² and W. Meleis²

¹Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701
magda@uark.edu

²Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
rappaport@neu.edu

Abstract

The computational solution of large-scale linear systems of equations necessitates the use of fast algorithms, but is also greatly enhanced by employing parallelization techniques. The objective of this work is to demonstrate the speedup achieved by the MPI-based parallel implementation of the Steepest Descent Fast Multipole Method (SDFMM). Although this algorithm has already been optimized to take advantage of the structure of the physics of scattering problems, there is still the opportunity to speed up the calculation by dividing component tasks into pieces and using multiple processors to solve them in parallel. The SDFMM has three bottlenecks ordered as (1) filling the sparse impedance matrix associated with the near field moment method interactions, (2) the matrix vector multiplications associated with this sparse matrix (3) the far field interactions associated with the fast multipole method. The parallel implementation task is accomplished basically using a thirty-one node Intel Pentium Beowulf cluster and is also validated on a 4-processor Alpha workstation. The Beowulf cluster consists of thirty-one nodes of 350MHz Intel Pentium IIs with 256 MB of RAM and one node of a 4x450MHz Intel Pentium II Xeon shared memory processor with 2GB of RAM with all nodes

connected to a 100 BaseTX Ethernet network. The Alpha workstation has a maximum of four 667MHz processors. The parallelized computer code is tested for different cases of the anti-personnel landmine detection application. Our numerical results show linear speedup in filling the sparse impedance matrix that tremendously reduced the overall code's runtime. Using the 32-processors on the Beowulf cluster lead to achieve a 7.2 overall speedup and significant reduction in the runtime is gained using the 4-processors on the Alpha workstation.

I. Introduction

The SDFMM is an integral equation-based fast algorithm that is a hybridization of (1) the Method of Moment (MoM), (2) the Fast Multipole Method (FMM), (3) the Steepest Descent Integration rule (SDP) [1]-[3]. Recently the SDFMM computer code has been successfully modified to handle subsurface sensing applications, in particular, the scattering from a landmine modeled as a PEC and/or penetrable spheroid buried under a two dimensional randomly rough ground [4]-[5]. The SDFMM has computational complexity for the CPU time and for the memory requirement equal to only $O(N)$ per iteration versus $O(N^2)$ for the MoM, where N is the total number of the unknowns [1]-[3]. The significant reduced complexity of the SDFMM over several other computational electromagnetics techniques has helped in achieving a fast and successful running for the Monte Carlo simulations [5]. However, the Monte Carlo sample needs in some cases to be greatly increased, e.g. when the ground random roughness increases the size of the Monte Carlo sample needs to be increased to achieve a converging solution. This could dramatically increase the required run time, especially when the dielectric constant of the ground is large and/or the penetrable buried object is electrically large. This necessitates a parallel

implementation of the SDFMM code to further reduce the overall runtime of large-scale 3-D scattering problems.

In this work, we used the MPI library for the parallel implementation of the SDFMM code [6]. The advantage of using the Beowulf cluster is that the system can be completely dedicated for the parallelization task, which is demonstrated in this work by executing small-scale cases due to memory limitations. Our emphasis is to demonstrate the overall speedup that can be achieved using the thirty-two processors [7]. Porting the parallelized code to the national supercomputers, where hundreds of processors and adequate RAM are available, will potentially facilitate the computations of large-scale problems [7]-[9].

The parallelization technique will be described in Section II, the numerical results will be shown in Section III and concluding remarks will be given in Section IV.

II. Parallelization Methodology

The SDFMM makes use of the equivalence theorem to calculate the electric and magnetic fields inside and outside a 3-D penetrable object buried under the rough surface interface [4]-[5]. The 3-D arbitrary object is modeled by scatterer R_3 that is immersed in scatterer R_2 which represents the rough ground which is immersed in the free space region represented by R_1 . The three regions, R_1 , R_2 and R_3 have permittivity and permeability given by ϵ_1 and μ_1 , ϵ_2 and μ_2 , and ϵ_3 and μ_3 , respectively, representing free space, soil medium and penetrable buried object. There are two final sets of unknown equivalent electric and magnetic surface currents in the following formulations. They are \bar{J}_1, \bar{M}_1 on the exterior of the rough ground interface between R_1 and R_2 , and \bar{J}_3, \bar{M}_3 on the exterior of the buried object interface between R_2 and

R_3 . Upon applying the boundary conditions, continuity of tangential components of the electric and magnetic fields on these interfaces, new integral equation formulations are obtained as [4]-[5]:

$$\bar{E}^{inc}(\bar{r})\Big|_{\text{tang.}} = \left[(L_1 + L_2)\bar{J}_1 - (K_1 + K_2)\bar{M}_1 - L_3\bar{J}_3 + K_3\bar{M}_3 \right]_{\text{tang.}} \quad (1a)$$

$$\bar{H}^{inc}(\bar{r})\Big|_{\text{tang.}} = \left[(K_1 + K_2)\bar{J}_1 + \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{M}_1 - K_3\bar{J}_3 - \frac{L_3}{\eta_2^2} \bar{M}_3 \right]_{\text{tang.}} \quad (1b)$$

$$0 = \left[-L_2\bar{J}_1 + K_2\bar{M}_1 + (L_3 + L_4)\bar{J}_3 - (K_3 + K_4)\bar{M}_3 \right]_{\text{tang.}} \quad (1c)$$

$$0 = \left[-K_2\bar{J}_1 - \frac{L_2\bar{M}_1}{\eta_2^2} + (K_3 + K_4)\bar{J}_3 + \left(\frac{L_3}{\eta_2^2} + \frac{L_4}{\eta_3^2} \right) \bar{M}_3 \right]_{\text{tang.}} \quad (1d)$$

In which the integro-differential operators L_i and K_i , $i=1, 2, 3$ and 4 , are given in detailed in [5]. In Eqs. 1a-d, the unknown surface electric and magnetic currents are \bar{J}_1 , \bar{M}_1 , \bar{J}_3 , and \bar{M}_3 , while the tangential component of the incident electric and magnetic fields on the rough surface are given by $\bar{E}^{inc}(\bar{r})\Big|_{\text{tang.}}$ and $\bar{H}^{inc}(\bar{r})\Big|_{\text{tang.}}$, respectively. The intrinsic impedance in each region is $\eta_i = \sqrt{\mu_i / \epsilon_i}$, $i=1, 2$, and 3 , where the dielectric permittivity and permeability in each region are ϵ_i and μ_i , respectively. The equivalent electric and magnetic currents are approximated using the RWG vector basis functions [10]-[11]. Upon applying Galerkin's method for testing and substituting the RWG surface current approximations in 1a-d, the original integral equations are transformed into a set of linear system of equations given by [4]-[5]:

$$\bar{\bar{Z}} \bar{I} = \bar{V} \quad (2a)$$

The impedance matrix $\bar{\bar{Z}}$ has order of $2(N+P) \times 2(N+P)$. The vector \bar{V} is a matrix of order $2(N+P) \times 1$ and composed of a submatrix of the tested tangential incident electric field \bar{E}^{inc} of order $N \times 1$ and a submatrix of the tested normalized magnetic field $\eta_1 \bar{H}^{inc}$ of order $N \times 1$, and a null submatrix of order $2P \times 1$. The quantities N and P are the number of basis functions (total number of edges of the triangular patches) on the surfaces of the rough ground and the buried object, respectively. If the MoM is used to solve Eq. (2a), it requires computing and storing all elements of the matrix $\bar{\bar{Z}}$ and then multiplying them by the vector \bar{I} , which is considered a very expensive computational process that will prohibit solving (2a) for large-scale problems. However, upon using the SDFMM [1]-[5], the matrix $\bar{\bar{Z}}$ becomes sparse and the system of equations in (2a) can be written as:

$$\bar{\bar{Z}}' \bar{I} + \bar{\bar{Z}}'' \bar{I} = \bar{V} \quad (2b)$$

where the matrix $\bar{\bar{Z}}'$ is a sparse matrix whose non-zero elements need to be calculated and stored using the conventional MoM and then multiplying them by the vector \bar{I} (near field interactions) while the matrix-vector multiply $\bar{\bar{Z}}'' \bar{I}$ is computed in one step without calculating or storing any elements of the matrix $\bar{\bar{Z}}''$. This is achieved by using the FMM hybridized with the SDP integration rule. This implies that there are three bottlenecks in the SDFMM code that can significantly enhance its performance by parallelization: (i) the subroutines that calculates the elements of the sparse matrix $\bar{\bar{Z}}'$; (ii) the subroutine that executes the matrix vector multiplication $\bar{\bar{Z}}' \bar{I}$ in each iteration of the solver; (iii) the subroutine that execute the fast multipole method for $\bar{\bar{Z}}'' \bar{I}$ (far field fast multipole interactions).

The computer code has been parallelized by exploiting the underlying available data parallelism. The key data structure in subroutine (i) is the sparse matrix $\bar{\bar{Z}}'$, which is stored as blocks of nonzero elements. These blocks are distributed among all processors, and no additional communication is needed. When this routine is parallelized we achieved near-linear speedups on 32 processors. In the matrix-vector multiplication $\bar{\bar{Z}}'\bar{I}$, the computation is parallelized by distributing \bar{I} to all processors in each iteration. The resulting vector components produced by the multiplication are then distributed to all processors. For bottleneck (iii), there are two involved subroutines to compute the far field interactions and they consist of a series of loops with complex interdependences. Each loop is separately parallelized, with collective communication used to distribute the results to all processors after executing each subroutine. In addition these two subroutines are executed in parallel, followed by another distribution of the results to all processors. Load balance between these two subroutines is achieved using a detailed performance model based on the serial execution time of each routine, the time required for collective communication operations, and the amount of communication overhead needed. Processors are assigned to these two subroutines such that the predicted maximum runtime for the two of them is minimized. The structure of the parallelized SDFMM application is shown in Fig. 1.

II. Numerical Results

We evaluated the parallel implementation of the SDFMM computer code on a 32-node Intel Pentium-based Beowulf cluster. Thirty one nodes of the Beowulf cluster are 350MHz Intel Pentium IIs with 256 MB of RAM in addition to one node of a 4x450MHz Intel Pentium II Xeon shared memory processor with 2GB of RAM. The nodes are connected to a 100 BaseTX

Ethernet network and they use the SuSE 6.1 operating system with Linux kernel 2.2.13, and the MPICH 1.2.1 implementation of the MPI library. Moreover, we tested the parallelized code on a 4-node shared memory Compaq Alpha-based workstation (667Mhz Alpha 21264) of 16GB total RAM. The processor uses the UNIX OSF/1 V5.1 operating system with the MPICH 1.1.2 MPI library.

Our benchmark includes three small-scale cases executed on the 256MB Intel cluster, and in addition one moderate-scale case that is executed on the Alpha workstation. To evaluate the speedup achieved by the parallel code, we considered a range of values for the ground roughness and/or for the buried object. All results obtained by executing the parallel version of the code are validated with those computed by the serial version of the code [4]-[5]. In all computations a 0.1% tolerance is assumed for the TFQMR iterative solver. The scattering problem configurations used in [5] are employed here, but for only one rough surface realization. The rough ground is characterized by Gaussian statistics with zero mean for the height, thus the roughness parameters can be described by the rms height (σ) and the correlation length (l_c). In all cases, the relative dielectric constant of the ground soil and the penetrable buried object (anti-personnel mine) are $\epsilon_r = 2.5 - j0.18$ and $\epsilon_r = 2.9 - j0.0092$, respectively. A Gaussian beam with horizontal polarization is employed for the incident waves at normal incidence for Cases 1-3 and at 10° from normal direction for Case 4 [5].

In the small-scale Cases 1-3, the dimensions of the modeled ground are assumed to be $3\lambda_0 \times 3\lambda_0$ leading to almost 8,800 of total number of surface current unknowns, while these dimensions are increased to be $8\lambda_0 \times 8\lambda_0$ for the moderate-scale Case 4 leading to 60,320 unknowns, where λ_0 is the free space wavelength. In Case 1, the scattered electric fields from a rough ground alone (no buried target) with $\sigma = 0.3\lambda_0$ and $l_c = 0.5\lambda_0$ are calculated at height of

$1.2\lambda_0$ above the ground. In Case 2, the scattered electric fields from a rough ground with a buried penetrable sphere are calculated at height of $0.5\lambda_0$ above the ground. The ground roughness are assumed to be $\sigma = 0.1\lambda_0$ and $l_c = 0.5\lambda_0$ and the sphere has radius of $a = 0.16\lambda_0$ with burial depth equal to $z = -0.32\lambda_0$ measured from its center to the mean plane of the ground. The sphere in Case 2 is replaced by a spheroid of dimensions $a = 0.3\lambda_0$ and $b = 0.15\lambda_0$ in Case 3 that is buried at $z = -0.3\lambda_0$ with ground roughness equal to $\sigma = 0.04\lambda_0$ and $l_c = 0.5\lambda_0$.

Both the overall speedup and the initial speedup (filling matrix $\bar{\bar{Z}}'$) are plotted versus the number of processors for Cases 1, 2 and 3 in Figs. 2a, 2b and 2c, respectively. The speedup is defined as the ratio of the serial runtime to the parallel runtime. The results in these figures show the significant speedup in the initial time (set up) that is consumed to fill the sparse matrix $\bar{\bar{Z}}'$ as explained in Section II. This initial speedup dramatically affects the overall speedup of the code as shown in these figures. In addition, the results show that almost the same overall speedup can be achieved by employing only twelve instead of thirty-two processors.

The efficiency for a given number of processors is defined as the ratio of the speedup to the number of processors. In each case, the peak speedup is achieved when running on 32 processors, where for case 1, the peak speedup is 7.1 as shown in Fig. 2a, with a reduction in runtime from 99 minutes on one processor to 14 minutes on 32 processors. For Case 2, the peak speedup is 6.2 as shown in Fig. 2b, with a reduction in runtime from 90 minutes to 14 minutes while for Case 3, the peak speedup is 7.2 as shown in Fig. 2c, with a reduction in runtime from 88 minutes to 12 minutes. Over these three cases, the average speedup on 32 processors is 6.8, giving an efficiency of 0.21. Based on the serial runtimes, 88% of the code is executed in parallel. Therefore by Amdahl's Law [12], the peak speedup achievable is 8.3. We conclude that

communication overhead and load imbalance among the processors accounts for the reduction in speedup from 8.3 to 6.8. An interesting comparison between the speedup achieved in each one of the bottlenecks (i)-(iii) mentioned in Section II, is shown in Fig. 3. These results show that overall speedup is almost the same as the speedup achieved in the matrix-vector multiplication $\bar{\bar{Z}}\bar{I}$ that is the bottleneck (ii).

In the second set of experiments, we solved the moderate-scale problem of Case 4 (60,320 unknowns) on the Alpha SMP using all four available processors. The penetrable spheroid of dimensions $a = 0.3\lambda_0$ and $b = 0.15\lambda_0$ is buried at $z = -0.3\lambda_0$ under the $8\lambda_0 \times 8\lambda_0$ rough ground with $\sigma = 0.04\lambda_0$ and $l_c = 0.5\lambda_0$. The magnitude of the total scattered electric field from the ground with the buried target is shown in Fig. 4a. However, the magnitude of the scattered electric fields for just the buried spheroid is computed by subtracting the return from the rough ground using complex vector representation from the total return from the ground with the buried target [4]-[5]. The output is shown in Fig. 4b. The results of Fig. 4a and 4b clearly demonstrate that the signature of the buried plastic landmine is relatively small compared with the return from the ground which is considered a major source of clutter in landmine detection application. Moreover, the distortion observed in Fig. 4b is due to the roughness of the ground which is modeled here as only one random rough surface realization, however the Monte Carlo simulations case was presented in [5]. The serial version took 96 minutes to run this case while the parallel version took 37 minutes, giving a speedup of 2.5 and an efficiency of 0.63. The predicted peak speedup on the four processors is 2.9. This implies that executing the parallel code on the 4-Alpha 667MHz processor gives a remarkable reduced absolute runtime for this moderate-scale case. This achievement can be exploited to execute large-scale scattering problem as mentioned in Section II. For the memory requirements, the serial version of the code

requires 950MB of RAM while the parallel version requires 1154MB of RAM distributed over the four processors as 288, 290, 289 and 287MB, respectively. Table I summarizes the parameters and output results for all cases presented in this section.

Table I

Case #	# of Unknowns	σ	Object	System	# of Processors	Serial/Par. Time (min.)	Speedup (overall)
1	8,800	$0.3\lambda_o$	None	Cluster	32	99/14	7.1
2	8,800	$0.1\lambda_o$	Sphere	Cluster	32	90/14	6.2
3	8,800	$0.04\lambda_o$	Spheroid	Cluster	32	88/12	7.2
4	60,320	$0.04\lambda_o$	Spheroid	Alpha Server	4	96/37	2.5

The results described in this section demonstrate that by implementing the fine grained parallelism, we have achieved significant speedups when using a single rough surface realization (one run of the code). However, when the number of rough surface realizations is much larger than the number of available processors, larger speedups are possible. In this case we assign a group of these realizations (runs of the code) to be executed in parallel on each processor. Since the computations are independent, this coarse grained parallelism gives a perfect speedup that is only limited by the number of available processors with the advantage that no communication is required. This situation occurs when we need to run Monte Carlo simulations [5]. In other subsurface scattering configurations, we may need to obtain multiple views of a target buried under the same rough surface realization [4], which requires running the code only few times. A combination of the fine-grained parallelism and coarse-grained parallelism can make efficient use of all available processors. Reasonable speedup can be achieved in this case where the number of needed runs is comparable to, or smaller than, the number of the available processors.

IV. Conclusions

Remarkable speedup has been achieved as the SDFMM is parallelized using the MPI library. The linear speedup obtained for the bottleneck associated with filling the impedance sparse matrix indicates a significant accomplishment as it dramatically affects the overall speed of the computer code. Reasonable speedups are obtained for the second and third bottlenecks associated with the matrix vector multiplication and the far field approximations, respectively. However, both speedups saturate upon using a number of processors less than the total number of nodes available on the system. This mechanism causes the overall runtime of the code to be larger than what we anticipated. Thus more parallelization effort is needed especially for the second bottleneck (matrix vector multiplication for the sparse matrix) in order to achieve as much linear overall speedup as possible.

ACKNOWLEDGMENTS

This research was sponsored in part by the College of Engineering at the University of Arkansas, in part by the Engineering Research Centers Program of the NSF under award number EEC-9986821, and in part by the Army Research Office Demining MURI grant # DAA 0-55-97-0013. This work benefited from the allocation of time at the Jouliau Cluster and at the Advanced Scientific Computation Center (NU-ASCC) at Northeastern University.

The authors would like to thank Prof. W. Chew and Prof. E. Michielssen at UIUC for allowing them the use and modification of the SDFMM computer code for the current application.

References

- [1] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [2] V. Jandhyala, B. Shanker, E. Michielssen, and W. C. Chew, "A fast algorithm for the analysis of scattering by dielectric rough surfaces," *J. Opt. Soc. Am. A*, vol. 15, no. 7, pp. 1877-1885, July 1998.
- [3] M. El-Shenawee, V. Jandhyala, E. Michielssen and W. C. Chew, "The steepest descent fast multipole method (SDFMM) for solving combined field integral equation pertinent to rough surface scattering," *Proc. of the IEEE APS/URSI '99 conf.*, Orlando, Florida, pp. 534-537, July 1999.
- [4] M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, "3-D subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces: Use of the steepest descent fast multipole method (SDFMM)," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 6, pp. 1174-1182, June 2001.
- [5] M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo simulations of electromagnetic wave scattering from random rough surface with 3-D penetrable buried object: Mine detection application using the SDFMM," *J. Opt. Soc. Am. A*, to appear in December 2001.
- [6] Message Passing Interface Forum. MPI: A Message-Passing Interface standard. *The International Journal of Supercomputer Applications and High Performance Computing*, no. 8, 1994.

- [7] S. V. Velamparambil, J. E. Schutt-Aine, J. G. Nickel, J. M. Song, and W. C. Chew, "Solving large scale electromagnetic problems using a linux cluster and parallel MLFMA," *IEEE Antennas Propagat. Symp.*, 1:636-639, July 1999.
- [8] S. Velamparambil, J. Song, and W. C. Chew, "On the parallelization of dynamic multilevel fast multipole method on distributed memory computers," *Proc. International Workshop on Innovative Architecture for Future Generation High-Performance Processors and Systems, Maui, Hawaii*, November 1999.
- [9] S. Li, C. H. Chan, L. Tsang, Q. Li, and L. Zhou, "Parallel Implementation of the sparse matrix/canonical grid method for the analysis of two-dimensional random rough surfaces (three-dimensional scattering problem) on a Beowulf system," *IEEE Trans. Geoscience Rem. Sensing*, vol. 38, no. 4, pp.1600-1608, July 2000.
- [10] L. Medgyesi-Mitschang, J. Putnam, and M. Gedera, "Generalized method of moments for three-dimensional penetrable scatterers," *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1383-1398, April 1994.
- [11] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Anten. and Prop.*, vol. AP 30, no. 3, pp.409-418, May 1982.
- [12] G. M. Amdahl, "Validity of the single processor approach to achieving large scale computing capabilities", *Proc. AFIPS Spring Joint computer Conference* 30, Atlantic City, N. J., pp. 483-485, April 1967.

List of Figures:

Figure 1 Structure of parallelized SDFMM showing major computational tasks and their interrelation.

Figure 2 Speedup of the Beowulf cluster: (a) Case 1, target-free rough ground, (b) Case 2, penetrable sphere buried under moderately rough ground surface, (c) Case 3, penetrable spheroid buried under slightly rough ground surface.

Figure. 3 Performance improvement for each of the separate component tasks and overall speedup of the SDFMM algorithm, as a function of the number processors in the Beowulf cluster.

Figure 4 The near electric field scattered above the rough ground at $z = 0.5\lambda_0$ for the spheroid of Case 4 ($a = 0.3\lambda_0$, $b = 0.15\lambda_0$, buried at $z = -0.3\lambda_0$ in conductive clay loam soil), computed using the 4-processor Alpha Server: (a) the ground with the buried spheroid (total field), (b) just the spheroid obtained by subtraction.

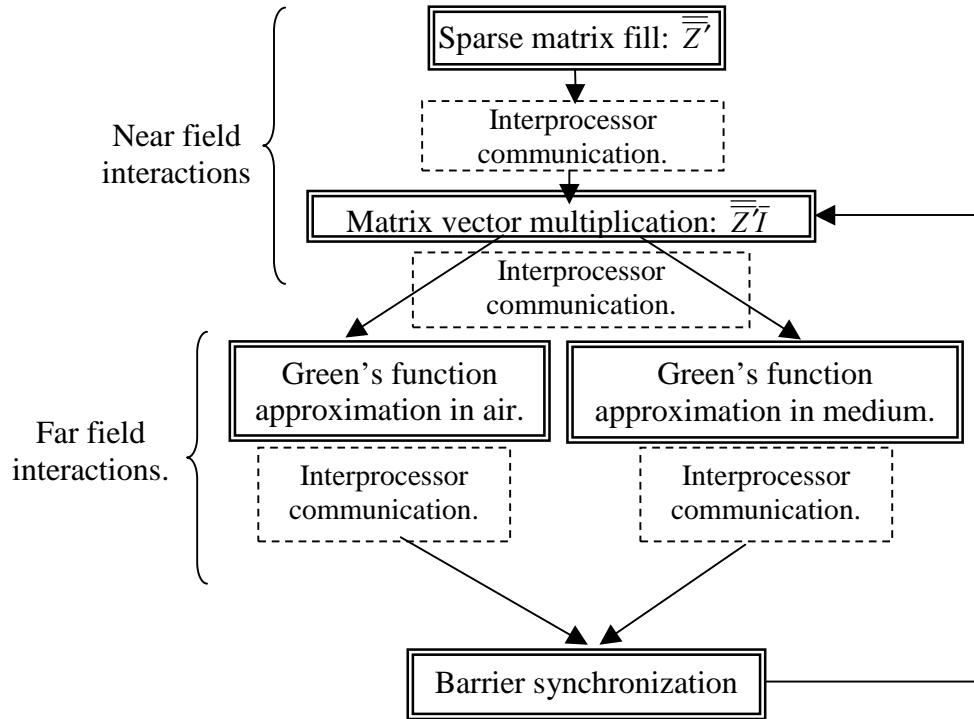


Fig. 1 Structure of parallelized SDFMM showing major computational tasks and their interrelation.

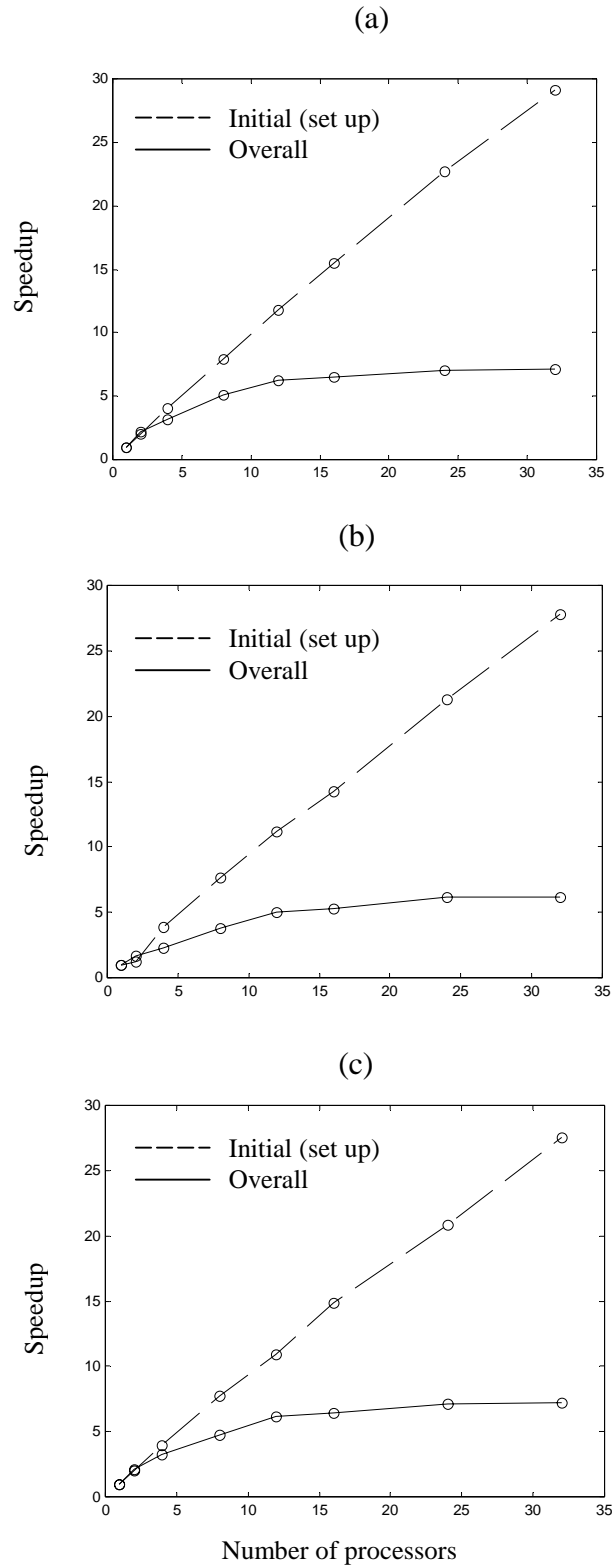


Fig. 2 Speedup of the Beowulf cluster: (a) Case 1, target-free rough ground, (b) Case 2, penetrable sphere buried under moderately rough ground surface, (c) Case 3, penetrable spheroid buried under slightly rough ground surface.

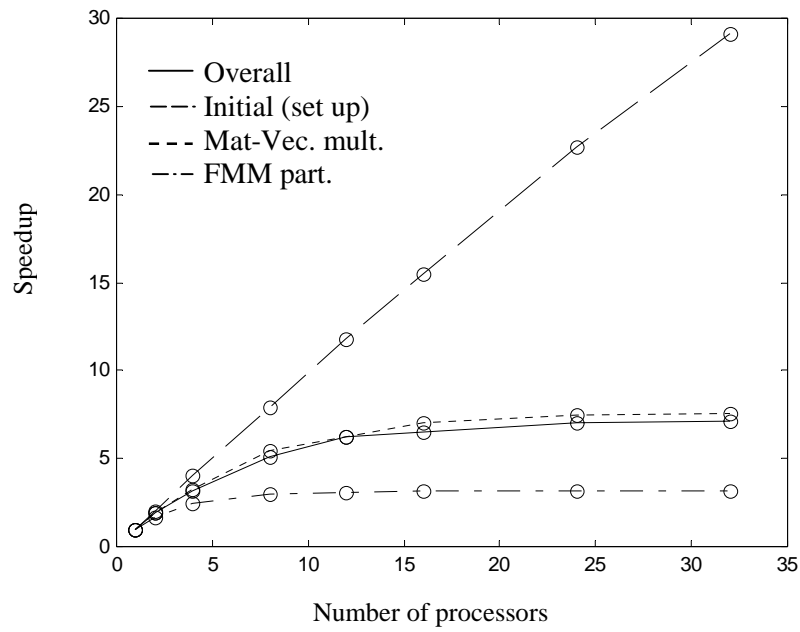


Fig. 3 Performance improvement for each of the separate component tasks and overall speedup of the SDFMM algorithm, as a function of the number processors in the Beowulf cluster.

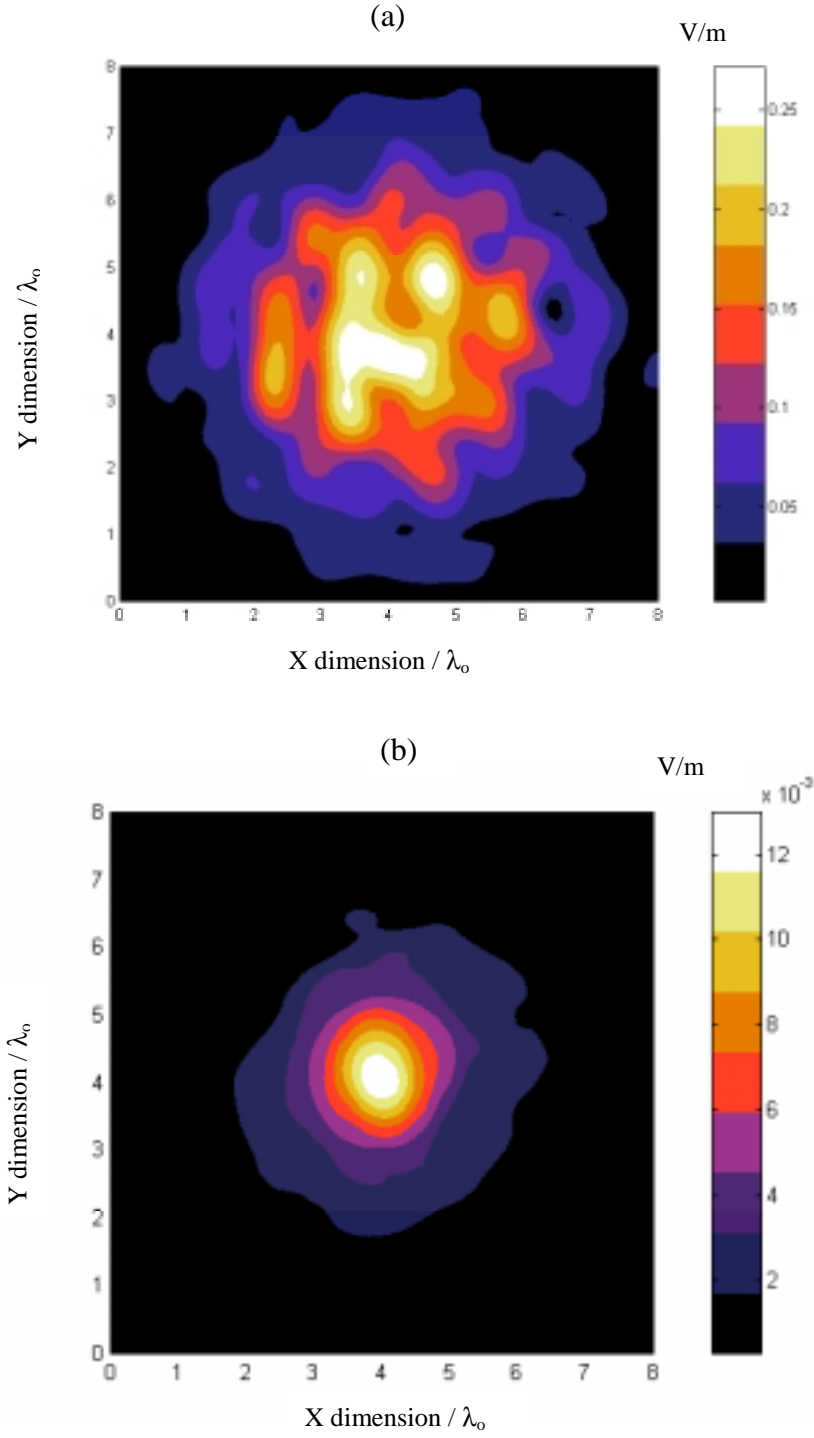


Fig. 4 The near electric field scattered above the rough ground at $z = 0.5\lambda_0$ for the spheroid of Case 4 ($a = 0.3\lambda_0$, $b = 0.15\lambda_0$, buried at $z = -0.3\lambda_0$ in conductive clay loam soil), computed using the 4-processor Alpha Server: (a) the rough ground with the buried spheroid (total field), (b) just the spheroid obtained by subtraction.

Modeling Clutter from Bosnian and Puerto Rican Rough Ground Surfaces for GPR Subsurface Sensing Applications Using the SDFMM

Magda El-Shenawee¹ and Carey Rappaport²

¹Department of Electrical Engineering
3217 Bell Engineering Center
University of Arkansas
Fayetteville, Arkansas 72701
Tel: (501)-575-6582, Fax: (501)-575-7967
magda@uark.edu

²Center for Subsurface Sensing and Imaging Systems
Department of Electrical and Computer Engineering
Northeastern University, 235 Forsyth BLD
Boston, MA 02115
Tel: (617)-373-2043, Fax: (617)-373-8627
rappaport@neu.edu

Abstract:

The Steepest Descent Fast Multipole Method (SDFMM) is used to analyze the distorting effect of random rough ground surfaces on scattered and transmitted electromagnetic waves. Two well-measured loamy soils: Bosnian and Puerto Rican clay loam are investigated, each with a variety of surface roughness. This study is important in understanding the effects of different soil properties and is meant to be an *a priori* phase of investigating scattering from buried targets under the rough ground. In this work, we investigated the scattering from rough soil ground without buried objects. The SDFMM is an integral equation-based fast algorithm that is well suited for two-dimensional penetrable rough surfaces (3-D scattering) in the frequency domain. The scattered and transmitted near electric field of an incident Gaussian beam are calculated at different locations above and below the mean plane of the dielectric rough interface. The

receiver locations above are chosen to simulate GPR measurement protocols. The obtained numerical results show that the scattered field undergoes more distortion than the transmitted field from both soil types. Moreover, the transmitted fields into the higher dielectric constant Puerto Rican soil experience more distortion than those transmitted into Bosnian soil.

Key words: Rough surface scattering, GPR, EM modeling, fast algorithms.

I. INTRODUCTION

When electromagnetic waves are used to sense buried objects, the effect of the random rough air/ground interface on both the transmitted and scattered fields is great, and must be carefully considered. In many cases, the scattering from the air/ground interface is larger than the scattering of a buried target especially when the ground surface roughness is comparable to the height and the burial depth of the target. Accurate modeling of the clutter is necessary to enable signal-processing algorithms to perform optimally in detecting buried targets. In this work, we investigate the clutter of different types of soil, which is scattering from rough surface without any buried targets.

The rough surface causes two types of distorting effects on a probing plane wave: a defocused scattered wave reflected back into the air, and a distorted transmitted wave that would continue until it encounters a target. Uncertainty is generated in both waves, and it is important to understand the relative amounts of distortion between the two, as well as its dependence on soil type. While there has been much published work on far field rough surface scattering [1]-[8], the near field effects— which are essential in inverse scattering and object sensing applications—

have not been well studied. However, there is published work for modeling near field from targets buried under planar surface [9]. As an effort to study the scattered fields in the near zone, we assumed that the incident waves are still far fields while point receivers are located in the near zone above and below the rough ground. No interactions between scattered fields and the transmitting antenna are accounted for in this work.

The surface scattering depends on the frequency dependent complex dielectric constant of the soil, which in turn depends on its composition, moisture and density, and also on the roughness of the surface. As moisture increases, so does the dielectric constant. And with a larger dielectric constant, the wavelength in soil decreases. For a given illuminated portion of ground, wetter soil will require finer computational discretization, and greater computational expense. However, using special-purpose numerical methods, the fast computation of wave interaction with a rough surface is feasible. We study two types of experimentally measured soil: Bosnian loam from the Alicia test site [10], and Puerto Rican clay loam [11].

The rough surface is characterized with Gaussian statistics in terms of random heights and auto-correlation function [12]. The roughness parameters (root mean square height and correlation length) are chosen to be in the moderate roughness range, with root mean square height ranging from 0.1 to 0.2 free space wavelengths and the correlation length kept constant at one free space wavelength.

In this work, a specialized fast algorithm is necessary, since multiple calculations of this realistic, but computationally intensive, model are needed for many different realizations of surface

roughness. There for we used the SDFMM, which is a hybridization of the Method of Moments (MOM), the Steepest Descent Path (SDP) method, and the Fast Multipole Method (FMM) [13], [14], to calculate the unknown surface current coefficients. The details of the SDFMM can be found in [15]-[18].

II. FORMULATION

In this work, the well-known PMCHW (Poggio, Miller, Chang, Harrington, and Wu) integral equations are implemented to calculate the electric and magnetic surface currents [16], [19]. For convenience, the integral equation are summarized here as:

$$\bar{E}^{inc}(\bar{r})\Big|_{tan.} = \left[(L_1 + L_2)\bar{J}_1 - (K_1 + K_2)\bar{M}_1 \right]_{tan.} \quad (1)$$

and

$$\bar{H}^{inc}(\bar{r})\Big|_{tan.} = \left[(K_1 + K_2)\bar{J}_1 + \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{M}_1 \right]_{tan.} \quad (2)$$

in which $\bar{E}^{inc}(\bar{r})$ and $\bar{H}^{inc}(\bar{r})$ are the incident electric and magnetic fields on the interface between air and soil as shown in Figure (1). The subscript *tan* means the tangential component.

The differentio-integral operators L_1, L_2, K_1 and K_2 are defined as [19]:

$$L_{1,2}\bar{X} = \int_S \left\{ i\omega\mu_{1,2}\Phi_{1,2}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{1,2}}\nabla\nabla' \cdot \bar{X}(\bar{r}')\Phi_{1,2} \right\} ds' \quad (3a)$$

$$K_{1,2}\bar{X} = \int_S \bar{X}(\bar{r}') \times \nabla\Phi_{1,2} ds' \quad (3b)$$

where, the vector \bar{X} represents the surface electric current \bar{J} or the surface magnetic current \bar{M} on the interface S . The dielectric permittivity and permeability in each region are ϵ_i and μ_i , $i=1$ and 2. The 3-D scalar Green's function in $\Phi_i(\bar{r})$ is given by :

$$\Phi_i(\bar{r} - \bar{r}') = \frac{\exp(-ik_i|\bar{r} - \bar{r}'|)}{4\pi|\bar{r} - \bar{r}'|} \quad (3c)$$

in which \bar{r} is the field point, \bar{r}' is the source point, $k_i = \omega\sqrt{\epsilon_i\mu_i}$ is the wave number in each region, $i=1$ and 2 . The equivalent electric and magnetic currents \bar{J} and \bar{M} on S , are approximated using the RWG vector basis functions $\bar{j}(\bar{r})$ [19], [20] as follows

$$\bar{J}(\bar{r}) = \sum_{n=1}^N I_{1n} \bar{j}_n(\bar{r}), \quad \bar{r} \in S \quad (4a)$$

$$\bar{M}(\bar{r}) = \eta_1 \sum_{n=1}^N I_{2n} \bar{j}_n(\bar{r}), \quad \bar{r} \in S \quad (4b)$$

where η_1 is the intrinsic impedance of the free space. Upon substituting Eq. 3-4 in Eqs. (1) and (2) and testing with the same basis functions, we get the $2N \times 2N$ system of equations as

$$\bar{\bar{Z}}\bar{I} = \bar{V} \quad (5)$$

in which \bar{V} represents the tangential component of the electric and magnetic fields, \bar{I} represents the unknown coefficients, and $\bar{\bar{Z}}$ is the impedance matrix [19]. Solving Eq.(5) using the method of moments is extremely expensive (computer memory and CPU time) when the size of the scatterer (the rough interface, in this case) is much larger than the free space wavelength. In reality, the size of the ground is infinite but for computational consideration the rough ground is assumed to be adequately larger than the footprint of the incident Gaussian beam [21]. As the number of unknown coefficients is very large, special computational procedures must be used such as the SDFMM, which is discussed in [15]-[18] and recently was implemented in [22].

III. NUMERICAL RESULTS

We consider the scatterer S to be a penetrable random rough surface with dimensions equal to $L \times L$, where $L = 8\lambda_0$, and λ_0 is the free space wavelength. Exciting the surface with a carefully tapered incident Gaussian beam minimizes the edge effect of the surface S on the surface currents [21]. Thus, infinitely large rough surfaces are sampled over a manageable region. The random rough surface is characterized with Gaussian statistics for the random heights and for the autocorrelation function, and is generated using the computer random number generator following the technique in [12]. Sixty Monte Carlo rough surface simulations were calculated for each pair of Gaussian parameters: root mean square height σ and correlation length l_c . For cases considered here, the electric field is assumed to be x-directed, and normally incident ($\theta^i = 0$) on the nominal ground surface.

Similar to the work in [16], which implemented the SDFMM to calculate the radar cross section of a penetrable random rough surface, we also are using the SDFMM but to calculate the near electromagnetic fields scattered above and transmitted into the ground.

Point receivers are located in the near zone at one wavelength above and below the mean ground with resolution equal to $x = y = 0.1\lambda_0$. The electric field is calculated using the nearfield formulas in [23]. The SDFMM is compared with the sparse matrix canonical grid (SMCG) method [1] and excellent agreement is observed as shown in Fig. 2. The incident angle is $\theta^i = 20^\circ$, soil relative dielectric constant is $\epsilon_r = 2.0 - i0.2$, and surface roughness parameters are rms height $\sigma = 0.02\lambda_0$ and correlation length $l_c = 0.5\lambda_0$. The comparison in Fig. 2 shows the normalized radar cross section (RCS) for the HH-polarization case with horizontal axis to be the scatter angle θ (elevation angle in degrees) measured from the z -axis.

In this work, two types of soils have been studied: Bosnian soil with density 1.4g/cm^3 and 3.8% moisture level with complex dielectric constant $\epsilon_r = 3 - i0.18$ at 1 Ghz [10], and Puerto Rican clay loam with density 1.2 g/cm^3 and 10% moisture with dielectric constant $\epsilon_r = 5.4 - i0.04$ at frequency $f = 960\text{ Mhz}$ [11]. Three values of the rough ground rms height are assumed: $\sigma = 0.1\lambda_0$, $0.12\lambda_0$, and $0.17\lambda_0$ along with constant correlation length $l_c = 1\lambda_0$.

In Figs. 3a-c, the mean scattered fields in air from Bosnian soil with different rms random heights are shown, while Figs. 3d-f shows the transmitted fields. Note that for a perfectly flat ground surface, $\sigma = 0$; the contours would be perfect circles. In fact, with enough Monte Carlo runs, the mean scattered and transmitted field would be expected to be circularly symmetric for normal incidence. However, these mean fields show significant asymmetry and distortion, indicating that the random rough surface causes so much variation in the scattered fields that 60 runs is insufficient for convergence. As the rms height of the rough surface increases, the distortion in both the mean scattered and transmitted field increases as well. The asymmetry of the mean scattered and transmitted wave patterns give a qualitative sense of the distorting effects of the rough surface. Moreover, the results show that the waves transmitted into the ground, Figs. 3d-f, are less distorted than the scattered fields above the ground. This could be explained by noting that for rays incident on locally tilted surface patches (the sides of small hills and valleys), the reflected rays are bent further away from the nominal surface normal than the transmitted rays. As the surface roughness increases, the slopes of the surface patches increase, and the reflected rays diverge more. In Figs. 4a-c and 4d-f, the scattered and transmitted fields from Puerto Rican soil with the same three rms random heights are shown. Comparing results of Fig. 3 and Fig. 4, the magnitude of the scattered fields from the Puerto Rican soil is observed to

be larger--and the transmitted fields smaller--than that of the Bosnian soil. This is consistent with the fact that the dielectric constant for Puerto Rican soil is almost double that of Bosnian soil. Greater distortion in the transmitted fields in the Puerto Rican soil is observed as compared with the Bosnian soil case (Figs. 4d-f versus Figs. 3d-f). This results from the larger electrical distance in the former medium from the rough surface scattering centers to the observation plane one wavelength in air below the ground. The destructive interference of the transmitted Gaussian beam is enhanced when the phase differences from various parts of the surface proportionally increase.

Although the conductivity in soil plays an important role in subsurface sensing, it does not appear to have a dominant effect in the distortion of the transmitted Gaussian beam. A test of 60 Monte Carlo SDFMM runs of a fictitious soil medium with the real dielectric constant of Bosnian soil, but with the loss of the Puerto Rican clay loam indicates that the transmitted field has about the same distortion as for the actual Bosnian soil.

The total CPU time needed for filling the impedance matrix, for the iterative solution using the transpose free quasi-minimal residual solver (TFQMR) [24], and for calculating scattered and transmitted near fields is plotted versus the rough surface rms height and shown in Fig. 5. The discretization rate of the surface current, number of electric and magnetic surface current unknowns, and the required computer memory to run the SDFMM code are given in Table 1. As expected, running the SDFMM code for the Puerto Rican soil required more CPU time and computer memory due to its larger relative dielectric constant requiring a finer discretization.

The NCSA SGI/CRAY Origin 2000 machines are used for the Puerto Rican case due its larger memory requirement, but the Compaq GS140 EV6 machine at Northeastern University is used for the Bosnian case, which reflects the results shown in Fig. 5.

III. CONCLUSIONS

The SDFMM fast algorithm is used to calculate the nearfield scattered from and transmitted into two types of rough ground soils. Bosnian and Puerto Rican soils were chosen for this study as *a priori* phase of our investigation of scattering from buried mines under the same types of soils [22]. The results show that the distortion in both the scattered and the transmitted fields increases as the ground roughness increases. We observed that the scattered fields experience more distortion than the transmitted ones. This suggests a future research to investigate how will the transmitted fields be changing with the increase of the incident angle. This result indicates that although rough surfaces randomly distort both the scattered and transmitted waves, the distortion is greater for the former. Thus the uncertainty of the surface clutter signal is greater than that of the probing signal, suggesting the relative importance of careful characterization of the scattered waves.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. W. Chew, Prof. E. Michielssen, and Dr. V. Jandhyala at UIUC for allowing us to use and modify the SDFMM computer code for the current application. This research was sponsored by the Army Research Office Demining MURI grant # DAA 0-55-97-0013 and in part by the Engineering Research Centers Program of the NSF under award number EEC-9986821. The computational work was partially supported by National

Computational Science Alliance under grant # pt1 and utilized the NCSA SGI/CRAY Origin 2000 and the allocation of time at the Northeastern University Advanced Scientific Computation Center (NU-ASCC).

REFERENCES

- [1] G. Zhang, L. Tsang, and K. Pak, "Angular correlation function and scattering coefficient of electromagnetic waves scattered by a buried object under a two-dimensional rough surface," *J. Opt. Soc. Am. A.*, pp. 2995-3002, December 1998.
- [2] Y. Zhang and E. Bahar, "Mueller matrix elements that characterize scattering from coated random rough surfaces," *IEEE Trans. on Antenn. and Propag.*, vol. 47, no. 5, pp. 949-955, May 1999.
- [3] A. Ishimaru and J. S. Chen, "Scattering from very rough metallic and dielectric surfaces: a theory based on the modified Kirchhoff approximation," *Waves in Random Media*, pp. 21-34, January 1991.
- [4] S. Mainguy and J. Greffet, "A numerical evaluation of Raleigh's theory applied to scattering by randomly rough dielectric surfaces," *Waves in random Media*, pp.79-101, January 1998.
- [5] D. A. Kapp and G. S. Brown, "A new numerical method for rough surface scattering calculations," *IEEE Trans. on Antenn. and Propag.*, vol. 44, no. 5, pp. 711-721, May 1996.
- [6] A. K. Fung, M. R. Shah, and S. Tjuatja, "Numerical simulations of scattering from three-dimensional randomly rough surfaces," *IEEE Trans. Geoscience Rem. Sensing*, vol. 32, no. 5, pp. 986-994, September 1994.
- [7] M. Nieto-Vesperinas and J. A. Sa'nchez-Gil, "Intensity angular correlations of light multiply scattered from random rough surfaces," *J. Opt. Soc. Am. A*, vol. 10, no. 1, pp. 150-157, January 1993.
- [8] J. T. Johnson, "Third order small perturbation method for scattering from dielectric rough surfaces," *J. Opt. Soc. Am. A*, vol. 16, no. 11, pp. 2720-2736, November 1999.

- [9] Z. Xiong and A. C. Tripp, "3-D electromagnetic modeling for near-surface targets using integral equations," *Geophysics*, vol. 62, no. 4, pp. 1097-1106, July-August 1997.
- [10] J. Curtis, "Dielectric properties of soils; various sites in Bosnia," *US Army Corp. of Eng., Waterways Experim. Station Data Rep.*, 1996.
- [11] J. E. Hipp, "Soil electromagnetic parameters as functions of frequency: soil density and soil moisture," *Proceedings of the IEEE*, vol. 62, No. 1, pp. 98-103, January 1974.
- [12] N. Garcia and E. Stoll, "Monte Carlo calculation for electromagnetic wave scattering from random rough surfaces," *Phys. Rev. Lett.*, vol. 52, no. 20, pp. 1798-1801, May 1984.
- [13] V. Rokhlin, "Rapid solution of integral equations of scattering theory in two dimensions," *J. Comput. Phys.*, vol. 36, pp. 414-439, 1990.
- [14] C. C. Lu and W. C. Chew, "Fast algorithm for solving hybrid integral equations," *IEE Proceedings-H*, vol. 140, no.6, December 1993.
- [15] V. Jandhyala, Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures, PhD Thesis, Department of Electrical Engineering, University of Illinois at Urbana-Champaign, 1998.
- [16] V. Jandhyala, B. Shanker, E. Michielssen, and W.C. Chew, "A fast algorithm for the analysis of scattering by dielectric rough surfaces," *J. Opt. Soc. Am. A*, vol. 15, no. 7, pp. 1877-1885, July 1998.
- [17] M. El-Shenawee, V. Jandhyala, E. Michielssen, and W.C. Chew, "An enhanced steepest descent fast multipole method for the analysis of scattering from two dimensional multilayered rough surfaces," *Proceedings of the URSI '98 conference*, Atlanta, pp. 182, June 1998.

- [18] M. El-Shenawee, V. Jandhyala, E. Michielssen, and W.C. Chew, "The steepest descent fast multipole method (SDFMM) for solving combined field integral equation pertinent to rough surface scattering," *Proceedings of the IEEE APS'99*, Orlando, July 1999.
- [19] L. Medgyesi-Mitschang, J. Putnam, and M. Gedera, "Generalized method of moments for three-dimensional penetrable scatterers," *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1383-1398, April 1994.
- [20] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Anten. And Prop.*, vol. AP 30, no. 3, pp.409-418, May 1982.
- [21] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.
- [22] M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, "3-D subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces: use of the steepest descent fast multipole method (SDFMM)," to be published in the *IEEE Trans. Geoscience Rem. Sensing*, June 2001.
- [23] C. A. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons Inc, ch. 6, pp.254-309, 1989.
- [24] R. W. Freund, "A transpose-free quasi-minimal residual algorithm for non-hermitian linear systems," *SIAM J. Sci. Comput.*, vol. 14, no. 2, pp. 470-482, March 1993.

List of Figures

Figure 1 Cross section of rough soil ground with Gaussian beam incident at θ^i .

Figure 2 Normalized RCS for a rough surface of root mean square height $\sigma=0.02\lambda_0$ and correlation length $l_c=0.5\lambda_0$, at incident angle $\theta^i = 20^\circ$. HH polarization.

Figure 3a Near field scattered at $z = 1\lambda_0$ above the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.1\lambda_0$ and correlation length $l_c=1\lambda_0$, and incident angle $\theta^i = 0$.

Figure 3b Near field scattered at $z = 1\lambda_0$ above the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.12\lambda_0$ and correlation length $l_c=1\lambda_0$, and incident angle $\theta^i = 0^\circ$

Figure 3c Near field scattered at $z = 1\lambda_0$ above the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.17\lambda_0$ and correlation length $l_c=1\lambda_0$, and incident angle $\theta^i = 0$.

Figure 3d Near field transmitted at $z = -1\lambda_0$ below the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.1\lambda_0$ and correlation length $l_c=1\lambda_0$, and incident angle $\theta^i = 0$.

Figure 3e Near field transmitted at $z = -1\lambda_0$ below the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.12\lambda_0$ and correlation length $l_c=1\lambda_0$, and incident angle $\theta^i = 0$.

Figure 3f Near field transmitted at $z = -1\lambda_0$ below the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.17\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

Figure 4a Near field scattered at $z = 1\lambda_0$ above the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.1\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

Figure 4b Near field scattered at $z = 1\lambda_0$ above the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.12\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

Figure 4c Near field scattered at $z = 1\lambda_0$ above the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.17\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

Figure 4d Near field transmitted at $z = -1\lambda_0$ below the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.1\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

Figure 4e Near field transmitted at $z = -1\lambda_0$ below the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.12\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

Figure 4f Near field transmitted at $z = -1\lambda_0$ below the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.17\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

Figure 5 Total CPU time versus the rms height of the rough surface.

Table 1

	Bosnian soil	Puerto Rican soil
Discretization rate	$14/\lambda_0$	$18.5/\lambda_0$
Number of surface current unknowns	73,382	132,610
Required computer memory	800MB	1.76GB

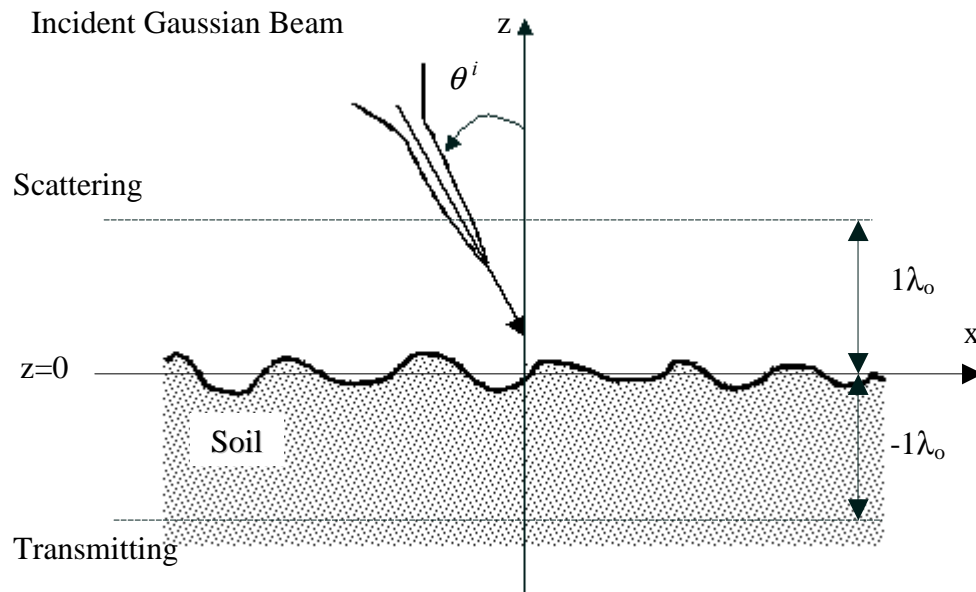


Figure 1 Cross section of rough soil ground with Gaussian beam incident at θ^i .

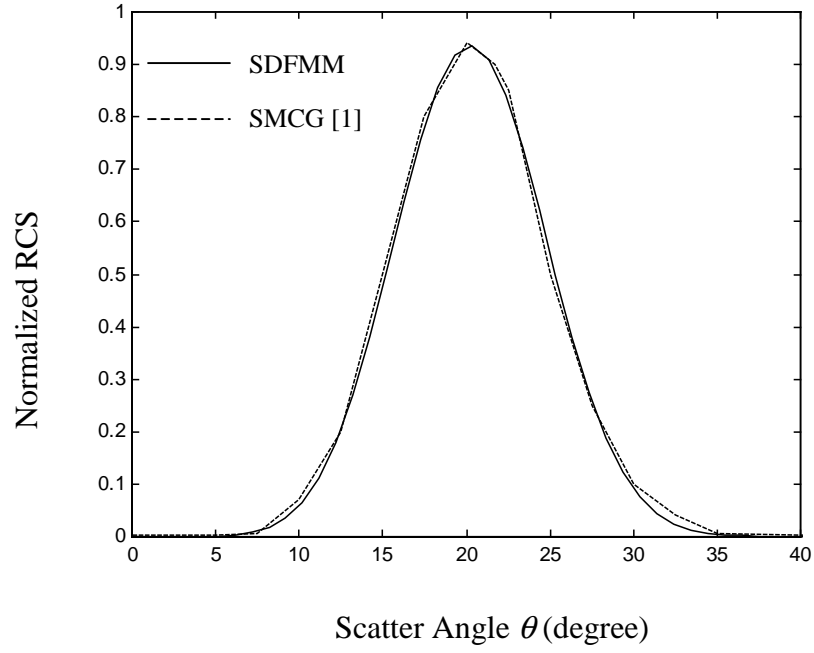


Figure 2. Normalized RCS for a rough surface of root mean square height $\sigma=0.02\lambda_0$ and correlation length $l_c=0.5\lambda_0$, at incident angle $\theta^i = 20^\circ$. HH polarization.

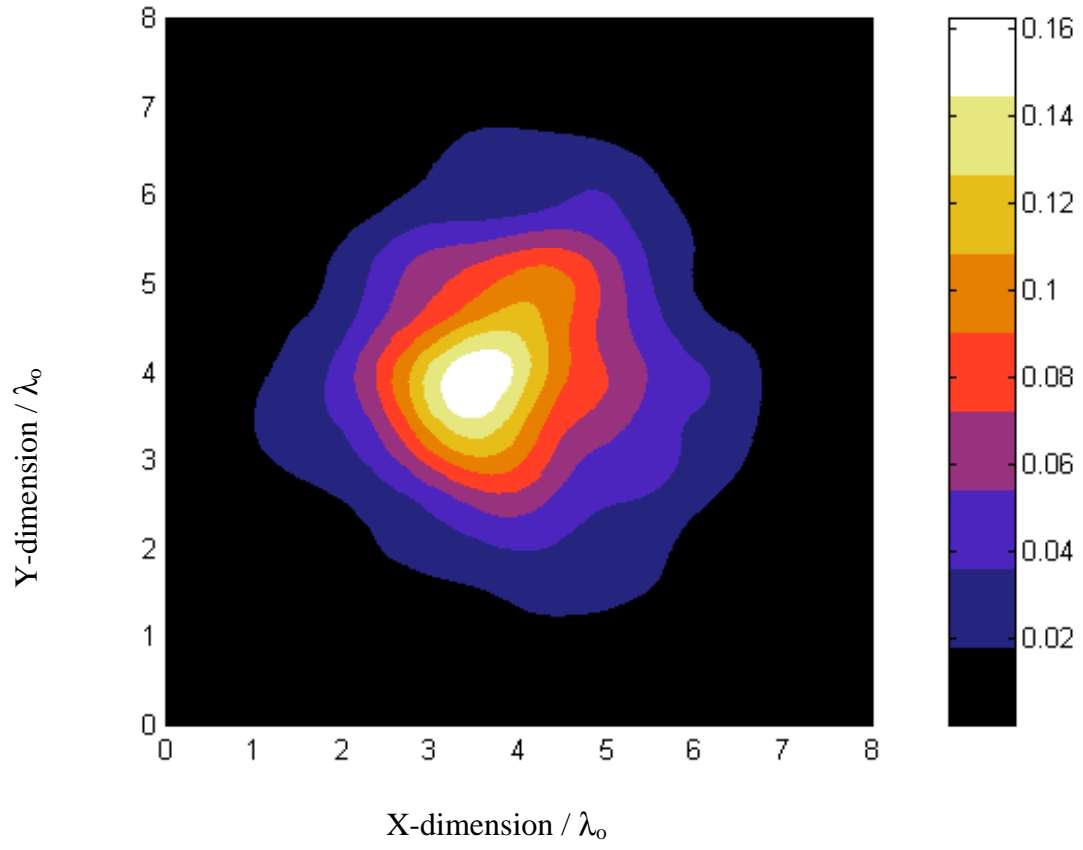


Figure 3a Near field scattered at $z = 1\lambda_0$ above the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.1\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

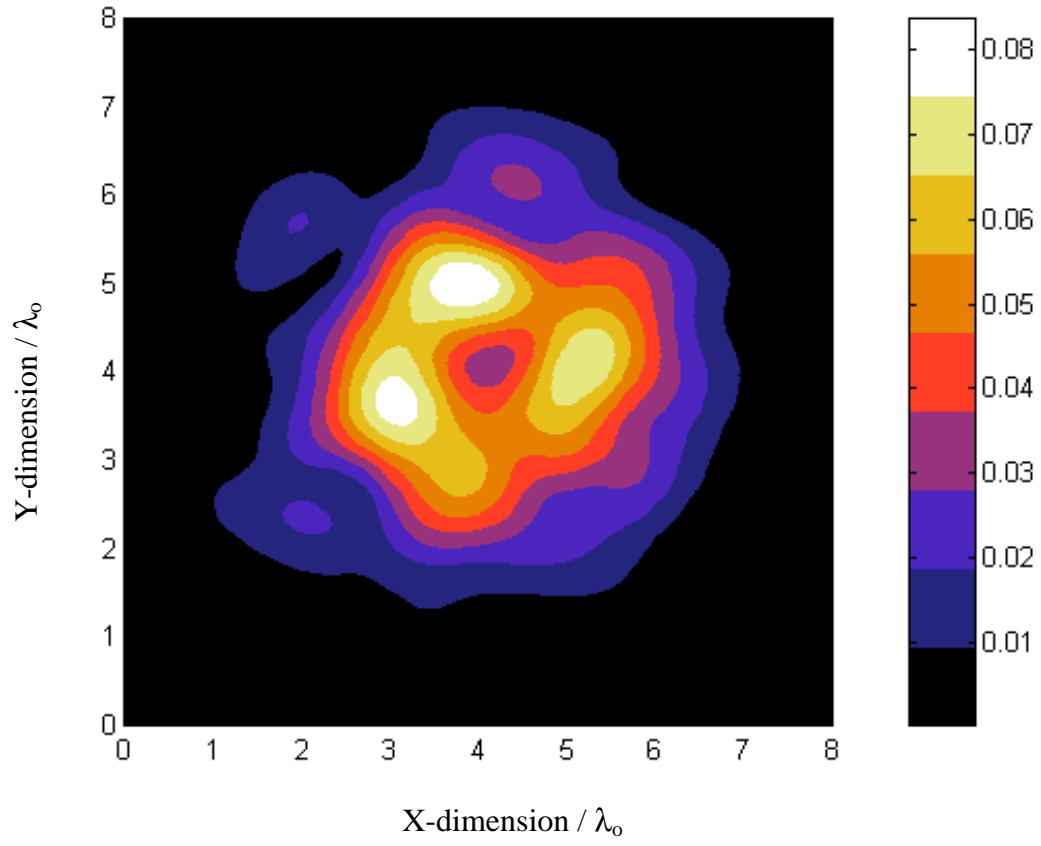


Figure 3b Near field scattered at $z = 1\lambda_0$ above the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.12\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

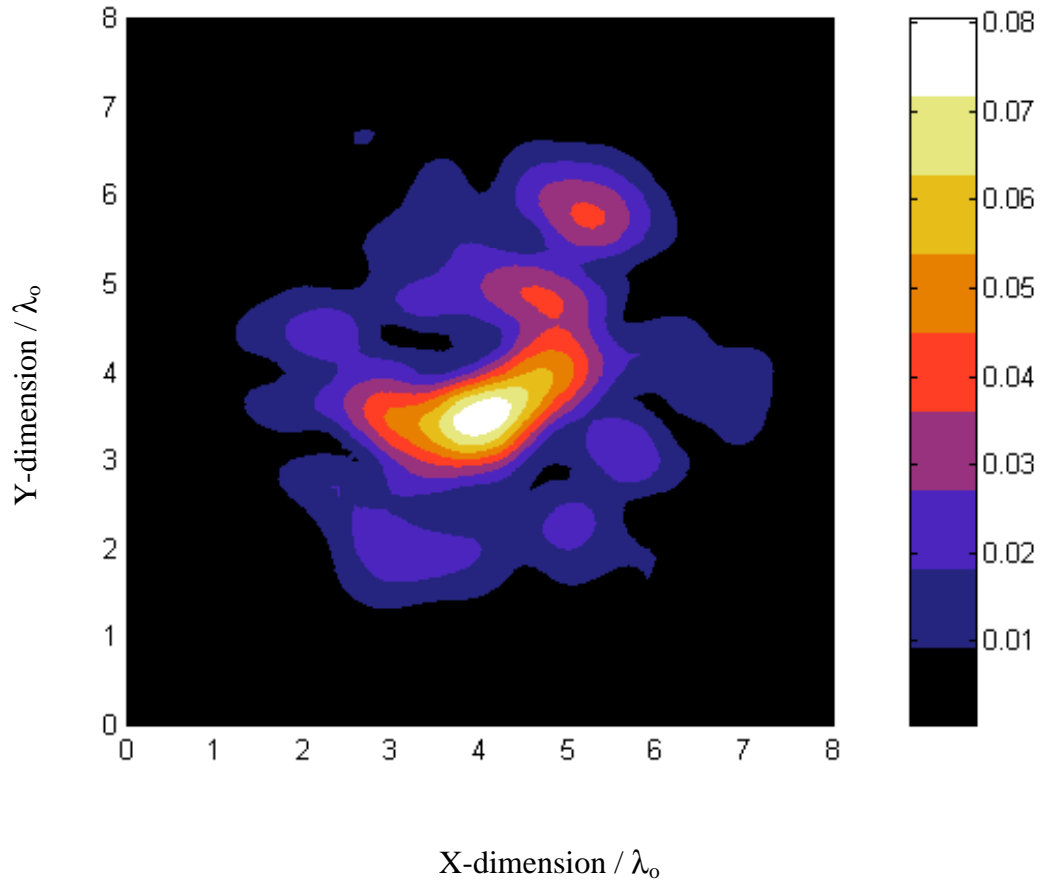


Figure 3c Near field scattered at $z = 1\lambda_0$ above the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.17\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

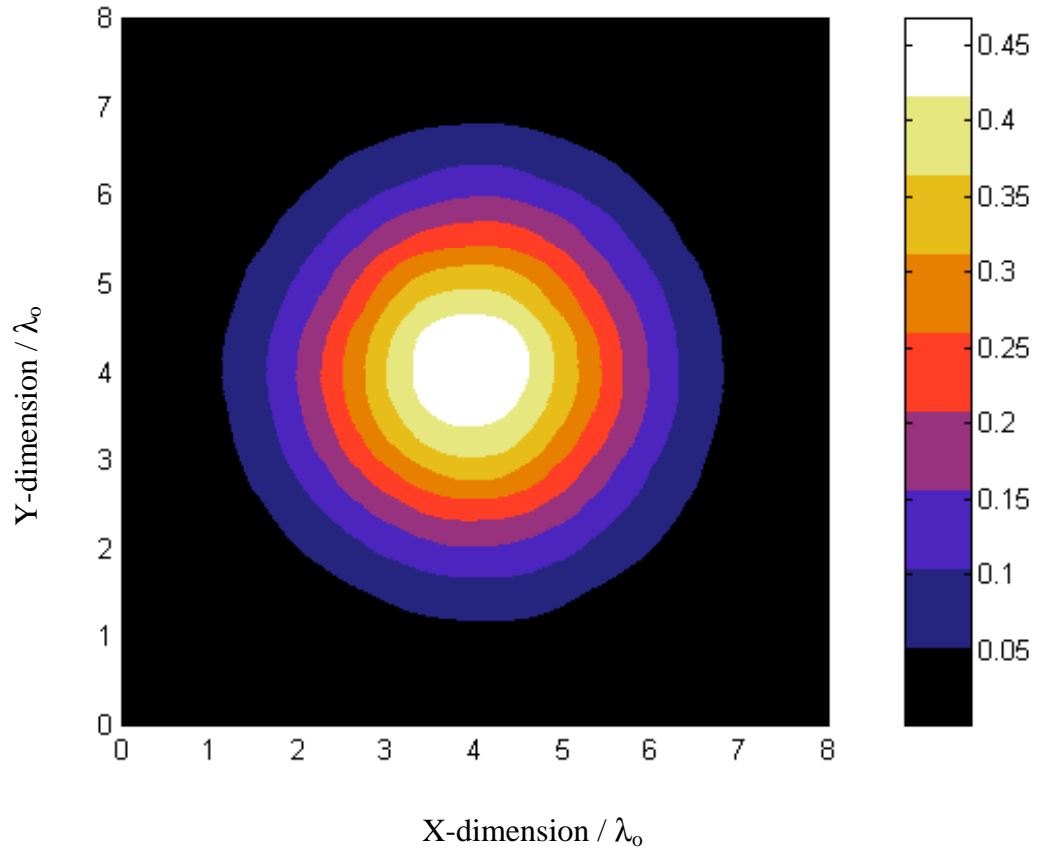


Figure 3d Near field transmitted at $z = -1\lambda_0$ below the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.1\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

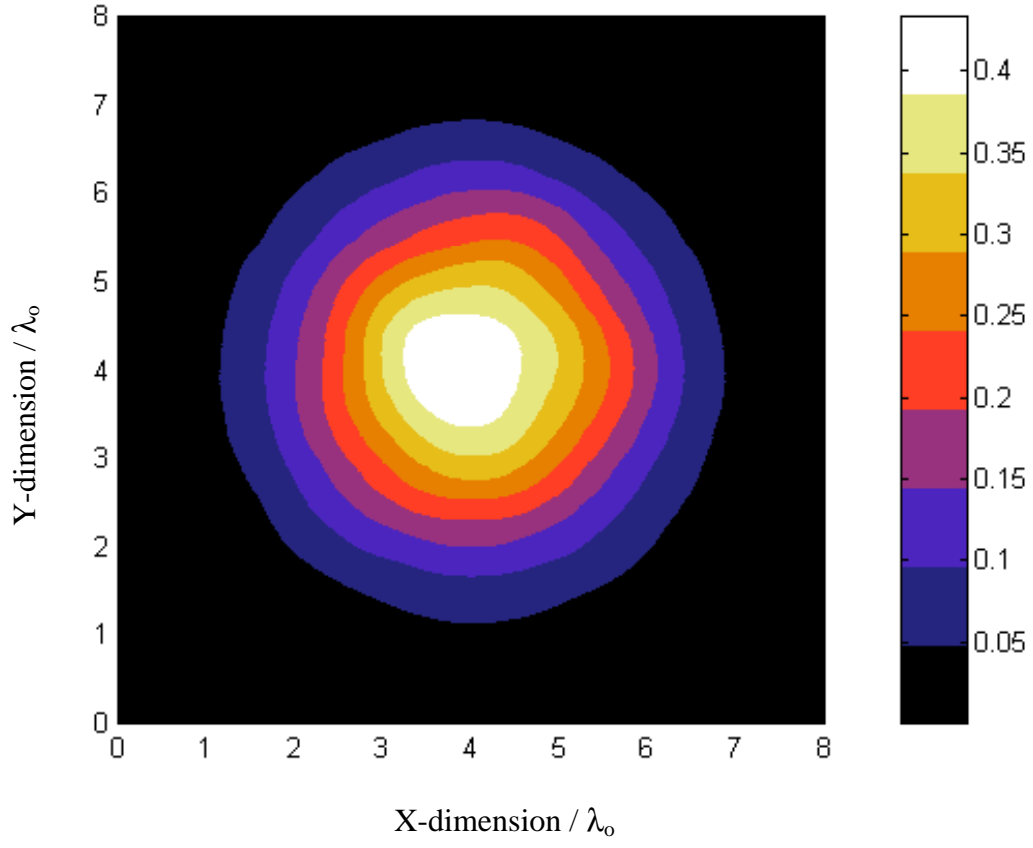


Figure 3e Near field transmitted at $z = -1\lambda_0$ below the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.12\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

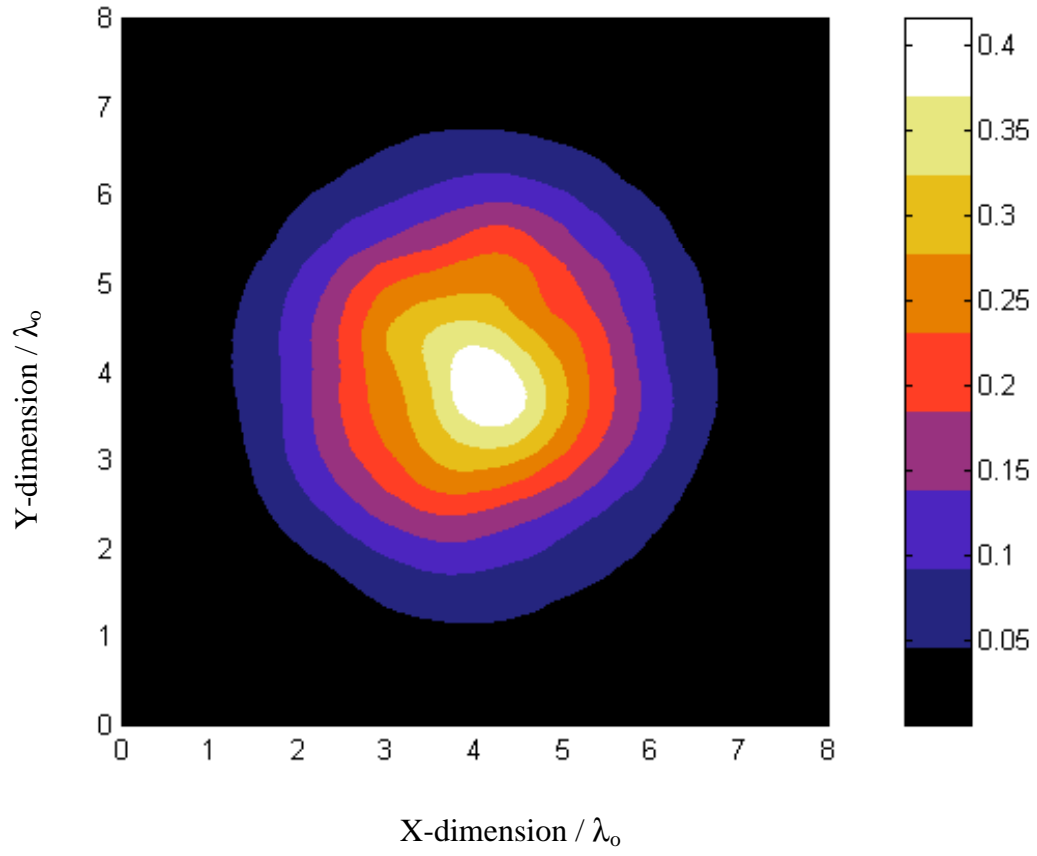


Figure 3f Near field transmitted at $z = -1\lambda_0$ below the mean ground of Bosnian soil rough surface of root mean square height $\sigma = 0.17\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$

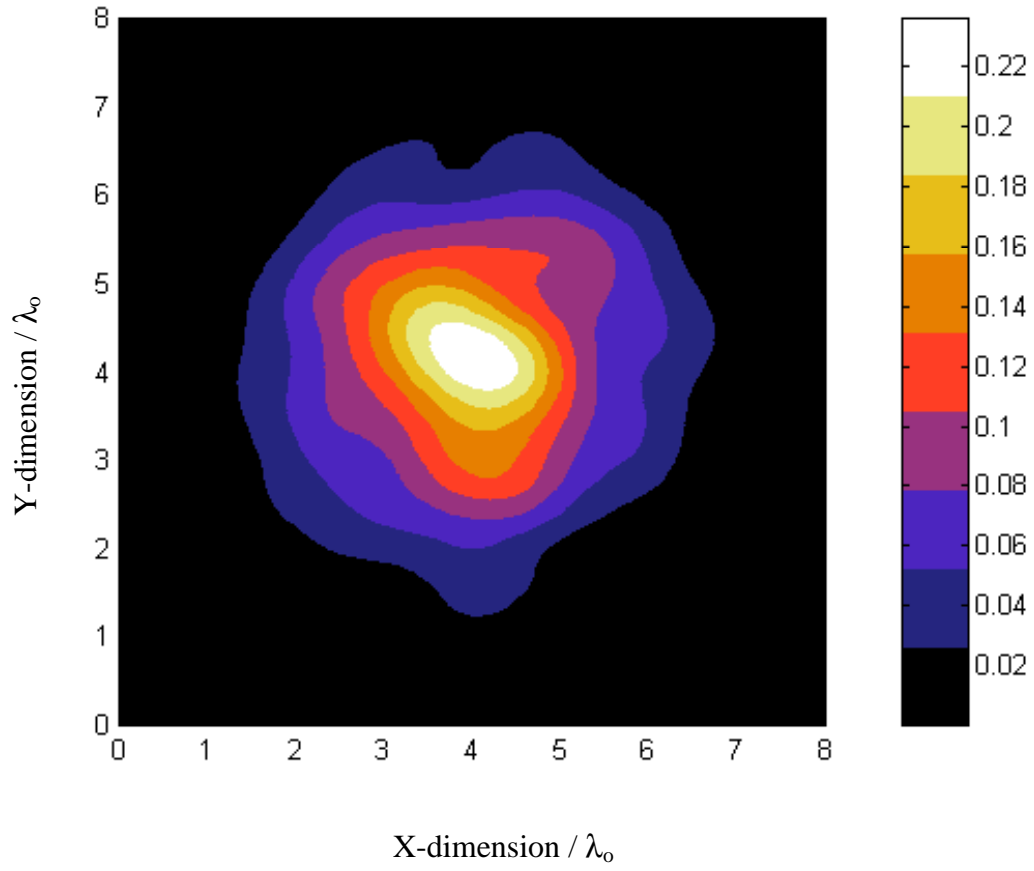


Figure 4a Near field scattered at $z = 1\lambda_0$ above the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.1\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

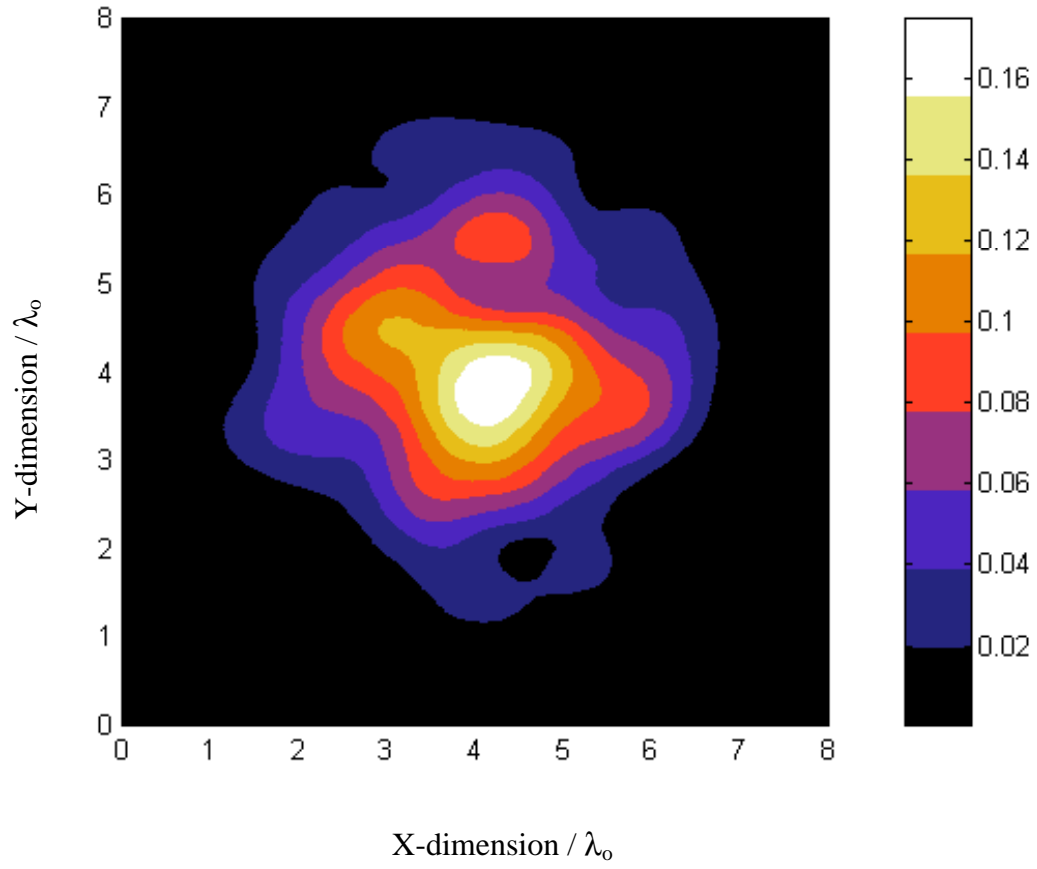


Figure 4b Near field scattered at $z = 1\lambda_0$ above the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.12\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

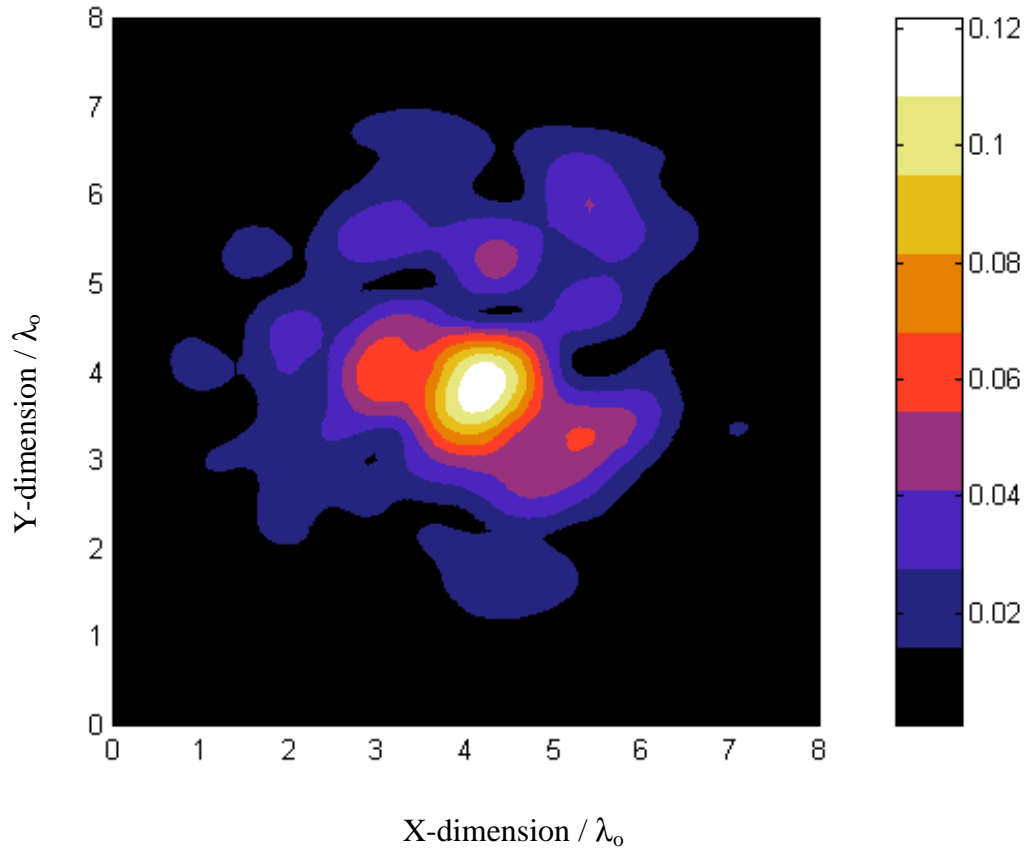


Figure 4c Near field scattered at $z = 1\lambda_0$ above the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.17\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

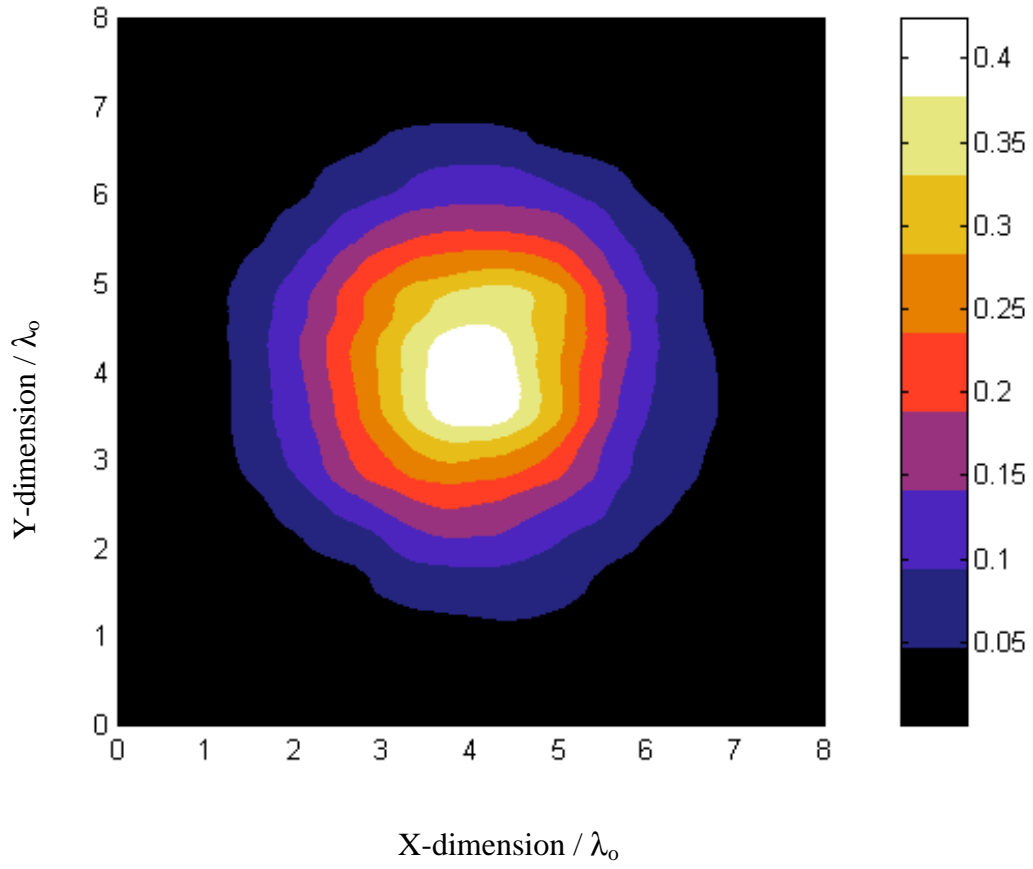


Figure 4d Near field transmitted at $z = -1\lambda_0$ below the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.1\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

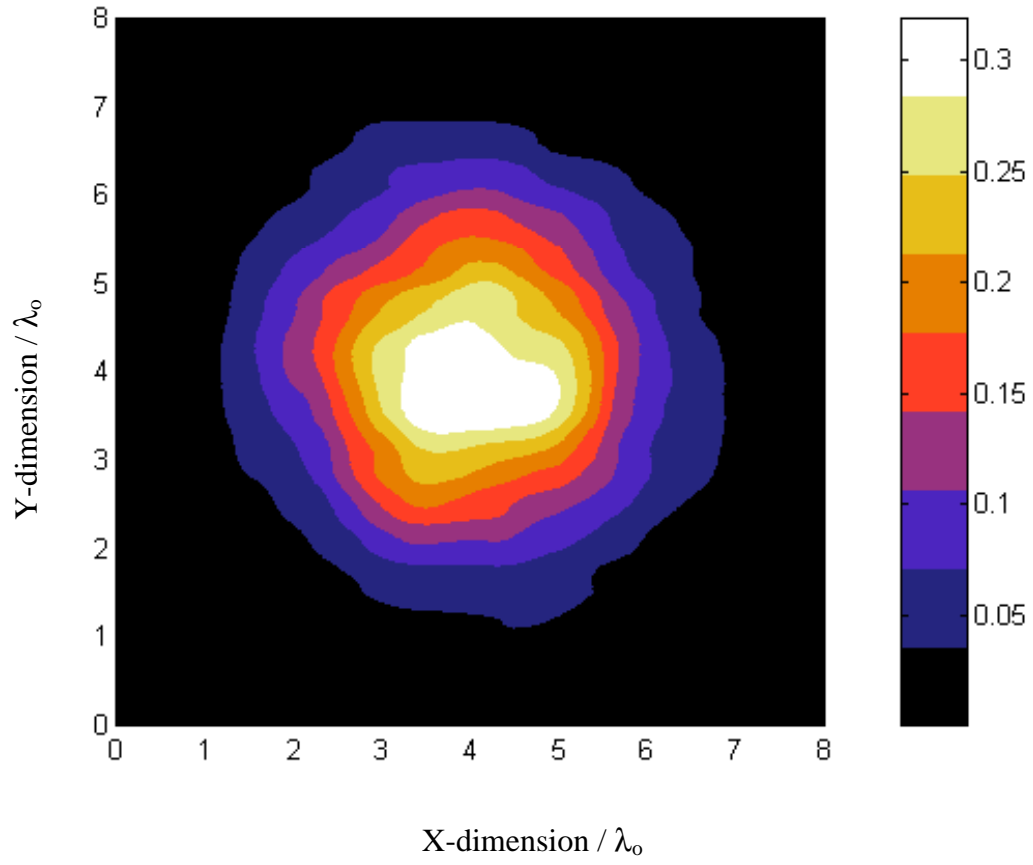


Figure 4e Near field transmitted at $z = -1\lambda_0$ below the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.12\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

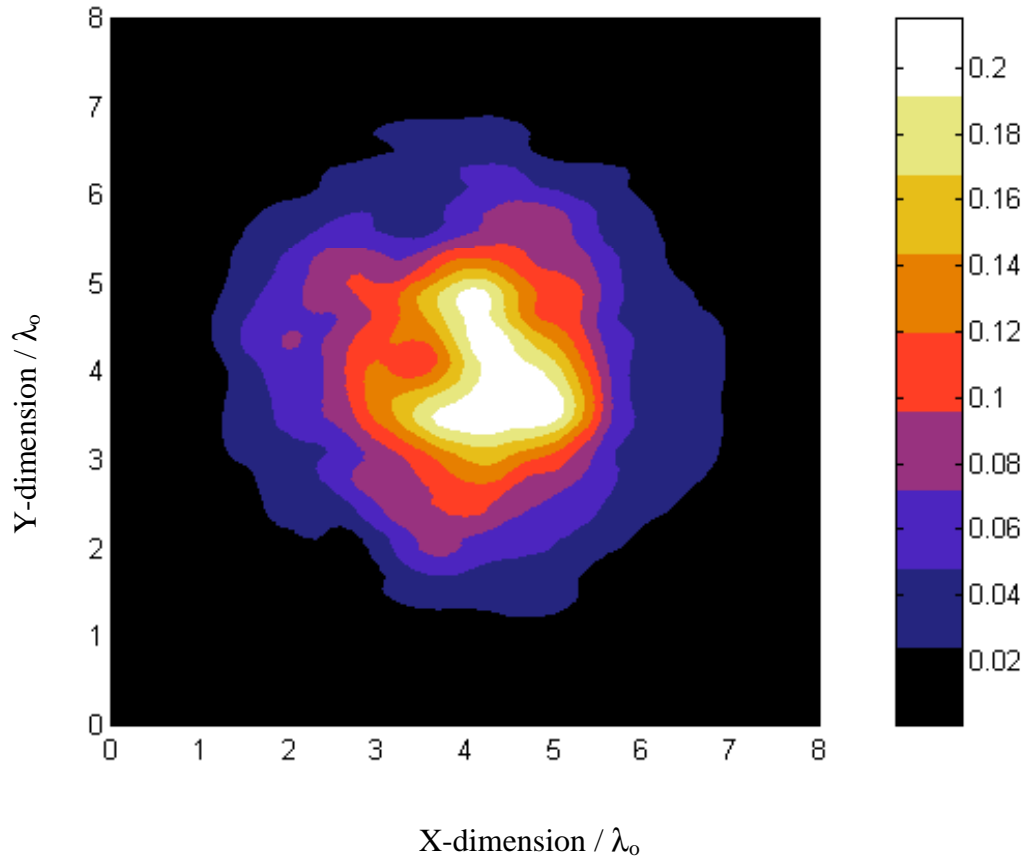


Figure 4f Near field transmitted at $z = -1\lambda_0$ below the mean ground of Puerto Rican soil rough surface of root mean square height $\sigma = 0.17\lambda_0$ and correlation length $l_c = 1\lambda_0$, and incident angle $\theta^i = 0$.

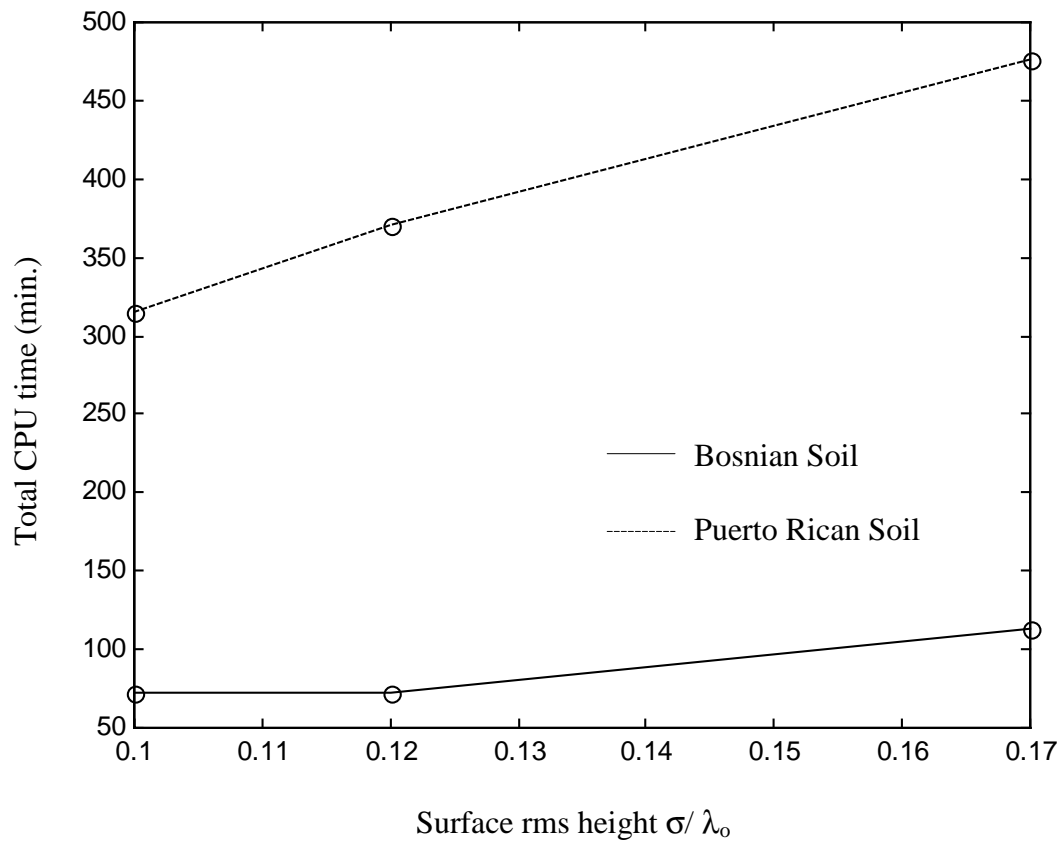


Figure 5 Total CPU time versus the rms height of the rough surface.

The Multiple Interaction Model for Non-Shallow Scatterers Buried Beneath Two-Dimensional Random Rough Surfaces

Magda El-Shenawee

Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701
Tel: 501-575-6582, Fax: 501-575-7967
magda@uark.edu

Abstract

The multiple interaction model is hybridized with the robust Steepest Descent Fast Multipole Method (SDFMM) to compute the signature of non-shallow penetrable scatterers buried beneath 2-D random rough surfaces. The most attractive feature of the multiple interaction model with using the SDFMM is removing the quasi-planar structure constraint for analyzing non-quasi-planar scatterers. The results show that the buried object's signature is largely due to the first interaction mechanism; however, the contribution of each additional interaction is explicitly calculated, though they may become insignificant especially for lossy background soil.

I. INTRODUCTION

Modeling electromagnetic scattering from realistic three-dimensional subsurface sensing applications requires huge number of computational operations that necessitates the use of fast algorithms [1]-[6]. Recently, the SDFMM [3] has been adopted to analyze the scattering from penetrable shallow objects buried beneath two-dimensional random rough ground [7],[8]. The SDFMM has the great advantage of $O(K)$ computational complexity for both the CPU time and computer memory [3] compared with the $O(K^2)$ for the MoM, where K is the total number of electric and magnetic surface current unknowns. However, there is a barrier that limits using the SDFMM in some applications; the whole scatterer should have a quasi-planar structure with total height equal to a fraction of a free-space wavelength. On the other hand, there are several

potential applications that have non-quasi-planar geometries; e.g., the multi-layered rough ground where the burial depth of the underground layer is larger than the wavelength.

The multiple interaction approach was previously used in investigating electromagnetic scattering problems either analytically from rough surfaces only (no buried objects), e.g. [9] and [10], or computationally from rough surfaces with buried objects, e.g. [11], [12]-[14]. In [9] and [10], each scattering element on the rough surface was assumed to be a second source for all scattering elements either on the same surface leading to the double scattering mechanism [9], or on the underground rough-layer as presented in [10]. On the other hand, in [11] a PEC sphere was buried beneath the rough ground while in [12],[13] a penetrable object was buried under or located above the rough ground. In addition, the multiple interaction approach is used with the Multilevel Fast Multipole Algorithm (MLFMA) for multiple targets located above the half space is presented in [14]. In these models, the coupling between the rough ground and the object is expressed as interaction matrices.

The multiple interaction approach with using the SDFMM is presented in this work. The basic idea is to decompose the 3-D non-quasi-planar structure (e.g., the rough ground with a buried scatterer) into two quasi-planar scatterers as shown in Fig. 1 where the SDFMM can be used separately for each scatterer. However, the interactions between the ground and the buried scatterer are not expressed in matrices (or submatrices) but are calculated concurrently using pre-computed and pre-stored information that facilitated the numerical evaluating of the near-field expressions given in [15]. In general, using the multiple interaction approach (with and/or without the SDFMM) enables investigating the contribution of each individual interaction mechanism between the ground and the buried object. This will greatly help in understanding the physics involved in subsurface sensing applications. It is necessary to distinguish between the multiple interaction model upon using the SDFMM, presented in this work, and our previous

complete SDFMM model, presented in [7] and [8]. All the self and mutual interactions between the rough ground and the buried object were expressed as submatrices in the total impedance matrix of the whole scatterer in [7]-[8].

II. FORMULATION

The four integral equations describing the unknown equivalent electric and magnetic surface currents for the problem of a single object buried beneath two-dimensional rough ground were derived and discussed in [7], [8]. The surfaces of the ground and the buried object were discretized using the RWG triangular patches [16]; and the set of linear system of equations was obtained [7]-[8]:

$$\begin{pmatrix} \bar{Z}_{g.g} & \bar{Z}_{g.obj} \\ \bar{Z}_{obj.g} & \bar{Z}_{obj.obj} \end{pmatrix} \begin{pmatrix} \bar{I}_g \\ \bar{I}_{obj} \end{pmatrix} = \begin{pmatrix} \bar{V}_g \\ 0 \end{pmatrix} \quad (1)$$

where $\bar{Z}_{g.g}$ represents the interactions between elements only on the ground surface; $\bar{Z}_{g.obj}$ represents interactions between elements on the ground surface and elements on the object surface; $\bar{Z}_{obj.g}$ represents interactions between elements on the object surface and elements on the ground surface; and $\bar{Z}_{obj.obj}$ represents interactions between elements only on the object surface. As discussed in [7], the total impedance matrix \bar{Z} has order of $2(N+P) \times 2(N+P)$, where N is the number of vector basis functions on the ground and P is the number of vector basis functions on the buried object. The factor of two accounts for both the electric and magnetic surface currents. The vector \bar{V}_g represents the tested tangential incident electric and magnetic fields on the ground surface. The unknown coefficients \bar{I}_g and \bar{I}_{obj} were solved for in [7] and [8] by completely applying the SDFMM to (1) leading to converting the dense matrix \bar{Z} into a sparse one. Conversely, in this work, the multiple interaction approach will be used to

solve two separate linear systems of equations for the unknown current coefficients and then iteratively update the incident fields on both the ground and the buried object as

$$\bar{Z}_{g.g} \bar{I}_g^{(n)} = \bar{V}_g^{(n)} \quad (2a)$$

$$\bar{Z}_{obj.obj} \bar{I}_{obj}^{(n)} = \bar{V}_{obj}^{(n)} \quad (2b)$$

where $n = 1, 2, 3, \dots$ is the number of the interactions between the ground and the buried object as depicted in Fig. 2. The algorithm begins with solving (2a) for $\bar{I}_g^{(0)}$, updating $\bar{V}_{obj}^{(1)}$ in (2b) by numerically evaluating the near-field surface integrations given by Eqs. 107-111 in Chapter 6 [15], solving (2b) for $\bar{I}_{obj}^{(1)}$, updating $\bar{V}_g^{(1)}$ in (2a), and finally solving (2a) for $\bar{I}_g^{(1)}$. These steps represent one interaction mechanism (i.e., ground-object-ground) that should be repeated until convergence of surface current solutions is achieved. The final updated electric and magnetic surface currents on the ground will be $\bar{J}_g = \bar{J}_g^{(0)} + \bar{J}_g^{(1)} + \bar{J}_g^{(2)} + \dots + \bar{J}_g^{(n)}$ and $\bar{M}_g = \bar{M}_g^{(0)} + \bar{M}_g^{(1)} + \bar{M}_g^{(2)} + \dots + \bar{M}_g^{(n)}$, respectively, and on the object $\bar{J}_{obj} = \bar{J}_{obj}^{(1)} + \bar{J}_{obj}^{(2)} + \bar{J}_{obj}^{(3)} + \dots + \bar{J}_{obj}^{(n)}$ and $\bar{M}_{obj} = \bar{M}_{obj}^{(1)} + \bar{M}_{obj}^{(1)} + \bar{M}_{obj}^{(2)} + \dots + \bar{M}_{obj}^{(n)}$, respectively. In order to accelerate the computations, the SDFMM can be used separately in (2a) and in (2b) to convert each dense impedance matrix $\bar{Z}_{g.g}$ and $\bar{Z}_{obj.obj}$ into sparse one leading to the multiple interaction model with the SDFMM. However, the efficient use of the SDFMM is contingent on the geometry of each scatterer separately, i.e., both the rough ground and the buried object should be quasi-planar structures. In this case, one interaction mechanism between the ground and the buried scatterer requires solving (2a) two times, solving (2b) only once, and numerically evaluating the near field expressions in [15] once for the object's and once for the ground's incident fields. This implies that SDFMM is not used in the near-field interaction computations given in [15]. Thus if K_1 and K_2 are the number of the RWG triangular patches on the rough

ground and on the object, respectively, then evaluating the near-field expressions in [15] using patch-patch interactions requires $2K_1K_2$ operations [16]. Moreover, all the information needed to evaluate these surface integrations are pre-calculated and pre-stored during the computations of $\bar{Z}_{g, g}$ and $\bar{Z}_{obj, obj}$. Thus the overall number of operations for the multiple interaction model with the SDFMM is proportional to $2N + 2N + 2P + 4K_1K_2 = (4N + 2P + 4K_1K_2)$. The quantity $(4K_1K_2)$ can be approximated by $(16NP/9 \approx 2NP)$ [16]. However, it is crucial to indicate that these $(2NP)$ operations are conducted only once in each interaction mechanism and they are not conducted in every iteration of the iterative solver like the $(4N+2P)$ operations. The Transpose Free Quasi Minimal Residual (TFQMR) iterative solver is used in this work [17]. The memory requirements in the multiple interaction model with the SDFMM is $O(2N+2P)$. On the other hand, the computational complexity of the complete SDFMM is $O(2N+2P)$ [3],[7],[8], the conventional MoM is of $O((2N+2P)^2)$, and the multiple interaction model is of $O(4N^2 + 4P^2)$ for both the CPU time and computer memory. In the case where the buried scatterer has a non-quasi-planar structure but instead has a small electrical size, the multiple interaction model can still be used by solving (2a) using the SDFMM while solving (2b) using the MoM (i.e. the multiple interaction model with partial use of the SDFMM). This scenario will slightly increase the total computational operations. However, if the non-quasi-planar buried scatterer has a very large electrical size, then using the MoM to solve (2b) will lead to a significant increase in the computational operations. In addition, the coupling $(2NP)$ operations per interaction will increase as well which limits the use of the multiple interaction model for this case.

III. NUMERICAL RESULTS

Basically we are interested in subsurface sensing problems where calculating the surface currents on the rough ground dominates the overall computations. In all results presented in this

section, the incident wave is assumed a Gaussian beam tapered towards the edges of the ground [18] with horizontally polarized incident electric field (i.e., in the y -direction). The half-beam width is $L/5$ where the ground has dimensions $L \times L$.

A variety of geometries represented by seven examples are used in this section. For Examples 1-5, the geometries are described in Fig. 2a-f where the ground media is lossy with relative dielectric constant assumed as $\epsilon_{2r} = 2.5 - j0.18$. However, to investigate the case of a shallow scatterer buried in a lossless medium, Example 6 uses the geometry in Fig. 2c but for larger dimensions of the ground given by $4.08\lambda_0 \times 4.08\lambda_0$ with $\epsilon_{2r} = 2.5$ while the underground layer is assumed to have $\epsilon_{3r} = 6.7 - j0.2$ and is buried at $z = -0.1\lambda_0$. The same geometry is also used in Example 7, but with even larger dimensions of the ground given by $8\lambda_0 \times 8\lambda_0$, same lossless soil of $\epsilon_{2r} = 2.5$, and lossy underground layer of $\epsilon_{3r} = 6.7 - j1.0$ buried at $z = -0.1\lambda_0$. The accuracy of the near-field expressions in [15] deteriorates for burial depths smaller than $0.1\lambda_0$ since the source and observation points become very close to each other.

A qualitative comparison (not presented here) showed excellent agreement between the surface currents obtained using the multiple interaction model with those obtained using the conventional MoM. Moreover, the convergence of the surface current solutions in Examples 1-7, is quantitatively demonstrated by plotting the normalized change in the currents $\|C^{(n)}\|/\|C^{(0)}\|$ versus the number of interactions between the ground and the buried scatterer. This is shown in Fig. 3. The vector C has dimensions of $(2N+2P)$ and it contains all the electric and magnetic surface currents on both the air/ground interface and the buried scatterer. The results show that a significant change occurs in the currents after one interaction ($n = 1$); however, much less significant change in the currents is observed after two interactions ($n = 2$), as shown for Examples 6 and 7, where the maximum relative change in the currents is less than 5%. However,

insignificant changes are observed for all cases at the third or higher interaction. This indicates that the contributions from the higher interactions depend on the physical properties of the scatterer, and in some cases, larger changes in the currents may occur at the second interaction. In Fig. 4, the error in the norm of the surface currents obtained using the multiple interaction model $\|C\|_{\text{Model}}$ and those obtained using the conventional MoM ($\|C\|_{\text{MoM}}$) are plotted versus the number of interactions for Examples 1-4. The results clearly validate the multiple interaction model.

In Fig. 5a, the scattered electric fields received above the ground at $z = 0.5\lambda_0$ and due just to the buried sphere are plotted versus the x -direction for $y = 2.04\lambda_0$ using the data of Example 4 (Fig. 2e). Three solutions are obtained for this example (i) the conventional MoM for the whole scatterer, (ii) the multiple interaction model (MIM), and (iii) the multiple interaction model with the SDFMM (MIM with SDFMM). The observed slight differences are attributed to the difference in defining the scattered fields due to just the buried sphere in both the conventional MoM and the multiple interaction model. Using the conventional MoM, we calculated the total scattered electric fields twice; with and without the buried sphere, then the results are subtracted from each other with complex vectors [7],[8]. On the other hand, using the multiple interaction model we use the obtained surface currents on the ground due to only the presence of the sphere, i.e., $\bar{J}_g^{(1)} + \bar{J}_g^{(2)} + \bar{J}_g^{(3)}$ and $\bar{M}_g^{(1)} + \bar{M}_g^{(2)} + \bar{M}_g^{(3)}$, and incorporate them into the near field expressions in [15] to compute the scattered electric fields above the ground. Notice that the quantities $\bar{J}_g^{(0)}$ and $\bar{M}_g^{(0)}$, the surface currents on the ground without a buried object, are not used in the calculations of the object signature. In Fig. 5b, the scattered electric fields received above the ground at $z = 0.5\lambda_0$ are plotted versus the y -direction for $x = 2.04\lambda_0$ using the data of Example 5 (Fig. 2f). The scattered fields due to just the air/ground rough interface are calculated using only the

ground surface currents without the presence of the buried layer, i.e., $\bar{J}_g^{(0)}$ and $\bar{M}_g^{(0)}$, while the scattered fields due to just the underground rough layer are calculated using only the ground surface currents $\bar{J}_g^{(1)} + \bar{J}_g^{(2)} + \bar{J}_g^{(3)}$ and $\bar{M}_g^{(1)} + \bar{M}_g^{(2)} + \bar{M}_g^{(3)}$. As expected, the signature of the air/ground rough interface is significantly larger than that of the underground rough layer. Table 1 summarizes all examples considered in this section obtained using the conventional MoM, the MIM, and the MIM with the SDFMM. All these three models are calculating the surface currents on the air/ground interface and on the buried object. The overall computer memory and CPU time are given for one interaction mechanism (i.e., ground-object-ground). In Example 7, the total CPU time is 192 min — 177 min for the TFQMR iterative solver and only 15 min for updating the incident fields on both the ground and the underground layer (interactions). This represents less than 8% of the total CPU time.

Table 1. CPU time and computer memory requirements for Examples 1-5.

Examples	# Unknowns		# Patches		CPU (min.)			Memory (MB)		
	$2N$	$2P$	K_1	K_2	MoM (conv.)	MIM	MIM with SDFMM	MoM (conv.)	MIM	MIM with SDFMM
1 (Fig. 2a)	8512	8352	2888	2808	118	49	†	1280	850	†
2 (Fig. 2c)	8512	8512	2888	2888	72	37	†	1305	880	†
3 (Fig. 2d)	8512	8512	2888	2888	78	38	†	1305	880	†
4 (Fig. 2e)	15402	2292	5202	764	93	66	38	1848	1467	225
5 (Fig. 2f)	15402	15402	5202	5202	‡	130	45	>6000	3256	395
7 (Fig. 2c)	59600	59600	20000	20000	‡	‡	192	‡	‡	1818

† The emphasis is to validate the MIM with the MoM for these small cases.

‡ The MoM could not be used for this case due to the large memory requirements (> 6 GB).

In all results presented in this section, a relative residual error of 10^{-5} is used in the TFQMR iterative solver [17] and the smallest FMM block size is assumed to be $0.32\lambda_0 \times 0.32\lambda_0$. In Examples 1-4, upon comparing the use of the MIM with the conventional MoM, the reductions in the total CPU time and computer memory, range from 30% to 60% and from 20% to 35%, respectively. On the other hand, the CPU time and computer memory are reduced by almost 65%

and 88%, respectively, when the MIM with the SDFMM is used versus the MIM as shown in Example 5. Moreover, in Example 7, only the MIM with the SDFMM could be used due to the huge required computer memory ($>6\text{GB}$) for both the MoM and the MIM.

It is necessary to emphasize that the most attractive feature of the MIM with the SDFMM is removing the quasi-planar structure constraint to enable its use in the non-quasi-planar applications described in Fig. 2. Moreover, there are no overlapping applications that require choosing between the MIM with the SDFMM which works more efficiently for non-shallow scatterers and the complete SDFMM in [7],[8] which works more efficiently for shallow scatterers. However, to obtain the signature of the buried scatterer, the complete SDFMM in [7],[8] requires executing the computer code twice, with and without the buried object, while the MIM with the SDFMM computes the object signature by running the computer code only once. Even though no convergence problems were encountered in any of the tested cases, even when a very shallow underground layer is buried in lossless soil, however for shallow scatterers, the complete SDFMM in [7],[8] should be used since it represents a potentially non-stationary algorithm for the whole matrix as discussed in [19] in addition to its superior computational complexity when used for the whole scatterer.

IV. CONCLUSIONS

The multiple interaction approach is used with the robust SDFMM to remove the quasi-planar structure constraint to enable analyzing non-shallow objects buried beneath the 2-D random rough ground. The results show that the first interaction mechanism between the ground and the buried scatterer significantly contributes to the surface current solutions while the contributions of additional interactions become insignificant especially when the non-shallow object is buried in lossy background soil.

ACKNOWLEDGMENTS

This research was sponsored by the College of Engineering at the University of Arkansas.

The SDFMM algorithm was originally developed at the UIUC by V. Jandhyala, E. Michielssen and W. Chew.

References

- [1] V. Rokhlin, "Rapid solution of integral equations of scattering theory in two dimensions," *J. Comput. Phys.*, vol. 36, pp. 414-439, 1990.
- [2] C. C. Lu and W. C. Chew, "A multilevel fast-algorithm for solving a boundary integral equation of wave scattering," *Microwave Opt. Tech. Let.*, vol. 7, pp. 466-470, July 1994.
- [3] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [4] S. Li, C. H. Chan, L. Tsang, Q. Li and L. Zhou, "Parallel implementation of the sparse matrix/canonical grid method for the analysis of two-dimensional random rough surfaces (three-dimensional scattering problem) on a Beowulf system," *IEEE Trans. Geosci. & Remote Sensing*, vol. 38, no. 4, pp. 1600-1608, July 2000.
- [5] D. Torrungrueng, H. Chou and J. T. Johnson, "A novel acceleration algorithm for the computation of scattering from two-dimensional large scale perfectly conducting random rough surfaces with the forward-backward method," *IEEE Trans. Geosci. & Remote Sensing*, vol. 38, no. 7, pp. 1656-1666, July 2000.
- [6] N. Geng, A. Sullivan and L. Carin, "Multilevel fast-multipole algorithm for scattering from conducting targets above or embedded in a lossy half space," *IEEE Trans. Geosci. & Remote Sensing*, vol. 38, no. 4, pp. 1561-1573, July 2000.
- [7] M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, "3-D subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces: Use of the steepest descent fast multipole method (SDFMM)," *IEEE Trans. Geosci. & Remote Sensing*, vol. 39, no. 6, pp. 1174-1182, June 2001.
- [8] M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo simulations of electromagnetic wave scattering from random rough surface with 3-D penetrable buried object: Mine detection application using the SDFMM," *J. Opt. Soc. Am. A*, to appear in December 2001.
- [9] E. Bahar and M. El-Shenawee, "Full Wave Single and Multiple Scattering from Rough Surfaces," *J. Comput. Phys.*, vol. 115, pp. 390-398, December 1994.
- [10] E. Bahar and Y. Zhang, "Diffuse like and cross-polarized fields scattered from irregular layered structures-full wave analysis," *IEEE Trans. Anten. & Propag.*, vol. 47, no. 5, pp. 941-948, 1999.
- [11] G. Zhang, L. Tsang and K. Pak, "Angular correlation function and scattering coefficient of electromagnetic waves scattered by a buried object under a two-dimensional rough surface," *J. Opt. Soc. Am. A*, vol. 15, no. 12, pp. 2995-3002, December 1998.
- [12] J. T. Johnson, "A study of the four-path model for scattering from an object above a halfspace," *Microwave Opt. Tech. Letters*, July 2001.

- [13] J. T. Johnson and R. J. Burkholder, "Coupled canonical grid/discrete dipole approach for computing scattering from objects above or below a rough interface," *IEEE Trans. Geosci. & Remote Sensing*, vol. 39, no. 6, pp. 1214-1220, June 2001.
- [14] Ling Li, Jiangqi, Zhijun Liu and Lawrence Carin, "MLFMA analysis of scattering from multiple targets in the presence of a half space," submitted in 2001 to the *IEEE Trans. on Anten. & Propag*
- [15] C. A. Balanis, *Advanced engineering electromagnetics*, John Wiley & Sons Inc., Ch. 6, pp. 254-309, 1989.
- [16] S. M. Rao, D. R. Wilton and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Anten. & Propag.*, vol. AP-30, no. 3, pp. 409-418, May 1982.
- [17] R. W. Freund, "A Transpose-free quasi-minimal residual algorithm for non-hermitian linear systems," *SIAM J. Sci. Comput.*, vol. 14, no. 2, pp. 470-482, March 1993.
- [18] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.
- [19] James C. West and J. Michael Sturm, "On iterative approaches for electromagnetic rough-surface scattering problems," *IEEE Trans. on Anten. & Propag.*, vol. AP-47, no. 8, pp. 1281-1288, August 1999.

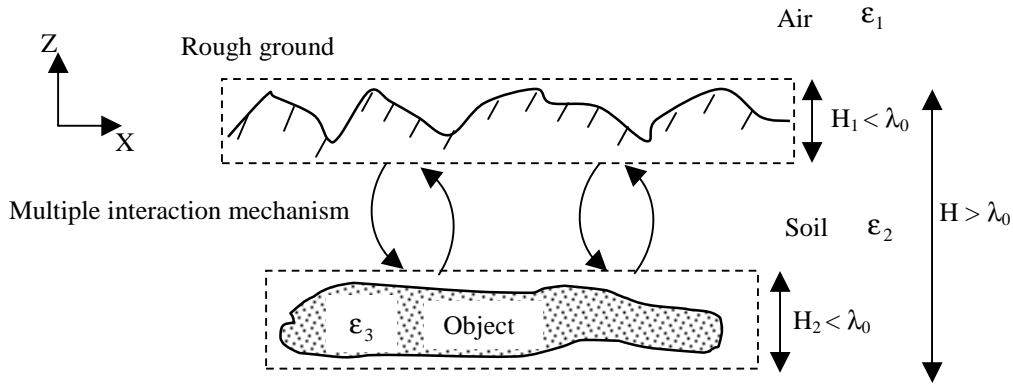
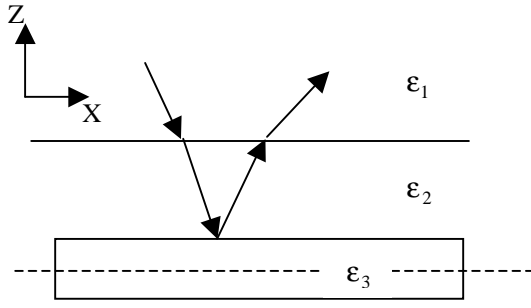
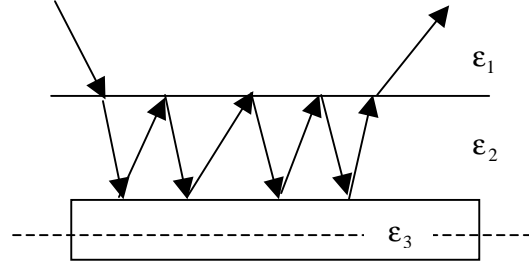


Figure 1. Non-quasi-planar structure decomposition into two quasi-planar structures showing their multiple interactions.



Buried cylinder-one interaction

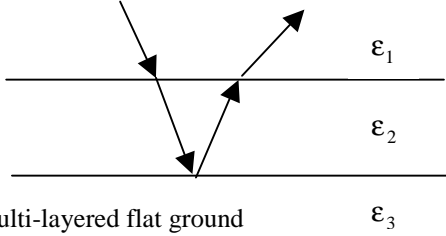
(a)



Buried cylinder-multiple interactions

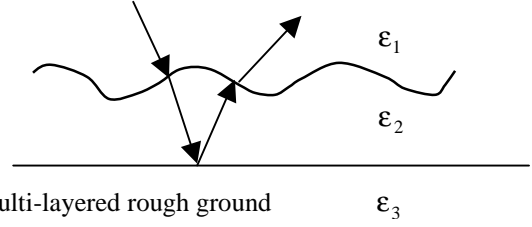
(b)

(a) a cylinder of length $b=3.04\lambda_0$, radius $a=0.15\lambda_0$ and $\epsilon_{3r}=7.9-j0.0029$ buried at $z=-0.65\lambda_0$ under a flat ground of dimensions $3.04\lambda_0 \times 3.04\lambda_0$ with $\epsilon_{2r}=2.5-j0.18$ for $\vartheta^i=0^\circ$ (Example 1), showing one ground-object-ground mechanism $n=1$, (b) the same data of (a) showing four ground-object-ground mechanisms $n=4$.



Multi-layered flat ground

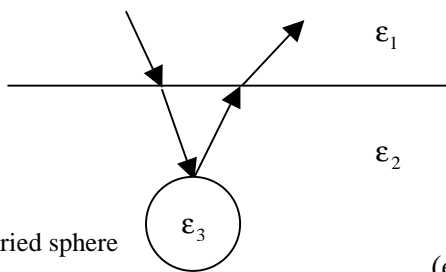
(c)



Multi-layered rough ground

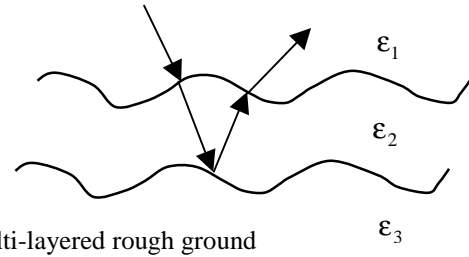
(d)

(c) the same flat air/ground interface of (a) and a flat under ground layer buried at $z=-0.45\lambda_0$ with $\epsilon_{3r}=4.2-j0.29$ for $\vartheta^i=10^\circ$ (Example 2), (d) the same data of (c) but with rough air/ground interface with rms height $\sigma=0.08\lambda_0$ and correlation length $l_c=0.5\lambda_0$ (Example 3).



Buried sphere

(e)



Multi-layered rough ground

(f)

(e) a sphere of radius of $a=0.5\lambda_0$ with $\epsilon_{3r}=4.5-j0.029$ buried at $z=-0.75\lambda_0$ (measured from its center) under a flat air/ground interface of dimensions $4.08\lambda_0 \times 4.08\lambda_0$ with $\epsilon_{2r}=2.5-j0.18$ for $\vartheta^i=0^\circ$ (Example 4), (f) the same data of (e) but with rough air/ground ($\sigma_1=0.06\lambda_0$, $l_{c1}=0.5\lambda_0$) and underground interface ($\sigma_2=0.05\lambda_0$, $l_{c2}=0.4\lambda_0$) buried at $z=-0.95\lambda_0$ (measured between the two mean planes) with $\epsilon_{3r}=3.7-j0.2$ (Example 5).

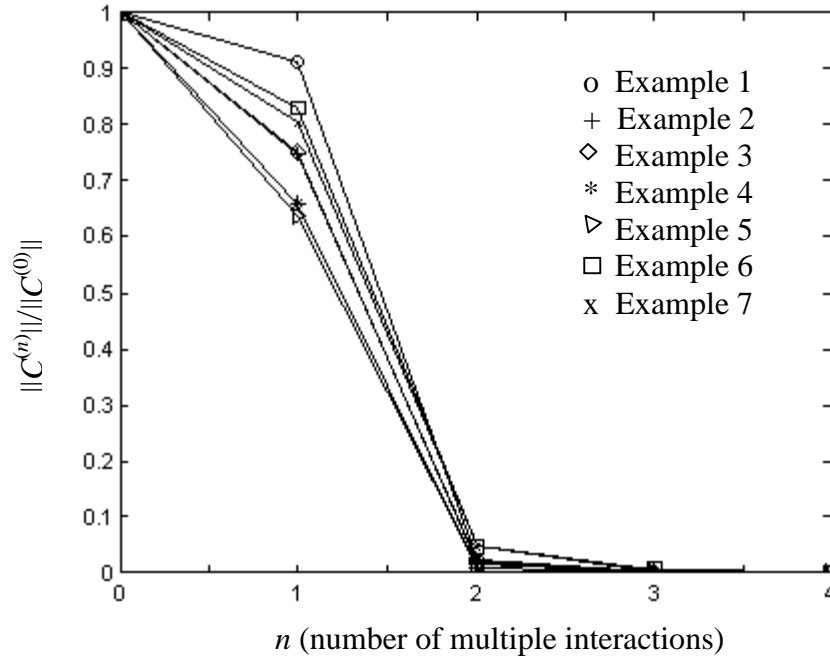


Figure 3. The normalized change in the surface currents $\|C^{(n)}\|/\|C^{(0)}\|$ versus the number of interactions for Examples 1-7, where the vector C contains all electric and magnetic surface currents.

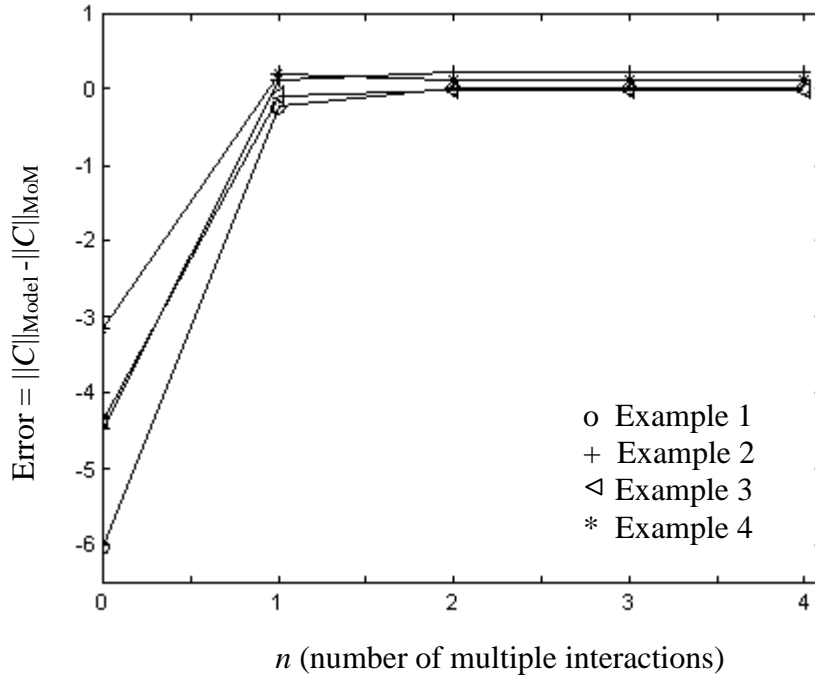


Figure 4. The error in the norm of the surface currents obtained using the multiple interaction model $\|C\|_{\text{Model}}$ and those obtained using the conventional MoM ($\|C\|_{\text{MoM}}$) versus the number of interactions for Examples 1-4, where the vector C contains all electric and magnetic surface currents.

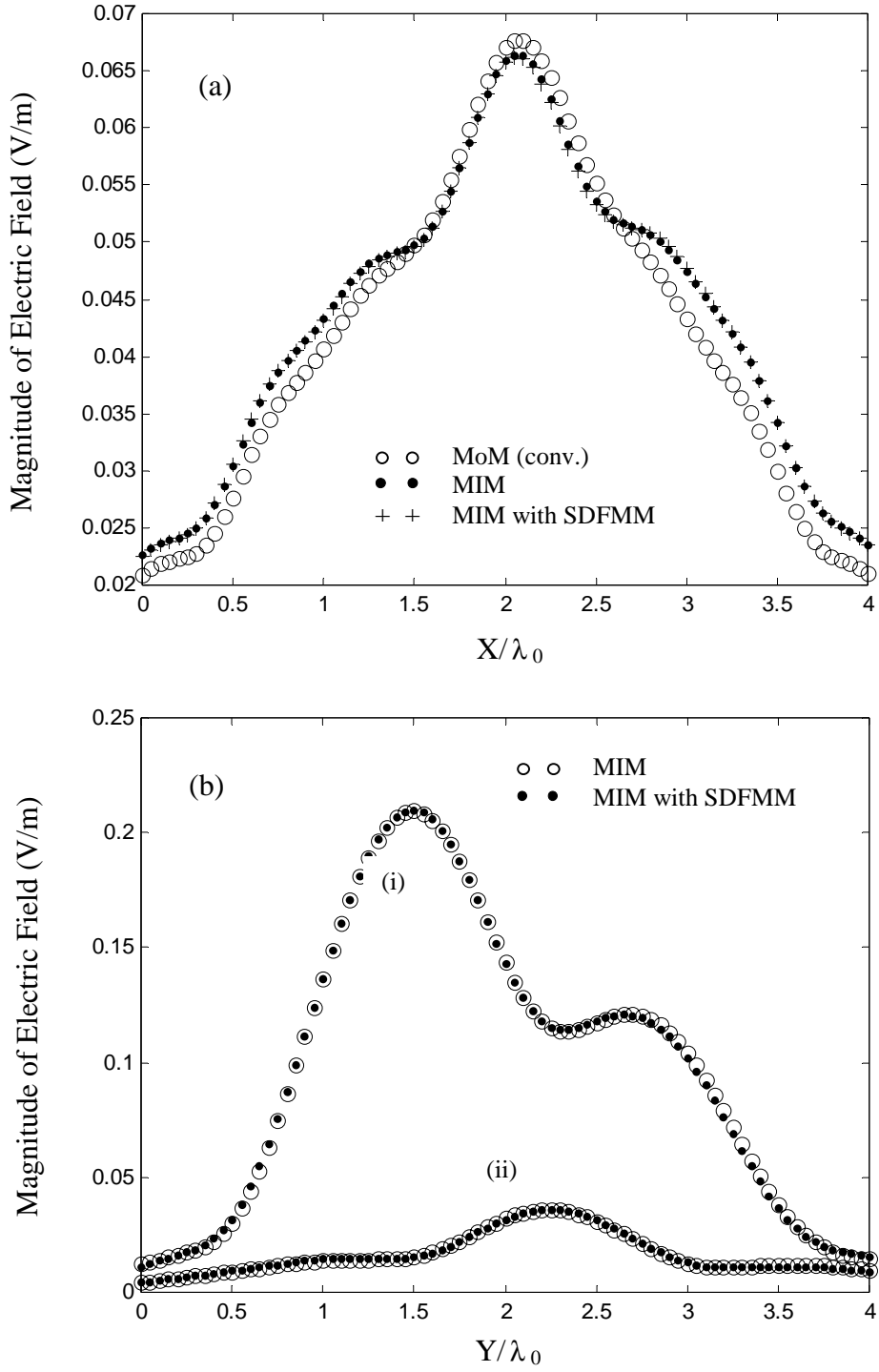


Figure 5. Magnitude of scattered electric field $|\bar{E}|$ received at $z = 0.5\lambda_0$ above the ground due to (a) just the buried sphere in Example 4 (Fig. 2e shown at $y = 2.04\lambda_0$) and (b) due to (i) just the rough air/ground interface and (ii) just the rough underground layer, in Example 5 (Fig. 2f. shown at $x = 2.04\lambda_0$).

Monte Carlo Simulations of Electromagnetic Wave Scattering from Random Rough Surface with 3-D Penetrable Buried Object: Mine Detection Application Using the SDFMM.

M. El-Shenawee¹, C. Rappaport² and M. Silevitch²

¹Department of Electrical Engineering

University of Arkansas

Fayetteville, Arkansas 72701

Tel: (501)-575-6582, Fax: (501)-575-7967

magda@uark.edu

²Center for Subsurface Sensing and Imaging Systems (CenSSIS)

Department of Electrical and Computer Engineering

Northeastern University, 235 Forsyth Bld

Boston, MA 02115

Tel: 617-373-2064, Fax: 617-373-8627

rappaport@neu.edu

ABSTRACT

In this work, we present a statistical study of the electric field scattered from a 3-D penetrable object buried under a 2-D random rough surface. Monte Carlo simulations using the Steepest Descent Fast Multipole Method (SDFMM) are conducted to calculate the average and the standard deviation of the near zone scattered fields. The SDFMM, originally developed at UIUC has been modified to calculate the unknown surface currents both on the rough ground and on the buried object due to excitation by a tapered Gaussian beam. The rough ground medium used in this study is an experimentally measured typical dry Bosnian soil with 3.8% moisture, while the buried object represents a plastic land mine modeled as an oblate spheroid with dimensions

and burial depth smaller than the free space wavelength. Both vertical and horizontal polarizations for the incident waves are studied. The numerical results show that the TNT mine signature is almost 5% of the total field scattered from the ground. Moreover, relatively recognizable object signatures are observed even when the object is buried under the tail of the incident beam. Interestingly, even for the small surface roughness parameters considered here, the standard deviation of the object signature is almost 30% of the signal itself, indicating significant clutter distortion due to the roughness of the ground.

I. INTRODUCTION

For subsurface sensing applications such as indicated in Fig. 1, electromagnetic wave scattering from rough ground is a major source of clutter for the measurements. Analyzing and characterizing this clutter is vital in the development of more efficient radar sensing systems. This random scattering of the electromagnetic fields necessitates the use of Monte Carlo simulations and makes it a key stochastic process to obtain the statistics of the scattered fields, and hence to extract information and draw meaningful conclusions from the numerical results. There is much published work on electromagnetic scattering from one- or two-dimensional random rough surfaces only (without buried objects) using Monte Carlo simulations, e.g., [1]-[5]. Moreover, there are several publications that model electromagnetic waves from objects buried under one-dimensional random rough surfaces, e.g., [6]-[8], or objects buried under flat half-spaces [9]. However, there are very few published works for an object buried under a two-dimensional random rough surface; [10], which attempts to model a perfect electrically conducting (PEC) buried object and, to the author's knowledge, only [11] attempts to model the case of a penetrable buried object.

The computational expenses, the CPU time and computer memory required to calculate the scattering fields from 3-D penetrable scatterers become excessive for either an electrically large scatterer and/or a scatterer with large dielectric constant. In addition, Monte Carlo simulations require that the calculations must be repeated for many rough surface realizations (depending on the ground roughness) until results converge. As it is known, solving a 3-D scattering problem is expensive even for one surface realization, so conducting Monte Carlo simulations with hundreds of realizations without using fast algorithms is prohibitive even using supercomputers. Implementing the fast multipole method (FMM) [12]-[16] makes the solution of a 3-D problem tractable with significant reduction in CPU time and computer memory. The integral equation-based Steepest Descent Fast Multipole Method (SDFMM), a hybridization of the Moment Method (MOM), the Fast Multipole Method (FMM) and the Steepest Descent Path integration rule (SDP), is used to calculate the unknown surface currents. The details of the SDFMM are published in [17]-[20].

In this work, we are using the same formulations we developed in [11], with primary emphasis on conducting efficient and fast Monte Carlo simulations to obtain the statistics of the scattered near fields. In Section II, the problem formulations are briefly summarized. Numerical results using the SDFMM are presented in Section III. Conclusions and future work are stated in Section IV.

II. PROBLEM FORMULATION

The SDFMM makes use of the equivalence theorem to calculate the electric and magnetic fields inside and outside a 3-D penetrable object buried under the rough surface interface, details are

presented in [11]. The 3-D arbitrary object is modeled by scatterer R_3 that is immersed in scatterer R_2 which represents the rough ground. The scatterer R_2 is immersed in the free space region represented by R_1 , as shown in Fig. 2. The three regions, R_1 , R_2 and R_3 have permittivity and permeability given by ϵ_1 and μ_1 , ϵ_2 and μ_2 , and ϵ_3 and μ_3 , respectively, representing free space, soil medium and penetrable buried object. There are two final sets of unknown equivalent electric and magnetic surface currents in the following formulations. They are $\mathbf{J}_1, \mathbf{M}_1$, on the exterior of S_1 , interface between R_1 and R_2 , and $\mathbf{J}_3, \mathbf{M}_3$ on the exterior of S_2 , interface between R_2 and R_3 . Upon applying the boundary conditions, continuity of tangential components of the electric and magnetic fields on S_1 and S_2 , new integral equation formulations are obtained as [11]:

$$\mathbf{E}^{inc}(\mathbf{r})\big|_{\text{tang.}} = [(L_1 + L_2)\mathbf{J}_1 - (K_1 + K_2)\mathbf{M}_1 - L_3\mathbf{J}_3 + K_3\mathbf{M}_3]_{\text{tang.}} \quad (1)$$

$$\mathbf{H}^{inc}(\mathbf{r})\big|_{\text{tang.}} = \left[(K_1 + K_2)\mathbf{J}_1 + \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \mathbf{M}_1 - K_3\mathbf{J}_3 - \frac{L_3}{\eta_2^2} \mathbf{M}_3 \right]_{\text{tang.}} \quad (2)$$

$$0 = [-L_2\mathbf{J}_1 + K_2\mathbf{M}_1 + (L_3 + L_4)\mathbf{J}_3 - (K_3 + K_4)\mathbf{M}_3]_{\text{tang.}} \quad (3)$$

$$0 = \left[-K_2\mathbf{J}_1 - \frac{L_2\mathbf{M}_1}{\eta_2^2} + (K_3 + K_4)\mathbf{J}_3 + \left(\frac{L_3}{\eta_2^2} + \frac{L_4}{\eta_3^2} \right) \mathbf{M}_3 \right]_{\text{tang.}} \quad (4)$$

In which the integro-differential operators L_i and K_i , $i=1, 2, 3$ and 4 , are given in detailed in [11]. In Eqs. 1-4, the unknown surface electric and magnetic currents are \mathbf{J}_1 , \mathbf{M}_1 , \mathbf{J}_3 , and \mathbf{M}_3 , while the tangential component of the incident electric and magnetic fields on the rough surface are given by $\mathbf{E}^{inc}(\mathbf{r})\big|_{\text{tang.}}$ and $\mathbf{H}^{inc}(\mathbf{r})\big|_{\text{tang.}}$, respectively. The intrinsic impedance in each region is

$\eta_i = \sqrt{\mu_i / \epsilon_i}$, $i=1, 2$, and 3 , where the dielectric permittivity and permeability in each region are ϵ_i and μ_i , respectively. Equations 1-4 are considered the extension of the PMCHW formulations that has been shown to yield a unique solution at internal resonances associated with the corresponding conducting scatterer [21]-[23]. The equivalent electric and magnetic currents are approximated using the RWG vector basis functions $\mathbf{j}(\mathbf{r})$ [21], [24] as:

$$\mathbf{J}_1(\mathbf{r}) = \sum_{n=1}^N I_{1n} \mathbf{j}_{1n}(\mathbf{r}), \quad \mathbf{M}_1(\mathbf{r}) = \eta_1 \sum_{n=1}^N I_{2n} \mathbf{j}_{1n}(\mathbf{r}), \quad \mathbf{r} \in S_1 \quad (5a)$$

$$\mathbf{J}_3(\mathbf{r}) = \sum_{m=1}^P I_{3m} \mathbf{j}_{2m}(\mathbf{r}), \quad \mathbf{M}_3(\mathbf{r}) = \eta_1 \sum_{m=1}^P I_{4m} \mathbf{j}_{2m}(\mathbf{r}), \quad \mathbf{r} \in S_2 \quad (5b)$$

where surfaces S_1 and S_2 are discretized into triangular patches with total number of edges (number of unknown coefficients) equal to N on S_1 and equal to P on S_2 . As shown in Eqs. (5a) and (5b), both the electric and magnetic surface currents has the same number of unknown coefficients (edges) on each surface. Upon applying Galerkin's method and substituting the above current approximations in Eqs. (1-4), the original integral equations are transformed into a set of linear system of equations given by:

$$\mathbf{Z}\mathbf{I} = \mathbf{V} \quad (6)$$

Notice that the matrix \mathbf{Z} has order of $2(N+P) \times 2(N+P)$, the vector \mathbf{V} is a matrix of order $2(N+P) \times 1$ and composed of a submatrix of the tested tangential incident electric field \mathbf{E}^{inc} of order $N \times 1$ and a submatrix of the tested normalized magnetic field $\eta_1 \mathbf{H}^{inc}$ of order $N \times 1$, and a null submatrix of order $2P \times 1$. Finally, the SDFMM is implemented in Eq. 6 reducing the

computational complexity for the CPU time and computer memory requirements from $O(K^2)$, for the MOM, to only $O(K)$ per iteration of iterative solver using the SDFMM [17]-[19], where $K=2(P+N)$ is the total number of the electric and magnetic surface current unknowns.

III. NUMERICAL RESULTS

The SDFMM code has been validated with the SMCG method (sparse canonical conjugate gradient method) [10] where the scatterer is a PEC sphere buried under a random rough surface and very good agreement is presented in [11]. More validations of the SDFMM code (not presented here) are conducted showing excellent agreement between the SDFMM code and both the Mie and MOM solutions for the case of just a penetrable sphere immersed in air (no rough surface) and excited by incident plane wave [21].

In all results presented in this section, the incident wave is assumed to be a Gaussian beam [25] located in the plane of incidence as shown in Fig.1. The half beam width of the Gaussian beam is assumed to be $L/5$, where the simulated rough ground has dimensions equal to $L \times L$, with $L = 8\lambda_0$ in this work. This leads to plane wave illumination of a spot size of diameter $3.2\lambda_0$ on the ground with tapered illumination towards the edges of the finite modeled ground. The rough surface is characterized with Gaussian statistics for the random heights and for the autocorrelation function. In this work, we are interested at the anti-personnel mine detection application which justifies the assumption of small roughness parameters of the ground given by the root mean square height $\sigma = 0.04\lambda_0$ and the correlation length $l_c = 0.5\lambda_0$. The dielectric constant of a typical Bosnian soil with 3.8% moisture is $\epsilon_r = 2.5 - i0.18$ at 1GHz [26], while the dielectric constant of the buried object (*TNT* material) is $\epsilon_r = 2.9 - i0.072$. The buried object is

modeled as an oblate spheroid with top view as a circle of radius $a=0.3\lambda_0$ and both front and side views are ellipses of dimensions $a=0.3\lambda_0$ and $b=0.15\lambda_0$. The burial depth of the object is $d = 0.3\lambda_0$ and is measured from its center to the mean plane of the rough surface. In this work the incident Gaussian beam is always pointing at the center of the rough ground with plane wave illumination spot of diameter equal to $3.2\lambda_0$ compared with $0.6\lambda_0$ diameter of the buried spheroid. Moreover, all results here represent near scattered electric fields and all subtraction processes are conducted with complex vectors.

In all results in this section, the discretization length of the surface current on the rough ground is assumed to be $0.08\lambda_0$ producing a total number of surface current unknowns as $2N = 59,600$ [11]. The number of nodes and patches on the buried penetrable oblate spheroid are 122 and 240, respectively, producing a total number of surface current unknowns as $2P=720$ [11]. The dimensions of the SDFMM finest block are assumed $0.32\lambda_0 \times 0.32\lambda_0$ with two blocks separating the near field (MOM) from the far field (SDFMM) interactions on the scatterer [17]-[19].

As mentioned earlier, we are primarily interested at scattered electric fields in the near zone, which can be calculated using the solved electric and magnetic surface currents [27] and are obtained at point receivers located $0.5\lambda_0$ above the nominal rough surface with resolution $0.1\lambda_0$, as shown in Fig. 1. The magnitude of the average scattered electric field and the standard deviation are calculated as:

$$Average = \left| \frac{1}{M} \sum_{i=1}^M \mathbf{E}_i \right| \quad (7a)$$

$$STD = \sqrt{\frac{1}{M} \sum_{i=1}^M |\mathbf{E}_i|^2 - \left| \frac{1}{M} \sum_{i=1}^M \mathbf{E}_i \right|^2} \quad (7b)$$

where, \mathbf{E}_i represents the complex vector electric field scattered from the i^{th} rough surface realization where $i=1, 2, \dots, M$, where the size of Monte Carlo set (M) is assumed to be 65 in all results in this section.

In Figs. 3-7, two incident elevation angles with one azimuth angle $\varphi^i = 0$, will be considered; $\vartheta^i = 0$ for the horizontally polarized incident waves where the electric field is in the y -direction (Figs. 3-6) and $\vartheta^i = 10^\circ$ for both polarizations (Figs. 7a-d).

In Figs. 3a and 3b, the magnitude of the average near scattered electric fields at $z = 0.5\lambda_0$, using Monte Carlo simulations, from the rough surface alone and from the rough surface with the buried spheroid, located at $x = y = 4.0\lambda_0$, are shown, respectively. Notice the slight difference between results in Fig. 3a and Fig. 3b, which is due to the small size of the buried object compared with the free space wavelength. Our objective is to analyze and extract information about the buried object that caused this slight difference. In Fig. 4a, the near scattered electric fields from the buried spheroid under two individual rough surface realizations selected from the 65 ones used in Fig. 3b, the average of the 65 surface realizations of rough surface with buried object (Fig. 3b), the average of the 65 surface realizations of rough surface alone (Fig. 3a), and the scattered electric field from flat ground with and without buried object are plotted at $Y = 4.0\lambda_0$. The results of Fig. 4a show the oscillations in the scattered electric fields from these individual rough surface realizations compared with the smooth curve obtained upon averaging the fields using Monte Carlo simulations. Notice that the average fields shown in Fig. 4a (solid

and dotted lines) are not perfectly symmetric around the point $X = 4.0\lambda_0$ compared with the symmetrical results of the flat ground. This can be attributed to the relatively small size of the Monte Carlo sample (65 realizations) and to the edge effect of the finite modeled rough ground. This qualitative comparison shown in Fig. 4a indicates the significant effect of the rough ground on the scattered signals even with the small roughness parameters chosen in this application. In order to obtain the scattered fields due just to the buried object, four types of subtractions are conducted with complex vectors and results are shown in Figs. 4b-e. First, we subtract the electric field scattered from flat ground alone from those scattered from flat ground with buried spheroid and results are shown in Fig. 4b. Second, we subtract the average scattered field from rough surface alone (solid curve in Fig. 4a) from the fields scattered from one rough surface realization with the buried object (surface #1 in Fig. 4a). The outcome is plotted in Fig. 4c, which shows very noisy results compared with Fig. 4b. Instead, the near electric fields scattered from only one surface realization is subtracted from those scattered from the same surface realization with buried spheroid, and results are plotted in Fig. 4d. Significant enhancement in results is shown in Fig. 4d compared with Fig. 4c giving more information about the object signature. Finally upon repeating the subtraction process of Fig. 4d for each of the 65 rough surface realizations, and then taking the statistical average of the outcome, results shown in Fig. 4e are obtained. Thus conducting Monte Carlo simulations of the 65 object signals obtained by subtraction, gives ideal image as clearly shown in Fig. 4e compared with Fig. 4b versus Figs. 4c and 4d. In practice signals scattered from a target buried under only one rough ground are measured. However the statistics shown in Fig. 4e are used to conclude that the *TNT* spheroid has signature almost equal to 5% of the total scattered fields. In reality, neither the profile of the rough ground or its soil dielectric constant is precisely known. However, if these parameters can

be estimated, then our fast model can be used to compute the scattered fields from the rough ground alone. As a result it can be used to remove the clutter from the measurement data by subtraction, as shown earlier, aiming to obtain results similar to Fig. 4d. Signal-processing techniques could be applied to this output rather than directly to the more noisy measurement data.

Upon keeping the incident Gaussian beam pointing at the center of the ground and changing the location of the buried spheroid from the center of the ground at $x = y = 4.0\lambda_0$ to $x = y = 5.0\lambda_0$ and to $x = 6.5\lambda_0, y = 4.0\lambda_0$, similar results to Fig. 4e are obtained at these locations. A qualitative comparison between the scattered electric fields from the buried spheroid versus the object location is shown in Fig. 5. As expected, these results show the degradation of object signature with respect to its closeness to the center of the incident Gaussian beam. Although the maximum excitation is not over the target, computing the difference fields clearly shows the target position. The standard deviation (*STD*) of the 65 scattered electric fields due just to the spheroid is calculated using Eq. 7b and plotted in Fig. 6 show that it is equal to almost 30% of the object signature shown in Fig. 4e. In Figs. 7a and 7b, the elevation angle of the incident Gaussian beam is varied to be $\vartheta^i = 10^\circ$, and the average scattered fields using Monte Carlo simulations are plotted for the vertical polarization from surface only and from surface with buried spheroid, respectively. Also, slight distortion is observed in the object signature at this oblique incident angle for both polarizations as shown in Figs. 7c and 7d.

The CPU time required for the SDFMM to calculate the surface current unknowns for each rough surface realization are 38 minutes to fill in the impedance matrix, 72 minutes for the

TFQMR (transpose-free quasi-minimal residual) iterative solver [28], and 3 minutes to calculate the near field scattered above the ground with $0.1\lambda_0$ resolution. The computer memory required to run the SDFMM is 950MB. These computations were conducted using one processor on Compaq GS140 EV6 machine.

As known, the size of Monte Carlo sample (number of rough surface realizations) increases dramatically with the increase of ground roughness parameters. For flat ground only one surface is needed while for very rough ground, hundreds of realizations are needed for efficient conduction of Monte Carlo simulations. The adequate sample size can be obtained by gradually increasing the number of realizations used in the calculations till no change is observed in numerical results. The CPU time dramatically increases with the increase of Monte Carlo sample size. This barrier can be eliminated by the possible parallel implementation of the SDFMM code similar to the SMCG method [29].

IV. CONCLUSIONS

We demonstrated the implementation of the SDFMM to calculate the unknown surface currents on a random rough surface with buried 3-D penetrable object. The significant speed of this algorithm allowed an efficient conduction of the Monte Carlo simulations. The average and the standard deviation of the near fields scattered from the ground with the buried object are calculated. Our results showed that the presence of the rough interface distorts the scattered fields from the buried object even for the small roughness parameters considered here. Interestingly, we also showed that the standard deviation statistic of the object signals is large relative to the signal itself. Recognizable object signature is observed even when the object is not

buried under the center of the ground where the incident Gaussian beam is pointing. Having determined some of the statistics of both clutter and target signals, it is possible to apply statistical signal processing for target detection, which is the subject of future work.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. W. Chew, Prof. E. Michielssen and Dr. V. Jandhyala at UIUC for allowing them the use and modification of the SDFMM computer code for the current application. This research was sponsored by the Army Research Office Demining MURI grant # DAA 0-55-97-0013 and in part by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821. This work benefited from the allocation of time at the Northeastern University Advanced Scientific Computation Center (NU-ASCC).

REFERENCES

- [1] L. Tsang, C. H. Chan, K. Pak, H. Sangani, A. Ishimaru and P. Phu, "Monte Carlo simulations of large-scale composite random rough-surface scattering based on the banded-matrix iterative approach," *J. Opt. Soc. Am. A*, vol. 11, no. 2, pp. 691-696, February 1994.
- [2] R. L. Wagner, J. Song and W. C. Chew, "Monte Carlo simulation of electromagnetic scattering from two-dimensional random rough surfaces," *IEEE Trans. Antennas Propagat.*, vol. 45, no. 2, pp. 235-245, February 1997.
- [3] C. H. Chan, L. Tsang, and Q. Li, "Monte Carlo simulations of large-scale one dimensional random rough-surface scattering at near grazing incidence: penetrable case," *IEEE Trans. on Anten. and Prop.*, vol. 46, no. 1, pp. 142-149, January 1998.
- [4] F. D. Hastings, J. B. Schneider and S. L. Broschat, "A Monte Carlo FDTD technique for rough surface scattering," *IEEE Trans. Antennas Propagat.*, vol. 43, no. 1183-1191, November 1995.
- [5] J. T. Johnson, L. Tsang, R. T. Shin, K. Pak, C. H. Chan, A. Ishimaru and Y. Kuga, "Backscattering enhancement of electromagnetic waves from two-dimensional perfectly conducting random rough surfaces: A comparison of Monte Carlo simulations with experimental data," *IEEE Trans. Antennas Propagat.*, vol. 44, no. 5, pp. 748-756, May 1996.
- [6] K. O'Neill, R. F. Lussky Jr. and K. D. Paulsen, "Scattering from a metallic object embedded near the randomly rough surface of a lossy dielectric," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, no. 2, pp. 367-376, March 1996.

- [7] G. Zhang, L. Tsang and Y. Kuga, "Studies of the angular correlation function of scattering by random rough surfaces with and without a buried object," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, no. 2, pp. 444-453, March 1997.
- [8] A. Madrazo and M. Nieto-Vesperinas, "Scattering of light and other electromagnetic waves from a body buried beneath a highly rough random surface," *J. Opt. Soc. Am. A*, vol. 14, no. 8, pp. 1859-1866, August 1997.
- [9] N. Geng, A. Sullivan and L. Carin, "Multilevel fast-multipole algorithm for scattering from conducting targets above or embedded in a lossy half space," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 4, pp. 1561-1573, July 2000.
- [10] G. Zhang, L. Tsang and K. Pak, "Angular correlation function and scattering coefficient of electromagnetic waves scattered by a buried object under a two-dimensional rough surface," *J. Opt. Soc. Am. A*, vol. 15, no. 12, pp. 2995-3002, December 1998.
- [11] M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, "3-D subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces: Use of the steepest descent fast multipole method (SDFMM)," *IEEE Trans. Geosci. Remote Sensing*, to be published in June 2001.
- [12] V. Rokhlin, "Rapid solution of integral equations of scattering theory in two dimensions," *J. Comput. Phys.*, vol. 36, pp. 414-439, 1990.
- [13] R. Coifman, V. Rokhlin and S. Wandzura, "The fast multipole method for the wave equation: A pedestrian description," *IEEE Antennas Propagat. Mag.*, vol. 35, no. 3, pp. 7-12, June 1993.
- [14] C. C. Lu and W. C. Chew, "Fast algorithm for solving hybrid integral equations," *IEE Proceedings-H*, vol. 140, no.6, December 1993.

- [15] C. C. Lu and W. C. Chew, "A multilevel fast-algorithm for solving a boundary integral equation of wave scattering," *Microwave Opt. Tech. Lett.*, vol. 7, pp.466-470, July, 1994.
- [16] J. M .Song and W. C. Chew, "Multilevel fast-multipole algorithm for solving combined field integral equations of electromagnetic scattering," *Microwave Opt. Tech. Lett.*, vol. 10, pp.14-19, 1995.
- [17] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [18] V. Jandhyala, E. Michielssen, B. Shanker and W.C. Chew, "A combined steepest descent-fast multipole algorithm for the fast analysis of three-dimensional scattering by rough surfaces," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, no. 3, pp. 738-748, May 1998.
- [19] V. Jandhyala, B. Shanker, E. Michielssen, and W. C. Chew, "A fast algorithm for the analysis of scattering by dielectric rough surfaces," *J. Opt. Soc. Am. A*, vol. 15, no. 7, pp. 1877-1885, July 1998.
- [20] M. El-Shenawee, V. Jandhyala, E. Michielssen and W. C. Chew, "The steepest descent fast multipole method (SDFMM) for solving combined field integral equation pertinent to rough surface scattering," *Proc. of the IEEE APS/URSI '99 conf.*, Orlando, Florida, pp. 534-537, July 1999.
- [21] L. Medgyesi-Mitschang, J. Putnam, and M. Gedera, "Generalized method of moments for three-dimensional penetrable scatterers," *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1383-1398, April 1994.
- [22] J. R. Mautz and R. F. Harrington, "H-field, E-field, and combined field solutions for conducting bodies of revolutions," *Arch. Elek. Ubertrag.*, vol. 32, pp.157-164, 1978.

- [23] P. L. Huddleston, L. N. Medgyesi-Mitschang, and J. M. Putnam, "Combined field integral equation formulation for scattering from dielectrically coated conducting bodies," *IEEE Trans. Antennas and Propag.*, vol. Ap-34, pp. 510-520, 1986.
- [24] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Anten. and Prop.*, vol. AP 30, no. 3, pp.409-418, May 1982.
- [25] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.
- [26] J. Curtis, "Dielectric properties pf soils; various sites in Bosnia," *US Army Corp. of Eng., Waterways Experim.*, Station Data Rep., 1996.
- [27] C. A. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons Inc, ch. 6, pp.254-309.
- [28] R. W. Freund, "A transpose-free quasi-minimal residual algorithm for non-hermitian linear systems," *SIAM J. Sci. Comput.*, vol. 14, no. 2, pp. 470-482, March 1993.
- [29] S. Li, C. H. Chan, L. Tsang, Q. Li, and L. Zhou, "Parallel implementation of the sparse matrix/canonical grid method for the analysis of two-dimensional random rough surfaces (three-dimensional scattering problem) on a Beowulf system," *IEEE Trans. Geoscience Rem. Sensing*, vol. 38, no. 4, pp.1600-1608, July 2000.

List of figures

Figure1. Cross section of 2-D rough surface ground with 3-D object buried under the interface.

Figure 2. Penetrable 3-D scatterer R_3 immersed in scatterer R_2 immersed in region R_1 .

Figure 3a. Average of near electric field scattered at $z = 0.5\lambda_0$ above rough ground of rms height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$, incident angle $\vartheta^i = 0^\circ$, for vertical polarization.

Figure 3b. Average of near electric field scattered at $z = 0.5\lambda_0$ above rough ground of rms height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$, incident angle $\vartheta^i = 0^\circ$, for vertical polarization. The oblate spheroid object has dimensions $a = 0.3\lambda_0$ and $b = 0.15\lambda_0$ and is buried at depth $d = 0.3\lambda_0$ under the mean plane of the surface.

Figure 4a. Near electric field scattered at $z = 0.5\lambda_0$ from two individual rough surface realizations with buried spheroid selected from Fig. 3b (surface #1 and #2), the average of 65 electric fields scattered from rough surface with and without the buried spheroid, Fig. 3b and Fig. 3a, respectively, and the scattered electric field from flat ground with and without the buried spheroid. All plotted at $Y = 4.0\lambda_0$, where the spheroid is located at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

Figure 4b. Scattered near electric field due to just the buried object under a flat ground obtained by subtraction. Incident angle $\vartheta^i = 0^\circ$ for horizontal polarization and the spheroid is buried at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

Figure 4c. Scattered near electric field due to just the buried object obtained by subtracting the average electric field of Fig. 3a from the field scattered from only one surface (with buried spheroid) selected from the 65 realizations used to obtain Fig. 3b. Incident angle $\vartheta^i = 0^\circ$ for horizontal polarization and the spheroid is buried at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

Figure 4d. Scattered near electric field due to just the buried object obtained by subtracting fields scattered from only one surface selected from the 65 realizations used to obtain Fig. 3a (surface only) from fields scattered from the same surface (with buried spheroid) selected from the 65 realizations used to obtain Fig. 3b. Incident angle $\vartheta^i = 0^\circ$ for horizontal polarization and the spheroid is buried at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

Figure 4e. Average near electric field scattered due to only buried object obtained by subtracting fields scattered from the 65 realizations used to obtain Fig. 3a from fields scattered from the same 65 realizations and used to obtain Fig. 3b then take the statistical average. Incident angle $\vartheta^i = 0^\circ$ for horizontal polarization. The spheroid is buried at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

Figure 5. Comparison between signatures of three objects located at $x = y = 4.0\lambda_0$, $x = y = 5.0\lambda_0$, and to $x = 6.5\lambda_0, y = 4.0\lambda_0$, respectively.

Figure 6. The standard deviation (*STD*) of near electric fields scattered due to just buried spheroid for the same data in Fig. 4e.

Figure 7a. Average of near electric field scattered at $z = 0.5\lambda_0$ above rough ground of rms height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$, incident angle $\vartheta^i = 10^\circ$ for vertical polarization.

Figure 7b. Average of near electric field scattered at $z = 0.5\lambda_0$ above rough ground of rms height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$, incident angle $\vartheta^i = 10^\circ$, and vertical polarization. The object has dimensions $a = 0.3\lambda_0$ and $b = 0.15\lambda_0$ and is buried at depth $d = 0.3\lambda_0$ under the mean plane of the surface.

Figure 7c. Average near electric field scattered due to just the buried object obtained by subtracting fields scattered from the 65 realizations used to obtain Fig. 7a from fields scattered from the same 65 realizations and used to obtain Fig. 7b and then take the statistical average.

Incident angle $\vartheta^i = 10^\circ$ for vertical polarization.

Figure 7d. Average near electric field scattered due to just the buried object for similar data shown in Fig. 7c but for horizontal polarization.

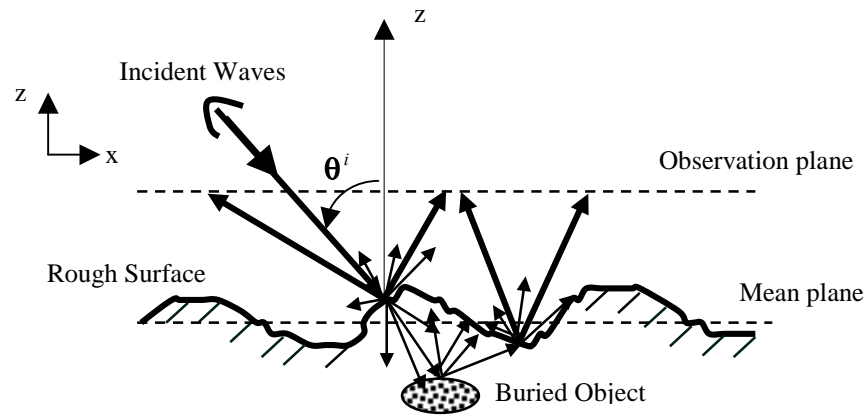


Figure1. Cross section of 2-D rough surface ground with 3-D object buried under the interface.

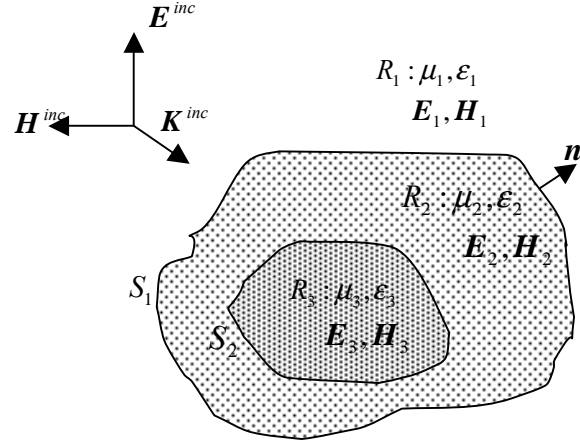


Figure 2. 3-D penetrable scatterer R_3 immersed in scatterer R_2 immersed in region R_1 .

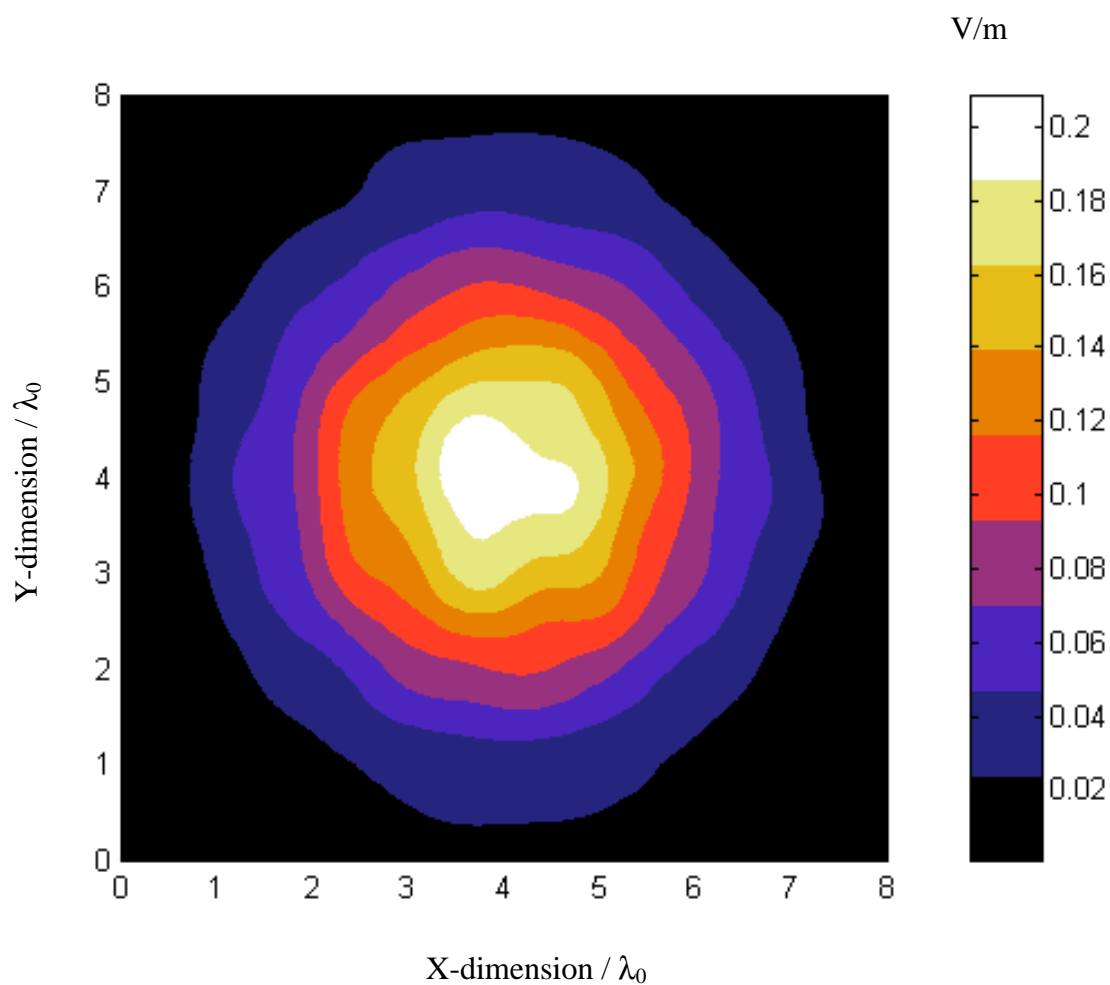


Figure 3a. Average of near electric field scattered at $z = 0.5\lambda_0$ above rough ground of rms height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$, incident angle $\vartheta^i = 0^\circ$, and horizontal polarization.

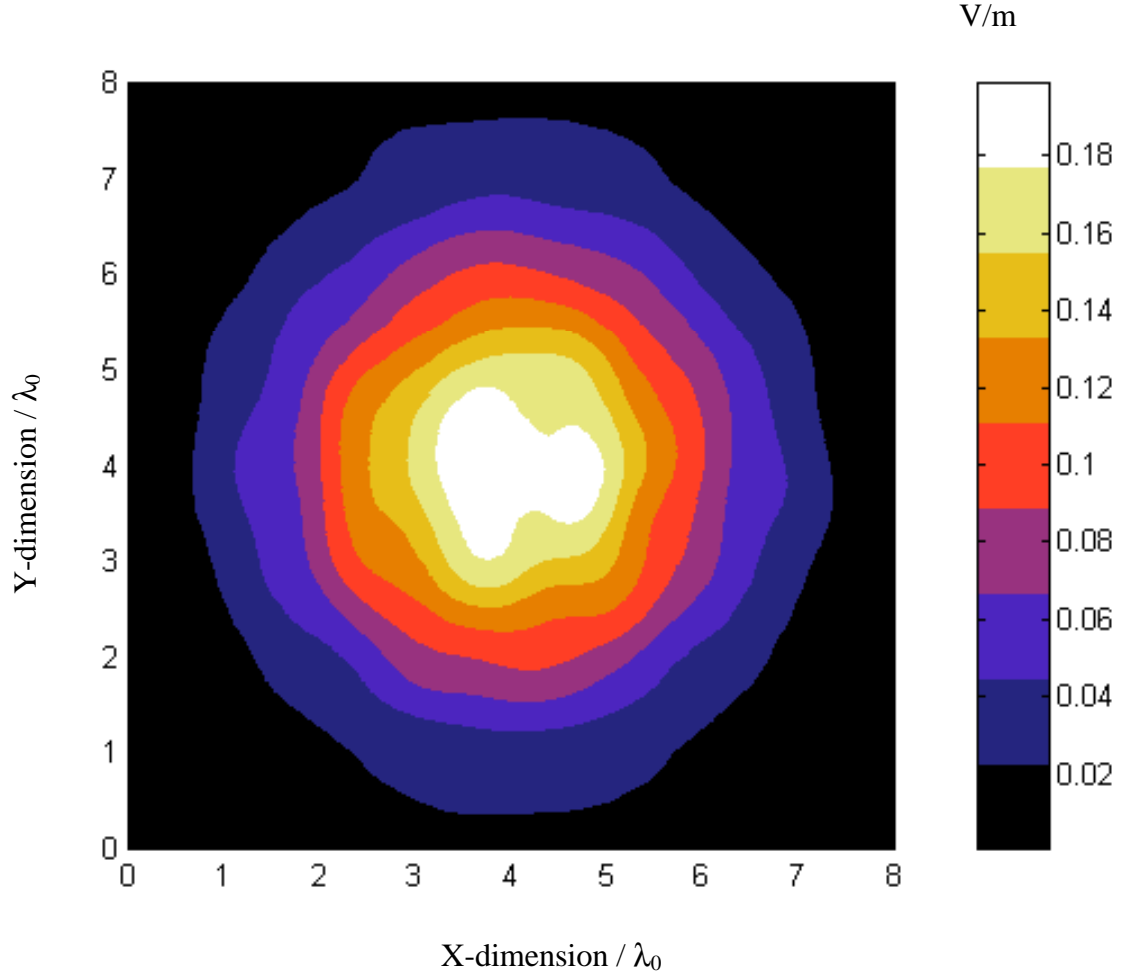


Figure 3b. Average of near electric field scattered at $z = 0.5\lambda_0$ above rough ground of rms height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$, incident angle $\vartheta^i = 0^\circ$, and horizontal polarization. The oblate spheroid has dimensions $a = 0.3\lambda_0$ and $b = 0.15\lambda_0$ and is buried at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

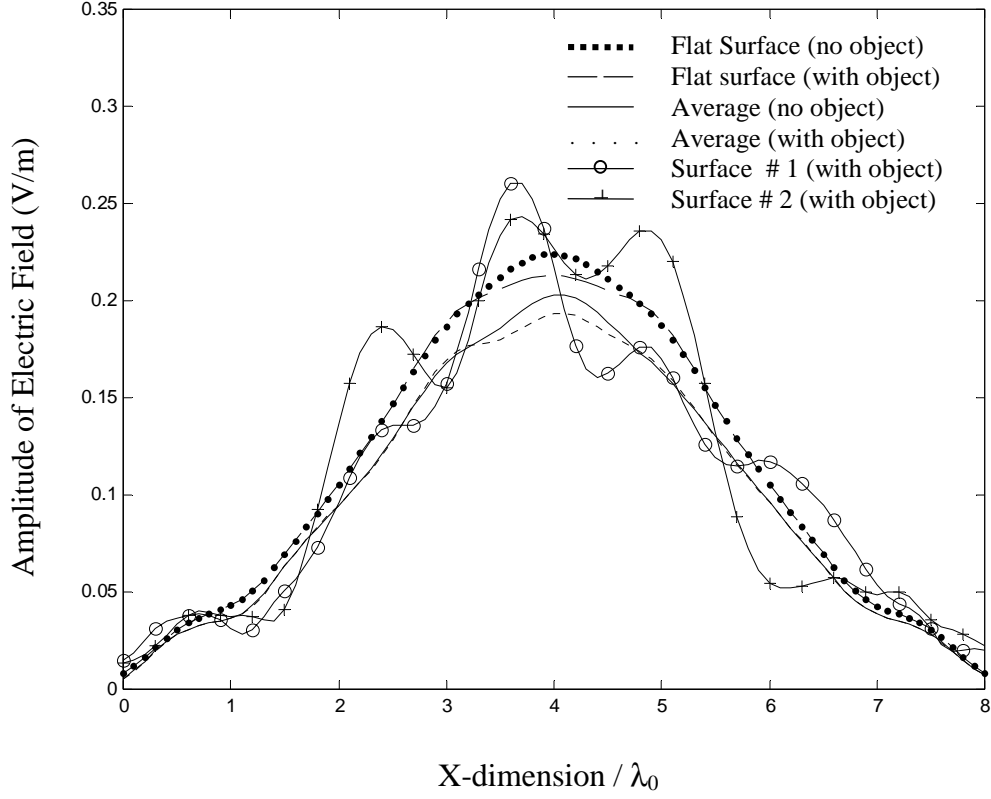


Figure 4a. Near electric field scattered at $z = 0.5\lambda_0$ from two individual rough surface realizations with buried spheroid selected from Fig. 3b (surface #1 and #2), the average of 65 electric fields scattered from rough surface with and without the buried spheroid, Fig. 3b and Fig. 3a, respectively, and the scattered electric field from flat ground with and without the buried spheroid. All plotted at $Y = 4.0\lambda_0$, where the spheroid is located at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

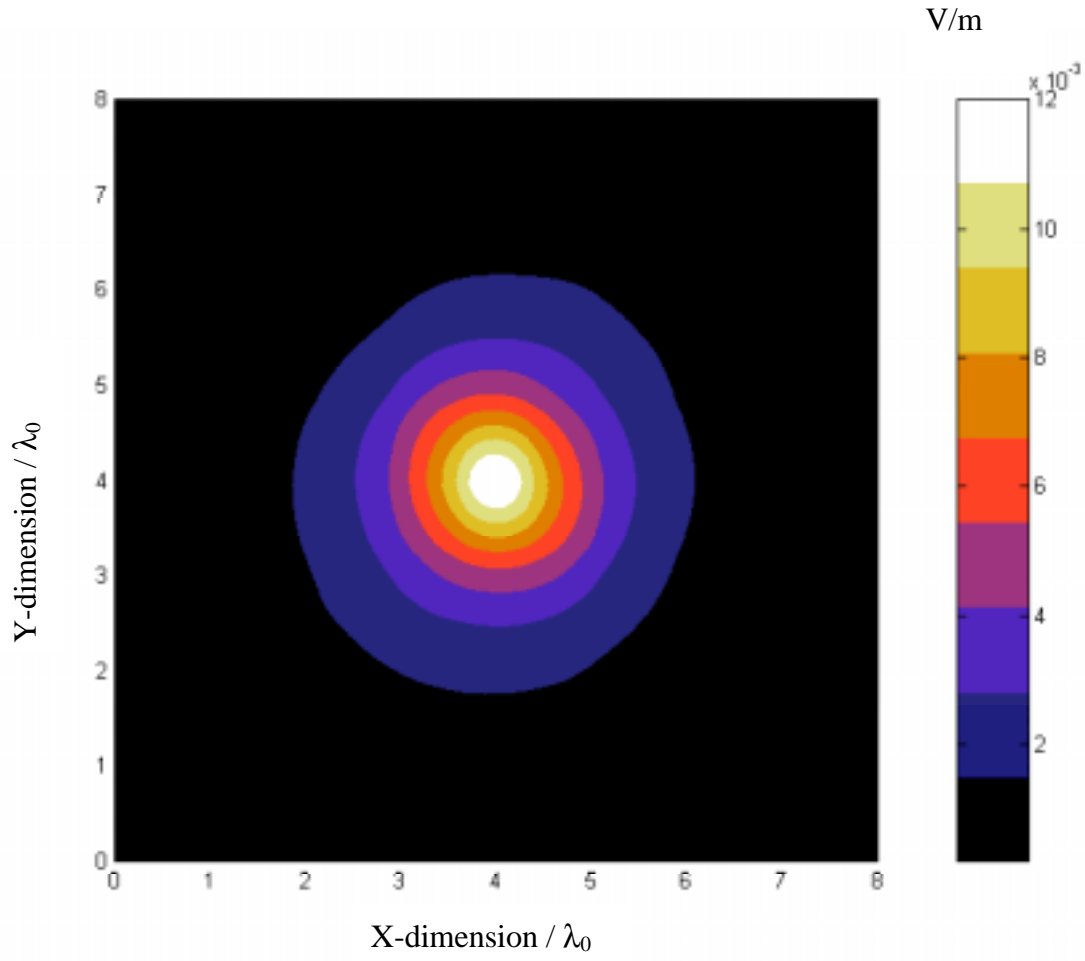


Figure 4b. Scattered near electric field due to just the buried object under a flat ground obtained by subtraction. Incident angle $\vartheta^i = 0^\circ$ for horizontal polarization and the spheroid is buried at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

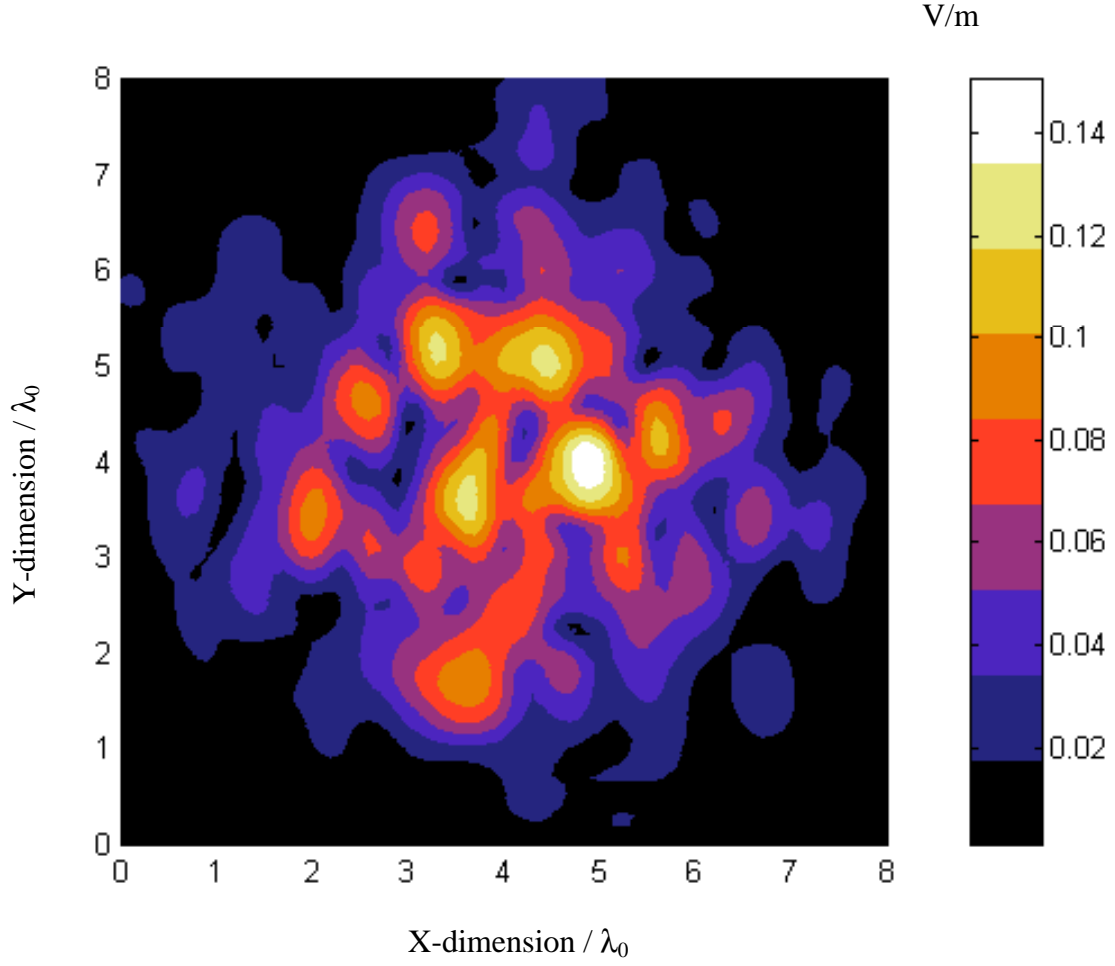


Figure 4c. Scattered near electric field due to just the buried object obtained by subtracting the average electric field of Fig. 3a from the field scattered from only one surface (with buried spheroid) selected from the 65 realizations used to obtain Fig. 3b. Incident angle $\vartheta^i=0^\circ$ for horizontal polarization and the spheroid is buried at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

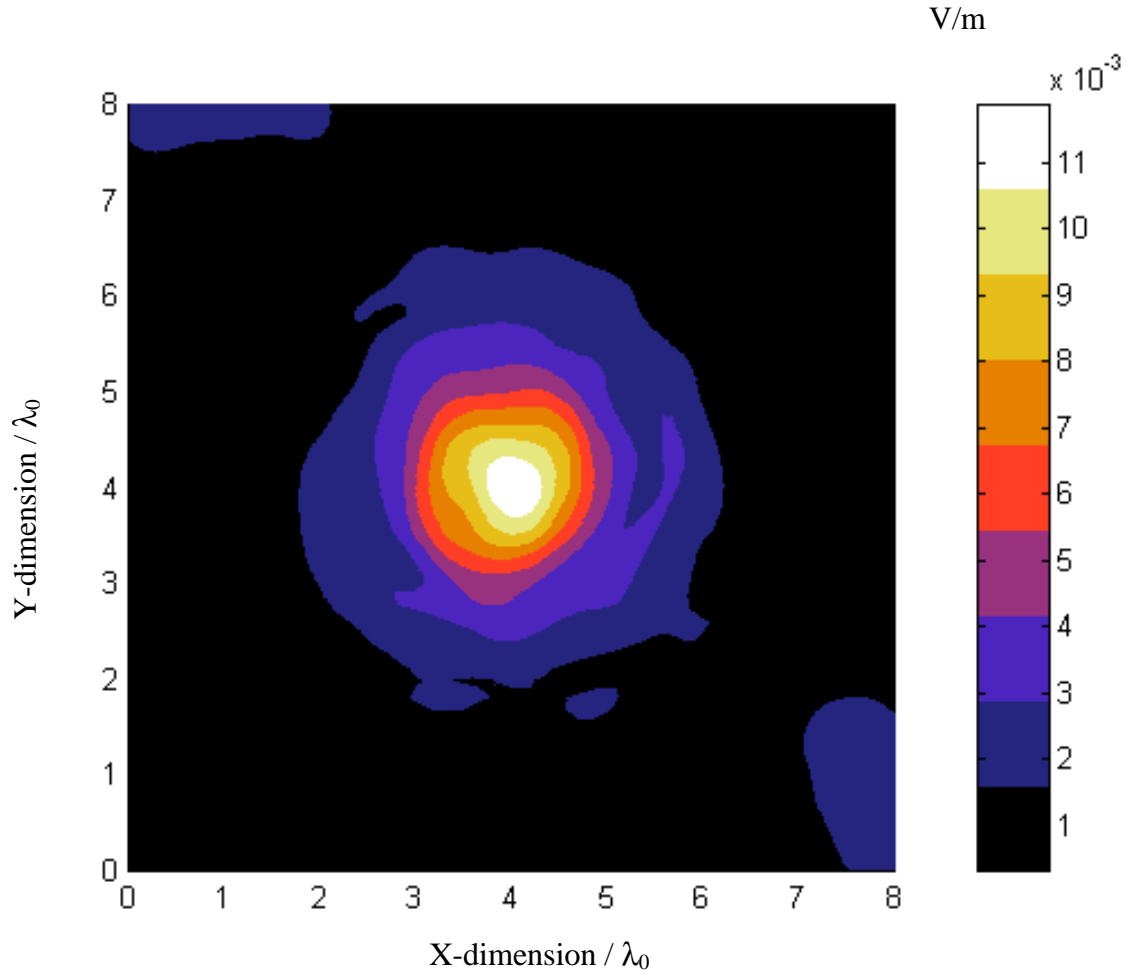


Figure 4d. Scattered near electric field due to just the buried object obtained by subtracting fields scattered from only one surface selected from the 65 realizations used to obtain Fig. 3a (surface only) from fields scattered from the same surface (with buried spheroid) selected from the 65 realizations used to obtain Fig. 3b. Incident angle $\vartheta^i = 0^\circ$ for horizontal polarization and the spheroid is buried at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

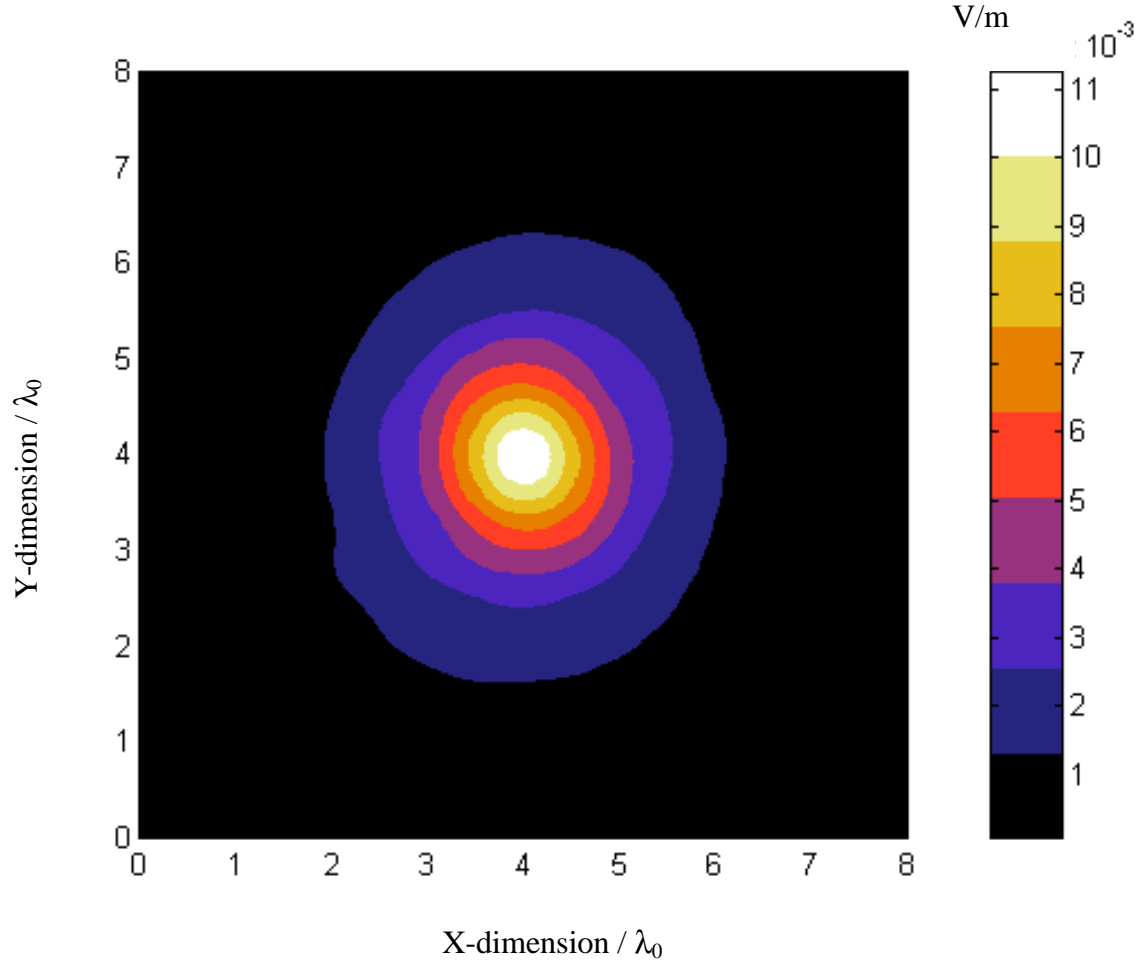


Figure 4e. Average near electric field scattered due to only buried object obtained by subtracting fields scattered from the 65 realizations used to obtain Fig. 3a from fields scattered from the same 65 realizations and used to obtain Fig. 3b then take the statistical average. Incident angle $\vartheta^i = 0^\circ$ for horizontal polarization. The spheroid is buried at $x = y = 4.0\lambda_0, z = -0.3\lambda_0$.

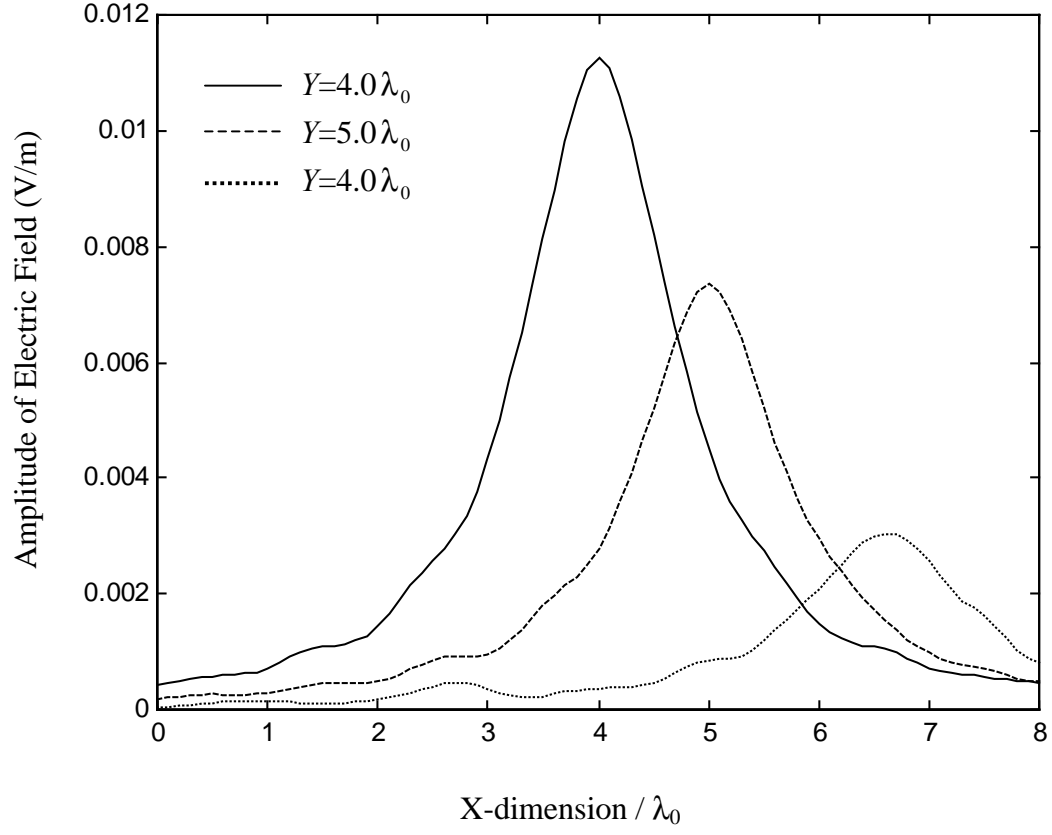


Figure 5. Comparison between signatures of three objects located at $x = y = 4.0\lambda_0$, $x = y = 5.0\lambda_0$, and to $x = 6.5\lambda_0, y = 4.0\lambda_0$, respectively.

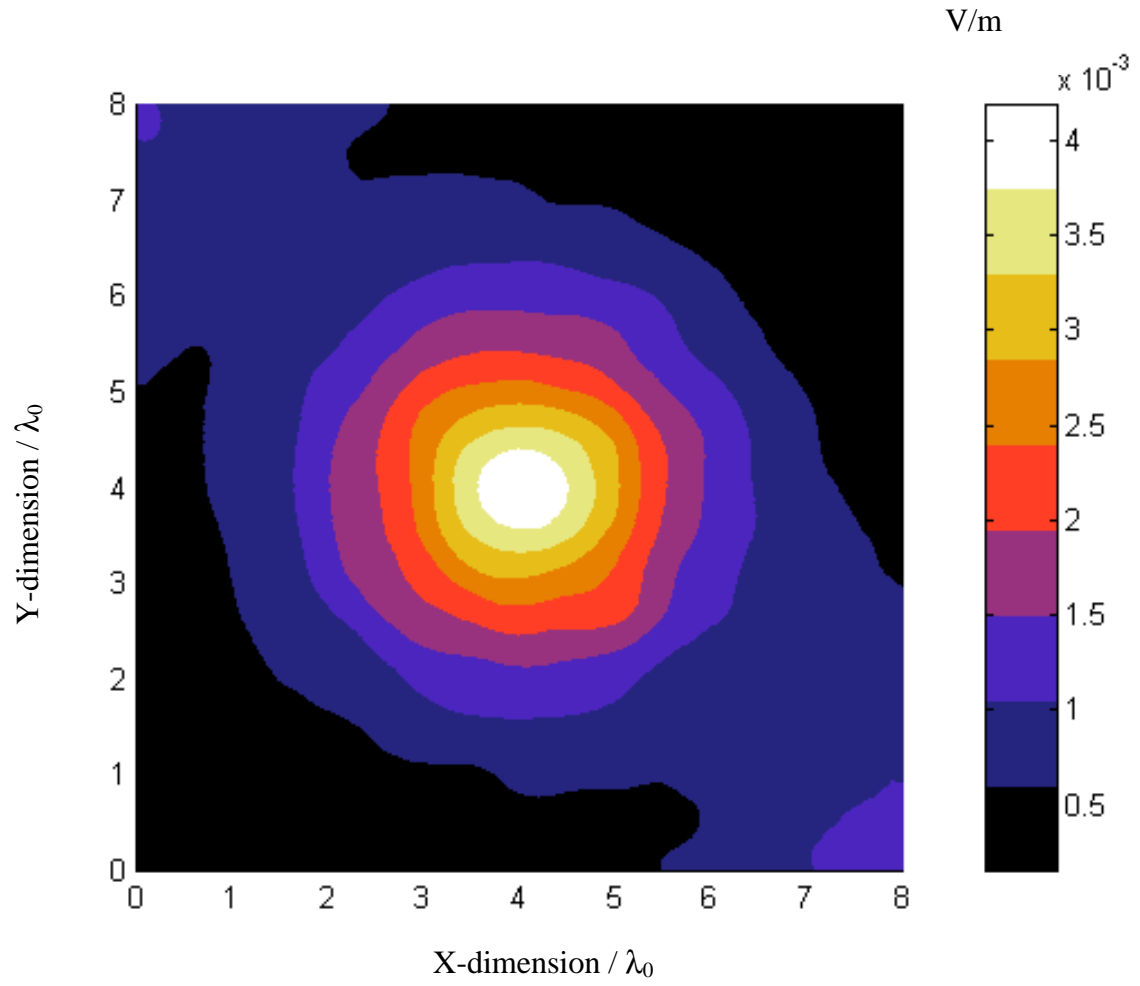


Figure 6. The standard deviation (*STD*) of near electric fields scattered due to just buried spheroid for the same data in Fig. 4e.

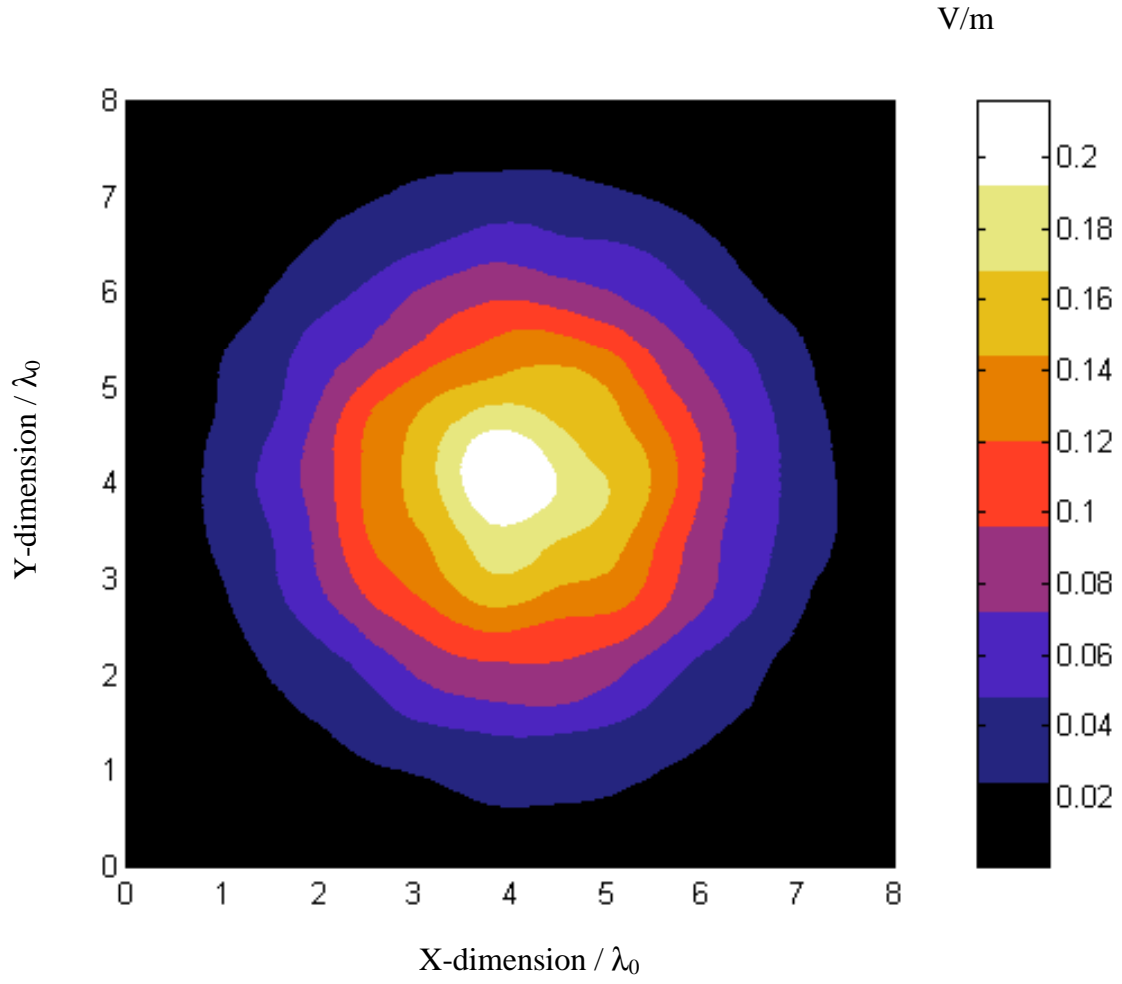


Figure 7a. Average of near electric field scattered at $z = 0.5\lambda_0$ above rough ground of rms height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$, incident angle $\vartheta^i = 10^\circ$ for vertical polarization.

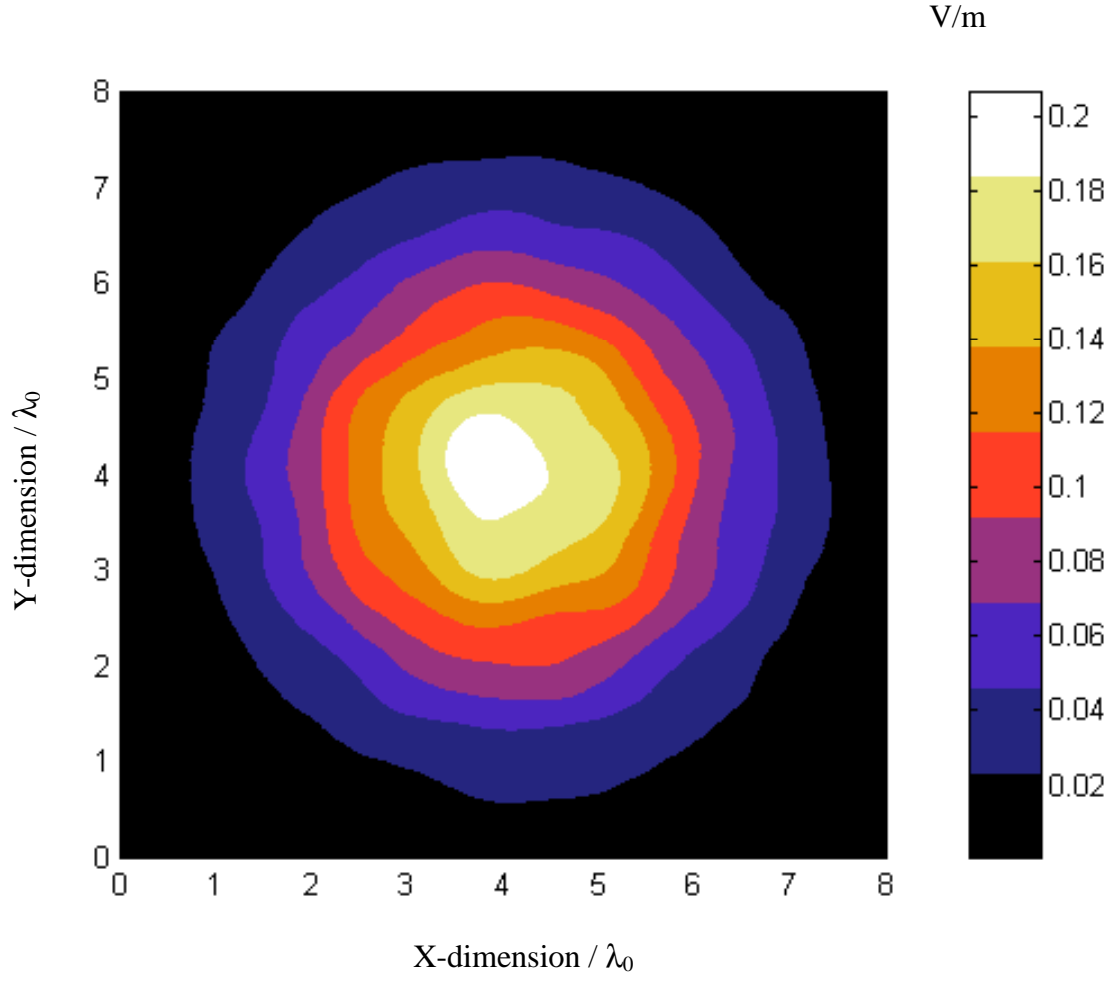


Figure 7b. Average of near electric field scattered at $z = 0.5\lambda_0$ above rough ground of rms height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$, incident angle $\vartheta^i = 10^\circ$, and vertical polarization. The object has dimensions $a = 0.3\lambda_0$ and $b = 0.15\lambda_0$ and is buried at depth $d = 0.3\lambda_0$ under the mean plane of the surface.

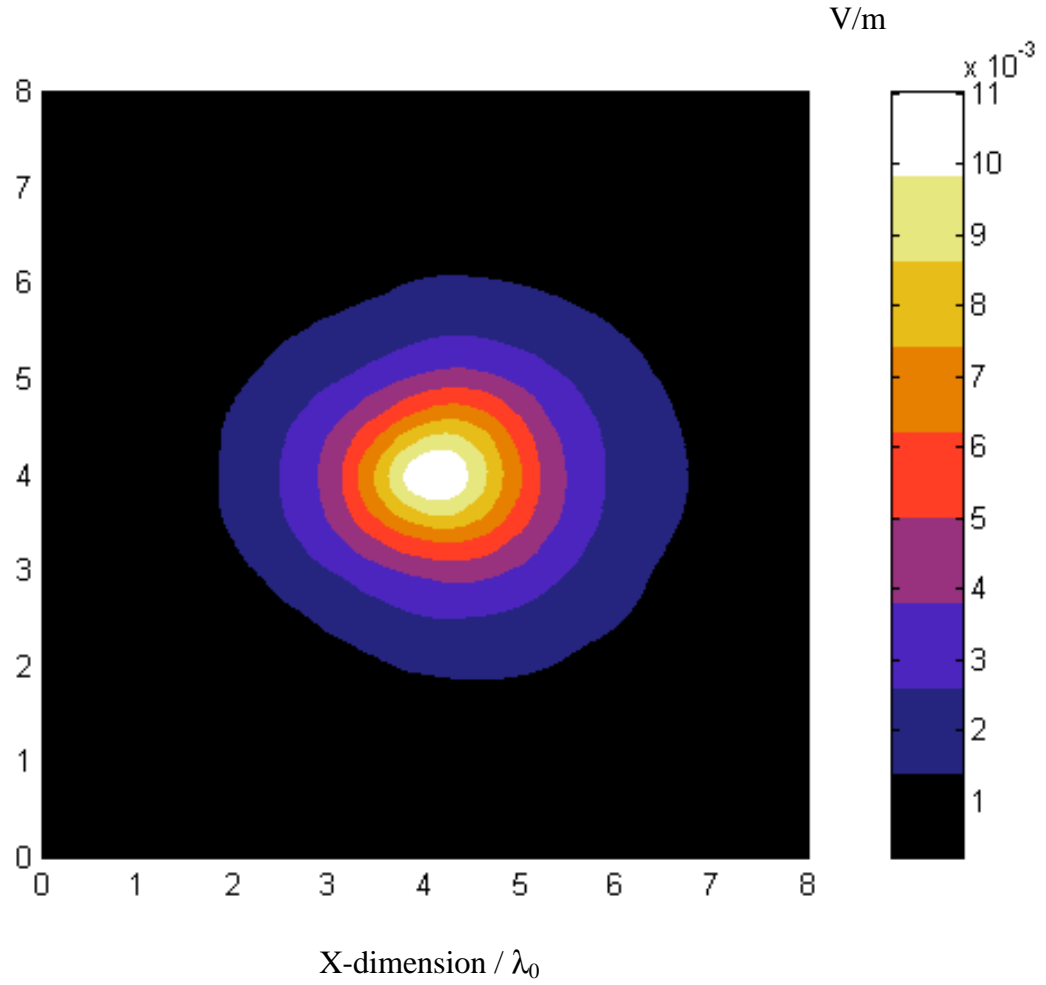


Figure 7c. Average near electric field scattered due to just the buried object obtained by subtracting fields scattered from the 65 realizations used to obtain Fig. 7a from fields scattered from the same 65 realizations and used to obtain Fig. 7b and then take the statistical average. Incident angle $\vartheta^i = 10^\circ$ for vertical polarization.

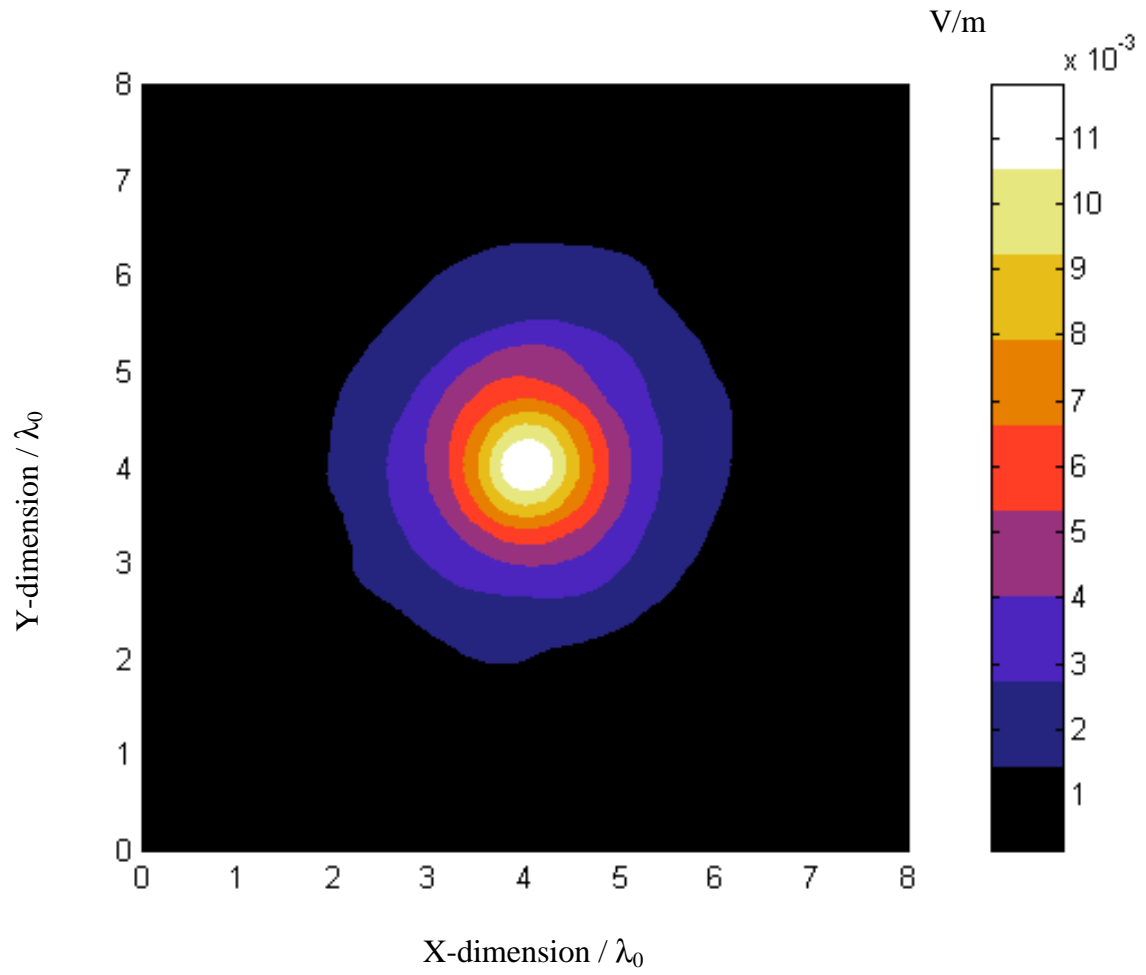


Figure 7d. Average near electric field scattered due to just the buried object for data shown in Fig. 7c but for horizontal polarization.

3-D Subsurface Analysis of Electromagnetic Scattering from Penetrable/PEC Objects Buried under Rough Surfaces: Use of the Steepest Descent Fast Multipole Method (SDFMM).

M. El-Shenawee¹, C. Rappaport², E. Miller² and M. Silevitch²

¹Department of Electrical Engineering

University of Arkansas

Fayetteville, Arkansas 72701

Tel: (501)-575-6582, Fax: (501)-575-7967

magda@uark.edu

²Center for Subsurface Sensing and Imaging Systems (CenSSIS)

Department of Electrical and Computer Engineering

Northeastern University, 235 Forsyth Bld

Boston, MA 02115

Tel: 617-373-2064, Fax: 617-373-8627

rappaport@cer.neu.edu

ABSTRACT

The electromagnetic scattering from a three-dimensional shallow object buried under a two-dimensional random rough dielectric surface is analyzed in this work. The buried object can be a perfect electric conductor (PEC) or can be a penetrable dielectric with size and burial depth comparable to the free space wavelength. The random rough ground surface is characterized with Gaussian statistics for surface height and for surface autocorrelation function. The PMCHW (Poggio, Miller, Chang, Harrington, and Wu) integral equations are implemented and extended in this work. The integral equation-based Steepest Descent Fast Multipole Method (SDFMM), that was originally developed at UIUC, has been used and the computer code based on this

algorithm has been successfully modified to handle the current application. The significant potential of the SDFMM code is that it calculates the unknown moment method surface electric and magnetic currents on the scatterer in a dramatically fast, efficient and accurate manner. Interactions between the rough surface interface and the buried object are fully taken into account with this new formulation. Ten incident Gaussian beams with same elevation angle and different azimuth angles are generated for excitation as a one possible way of having multiple views of a given target.

The scattered electric fields due to these ten incident beams are calculated in the near zone and their complex vector average over the multiple views is computed. The target signature is obtained by subtracting the electric fields scattered from the rough ground only from those scattered from the ground with the buried anti-personnel mine. Significant polarization dependency is observed for the PEC object signature compared with that of the penetrable object, which can be used in target discrimination. Moreover, fields scattered above the rough ground experience significantly more distortion than fields transmitted into the ground.

I. INTRODUCTION

Subsurface sensing of buried electromagnetically penetrable objects under random rough surfaces has many important applications, such as detection of mines, underground water, buried hazardous environmental waste, and underground petroleum. Modeling the wave scattering is essential for efficient sensor design and for inverse scattering and target reconstruction processing. Of particular interest here is the topic of anti-personnel landmine detection, as shown schematically in Fig. 1. The analysis of scattering and transmission of the electromagnetic waves

in the presence of a random rough dielectric interface and in the near field of the sensing systems is a crucial step for subsurface object detection problems in general and landmine remediation applications in particular. Generally, this fully three dimensional problem must be treated numerically. However the calculation of the required fields using conventional techniques (e.g. moment method, finite elements, or finite differences) is a computationally intensive undertaking especially for large dielectric constant soil. The computational complexity of the problem dramatically increases upon inserting objects under the rough interface, especially when these objects are penetrable.

To deal with this complex scenario, we apply the SDFMM, originally developed by V. Jandhyala, E. Michielssen, and W. Chew [1]-[3] to analyze 3-D scattering problems of quasi-planar structures, which is well suited to handle large computational domains of this sort. In particular, the SDFMM computer code has been successfully modified to handle objects with burial depth comparable to the wavelength and varieties of ground soil characteristics. Mine fields are located in different places all over the world where the dielectric constant of the soil significantly varies, for example, $\epsilon_r = 2.5 - i0.18$ for Bosnian soil with 3.8% moisture at 1GHz, $\epsilon_r = 5.4 - i0.04$ for Puerto Rican soil with 10% moisture at 960MHz, while $\epsilon_r = 9.18 - i1.26$ for Bosnian soil with 25% moisture at 1GHz. These modifications and their use in calculating scattering simulations are presented in this work. The SDFMM is a hybridization of three methods: the Method of Moments (MOM), the Steepest Descent Path (SDP) method, and the Fast Multipole Method (FMM). The basic concept of the SDFMM is to subdivide the rough surface and the buried object scatterers into small groups and each group includes some of the moment method surface current unknowns. The interactions between these current unknowns in

a given group or in nearby groups are calculated using the standard MOM, which incorporates multiplying the elements of the impedance matrix by the current coefficient vector. All other interactions are calculated in one step without conducting matrix vector multiplication by using the fast multipole method, which is also hybridized with the SDP integration rule. The details of the SDFMM can be found in [1]-[6].

In section II, we present the analysis of the problem based on extending the implementation of the PMCHW [7] integral equations for the three regions; air, soil, and buried object. The object can either be penetrable, as characterized by its dielectric constant, or a PEC. In section III, the numerical results for the electric fields in the near zone are shown. Conclusions and future work will be summarized in Section IV. The proof of an important and useful identity used in the computations is given in Appendix A.

II. FORMULATION

In this section, we formulate the problem of a penetrable arbitrary 3-D scatterer R_2 immersed in R_1 with an interior scatterer of different material R_3 as shown in Fig. 2a. The three regions, R_1 , R_2 and R_3 have permittivity and permeability given by ϵ_1 and μ_1 , ϵ_2 and μ_2 , and ϵ_3 and μ_3 , respectively. As shown in Figs. 2a-d, the region R_2 is bounded by surfaces S_1 (exterior) and S_2 (interior) with normal vectors \bar{n}_1 and \bar{n}_2 on S_1 and, \bar{n}_3 and \bar{n}_4 on S_2 . The region R_3 is bounded by surface S_2 . The normal vectors \bar{n}_1 , \bar{n}_2 and \bar{n}_3 , and \bar{n}_4 are chosen to point toward regions R_1 , R_2 , and R_3 , respectively. The electric and magnetic fields in regions R_1 , R_2 , and R_3 are \bar{E}_1, \bar{H}_1 , due to equivalent currents \bar{J}_1, \bar{M}_1 , as shown in Fig. 2b, \bar{E}_2, \bar{H}_2 due to equivalent currents

\bar{J}_2, \bar{M}_2 and \bar{J}_3, \bar{M}_3 , as shown in Fig. 2c, and \bar{E}_3, \bar{H}_3 , due to equivalent currents \bar{J}_4, \bar{M}_4 , as shown in Fig. 2d. Following the analysis in [7], [9], [10] with time dependence $\exp(i\omega t)$ assumed, the electric and magnetic fields at an arbitrary point \bar{r} in region R_1 are:

$$\theta_1(\bar{r})\bar{E}_1(\bar{r}) = \bar{E}^{inc}(\bar{r}) - \int_{S_1} (i\omega\mu_1(\bar{n}_1 \times \bar{H}_1)\Phi_1 - (\bar{n}_1 \times \bar{E}_1) \times \nabla'\Phi_1 - (\bar{n}_1 \cdot \bar{E}_1)\nabla'\Phi_1) ds' \quad (1a)$$

$$\theta_1(\bar{r})\bar{H}_1(\bar{r}) = \bar{H}^{inc}(\bar{r}) + \int_{S_1} (i\omega\epsilon_1(\bar{n}_1 \times \bar{E}_1)\Phi_1 + (\bar{n}_1 \times \bar{H}_1) \times \nabla'\Phi_1 + (\bar{n}_1 \cdot \bar{H}_1)\nabla'\Phi_1) ds' \quad (1b)$$

Similarly, the electric and magnetic fields in region R_2 are:

$$\begin{aligned} \theta_2(\bar{r})\bar{E}_2(\bar{r}) = & - \int_{S_1} (i\omega\mu_2(\bar{n}_2 \times \bar{H}_2)\Phi_2 - (\bar{n}_2 \times \bar{E}_2) \times \nabla'\Phi_2 - (\bar{n}_2 \cdot \bar{E}_2)\nabla'\Phi_2) ds' \\ & - \int_{S_2} (i\omega\mu_2(\bar{n}_3 \times \bar{H}_2)\Phi_2 - (\bar{n}_3 \times \bar{E}_2) \times \nabla'\Phi_2 - (\bar{n}_3 \cdot \bar{E}_2)\nabla'\Phi_2) ds' \end{aligned} \quad (1c)$$

$$\begin{aligned} \theta_2(\bar{r})\bar{H}_2(\bar{r}) = & \int_{S_1} (i\omega\epsilon_2(\bar{n}_2 \times \bar{E}_2)\Phi_2 + (\bar{n}_2 \times \bar{H}_2) \times \nabla'\Phi_2 + (\bar{n}_2 \cdot \bar{H}_2)\nabla'\Phi_2) ds' \\ & + \int_{S_2} (i\omega\epsilon_2(\bar{n}_3 \times \bar{E}_2)\Phi_2 + (\bar{n}_3 \times \bar{H}_2) \times \nabla'\Phi_2 + (\bar{n}_3 \cdot \bar{H}_2)\nabla'\Phi_2) ds' \end{aligned} \quad (1d)$$

Finally, the electric and magnetic fields in region R_3 , are:

$$\theta_3(\bar{r})\bar{E}_3(\bar{r}) = - \int_{S_2} (i\omega\mu_3(\bar{n}_4 \times \bar{H}_3)\Phi_3 - (\bar{n}_4 \times \bar{E}_3) \times \nabla'\Phi_3 - (\bar{n}_4 \cdot \bar{E}_3)\nabla'\Phi_3) ds' \quad (1e)$$

$$\theta_3(\bar{r})\bar{H}_3(\bar{r}) = \int_{S_2} (i\omega\epsilon_3(\bar{n}_4 \times \bar{E}_3)\Phi_3 + (\bar{n}_4 \times \bar{H}_3) \times \nabla'\Phi_3 + (\bar{n}_4 \cdot \bar{H}_3)\nabla'\Phi_3) ds' \quad (1f)$$

In Eqs. 1a-f, \bar{r} is the field point, \bar{r}' is the source point, and Φ_i is the 3-D scalar Green's function

$\exp(-ik_i|\bar{r} - \bar{r}'|)/4\pi|\bar{r} - \bar{r}'|$ in which $k_i = \omega\sqrt{\epsilon_i\mu_i}$ is the wave number in each region, $i=1, 2$ and

3. The symbol $\theta_i(\bar{r})$ is the Heaviside function that gives the jump condition at surface S ,

$(\theta_i(\bar{r})=1 \text{ for } \bar{r} \in R_i, 0.5 \text{ for } \bar{r} \in \partial R_i, 0 \text{ otherwise})$ [7], [9], [10], where the symbol ∂R_i

represents the boundary of region R_i . The equivalent currents \bar{J}_1 and \bar{M}_1 are related to the

tangential electric and magnetic field components in region R_1 as $\bar{J}_1 = \bar{n}_1 \times \bar{H}_1|_{S_1}$ and $\bar{M}_1 = \bar{E}_1 \times \bar{n}_1|_{S_1}$. Similarly, the equivalent currents $\bar{J}_2, \bar{J}_3, \bar{J}_4, \bar{M}_2, \bar{M}_3$, and \bar{M}_4 can be related to tangential electric and magnetic fields on S_1 and S_2 . Moreover, the normal electric and magnetic field components in region R_1 are related to the equivalent currents as $\bar{n}_1 \cdot \bar{E}_1 = -\nabla \cdot \bar{J}_1 / i\omega\epsilon_1$ and $\bar{n}_1 \cdot \bar{H}_1 = -\nabla \cdot \bar{M}_1 / i\omega\mu_1$. Similarly, the normal electric and magnetic field components in regions R_2 and R_3 can be expressed as functions of their associated equivalent currents. Applying the boundary conditions for the tangential electric and magnetic field components on surface S_1 leads to $\bar{J}_1 = -\bar{J}_2$ and $\bar{M}_1 = -\bar{M}_2$, and on surface S_2 leads to $\bar{J}_3 = -\bar{J}_4$ and $\bar{M}_3 = -\bar{M}_4$. Thus, upon equating the tangential component of the electric fields (\bar{E}_1 and \bar{E}_2) and magnetic fields (\bar{H}_1 and \bar{H}_2), on surface S_1 , we get:

$$\bar{E}^{inc}(\bar{r})|_{\text{tang.}} = [(L_1 + L_2)\bar{J}_1 - (K_1 + K_2)\bar{M}_1 - L_3\bar{J}_3 + K_3\bar{M}_3]_{\text{tang.}} \quad (2)$$

$$\bar{H}^{inc}(\bar{r})|_{\text{tang.}} = \left[(K_1 + K_2)\bar{J}_1 + \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{M}_1 - K_3\bar{J}_3 - \frac{L_3}{\eta_2^2} \bar{M}_3 \right]_{\text{tang.}} \quad (3)$$

and on surface S_2 , we get:

$$0 = [-L_2\bar{J}_1 + K_2\bar{M}_1 + (L_3 + L_4)\bar{J}_3 - (K_3 + K_4)\bar{M}_3]_{\text{tang.}} \quad (4)$$

$$0 = \left[-K_2\bar{J}_1 - \frac{L_2\bar{M}_1}{\eta_2^2} + (K_3 + K_4)\bar{J}_3 + \left(\frac{L_3}{\eta_2^2} + \frac{L_4}{\eta_3^2} \right) \bar{M}_3 \right]_{\text{tang.}} \quad (5)$$

where, operators L_j and $K_j, j=1, 2, 3$ and 4, are given by:

$$L_{1,2}\bar{X} = \int_{S_1} \left\{ i\omega\mu_{1,2}\Phi_{1,2}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{1,2}} \nabla \nabla' \cdot \bar{X}(\bar{r}') \Phi_{1,2} \right\} ds' \quad (6a)$$

$$L_{3,4}\bar{X} = \int_{S_2} \left\{ i\omega\mu_{2,3}\Phi_{2,3}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{2,3}}\nabla\nabla'\cdot\bar{X}(\bar{r}')\Phi_{2,3} \right\} ds' \quad (6b)$$

$$K_{1,2}\bar{X} = \int_{S_1} \bar{X}(\bar{r}') \times \nabla\Phi_{1,2} ds' \quad (6c)$$

$$K_{3,4}\bar{X} = \int_{S_2} \bar{X}(\bar{r}') \times \nabla\Phi_{2,3} ds' \quad (6d)$$

The vector \bar{X} represents the surface electric current \bar{J} and/or the surface magnetic current \bar{M} on surface S_1 or on surface S_2 . The intrinsic impedance in each region is $\eta_i = \sqrt{\mu_i/\epsilon_i}$, where the dielectric permittivity and permeability are ϵ_i and μ_i , respectively. Equations 2-5 are considered the extension of the PMCHW formulation which has been shown to yield a unique solution at internal resonances associated with the corresponding conducting scatterer [7], [11], [12].

The equivalent electric and magnetic currents $\bar{J}_{1,3}$ and $\bar{M}_{1,3}$ on S_1 and on S_2 , are approximated in Eqs. 2-5 using the RWG vector basis functions $\bar{j}(\bar{r})$ [7], [8] as follows

$$\bar{J}_1(\bar{r}) = \sum_{n=1}^N I_{1n} \bar{j}_{1n}(\bar{r}), \quad \bar{M}_1(\bar{r}) = \eta_1 \sum_{n=1}^N I_{2n} \bar{j}_{1n}(\bar{r}), \quad \bar{r} \in S_1 \quad (7a)$$

$$\bar{J}_3(\bar{r}) = \sum_{m=1}^P I_{3m} \bar{j}_{2m}(\bar{r}), \quad \bar{M}_3(\bar{r}) = \eta_1 \sum_{m=1}^P I_{4m} \bar{j}_{2m}(\bar{r}), \quad \bar{r} \in S_2 \quad (7b)$$

where η_1 is the intrinsic impedance of region R_1 , which is the free space in this work. For the sake of computational efficiency, both magnetic currents \bar{M}_1 and \bar{M}_3 are normalized with the same value of η_1 as shown in Eqs. 7a and 7b. The number of the unknown coefficients on surface S_1 with coefficient vectors I_1 and I_2 , which are associated with \bar{J}_1 and \bar{M}_1 , is denoted by N ,

while the number of the unknown coefficients on surface S_2 with coefficient vectors I_3 and I_4 , associated with \bar{J}_3 and \bar{M}_3 , is denoted by P , in Eqs. 7a and 7b. Next, we substitute Eqs. 7a and 7b in Eqs. 2-5, along with multiplying the H -field equations, Eqs. 3 and 5, by the same normalizing intrinsic impedance η_1 . Upon applying Galerkin's method for testing the E - and H -field equations, the original integral equations are thus transformed into a set of linear equations $\bar{\bar{Z}}\bar{I} = \bar{V}$ given by:

$$\begin{pmatrix} \bar{Z}_{11} & \bar{Z}_{12} \\ \bar{Z}_{21} & \bar{Z}_{22} \end{pmatrix} \begin{pmatrix} I^1 \\ I^2 \end{pmatrix} = \begin{pmatrix} \bar{V}_1 \\ 0 \end{pmatrix} \quad (8)$$

where, \bar{Z}_{11} is a submatrix of order $2N \times 2N$ which represents interactions between elements only on surface S_1 , \bar{Z}_{12} is a submatrix of order $2N \times 2P$ which represents interactions between elements on surface S_1 and elements on surface S_2 , \bar{Z}_{21} is a submatrix of order $2P \times 2N$ which represents interactions between elements on surface S_2 and elements on surface S_1 , and \bar{Z}_{22} is a submatrix of order $2P \times 2P$ which represents interactions between elements only on surface S_2 . Thus the total matrix $\bar{\bar{Z}}$ has order of $2(N+P) \times 2(N+P)$. Note that the vector \bar{V}_1 is a submatrix of order $2N \times 1$ and composed of the tested tangential incident electric field \bar{E}^{inc} and the tested normalized magnetic field $\eta_1 \bar{H}^{inc}$ on surface S_1 . The unknown current coefficients I^1 and I^2 are submatrices of order $2N \times 1$ and $2P \times 1$. The elements of the submatrix \bar{Z}_{11} are given by

$$\bar{Z}_{11} = \begin{pmatrix} \langle \bar{j}_1, (L_1 + L_2) \bar{j}_1 \rangle_{S_1} & \langle \bar{j}_1, -\eta_1 (K_1 + K_2) \bar{j}_1 \rangle_{S_1} \\ \langle \bar{j}_1, \eta_1 (K_1 + K_2) \bar{j}_1 \rangle_{S_1} & \left\langle \bar{j}_1, \eta_1^2 \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{j}_1 \right\rangle_{S_1} \end{pmatrix} \quad (9a)$$

in which $\langle \bar{A}, \bar{B} \rangle_s = \int_S \bar{A}^* \cdot \bar{B} ds$ denotes the complex inner product between vector functions \bar{A}

and \bar{B} on a surface S [7]. Similarly, the elements of the submatrices \bar{Z}_{12} , \bar{Z}_{21} , and \bar{Z}_{22} are:

$$\bar{Z}_{12} = \begin{pmatrix} \langle \bar{j}_1, -L_3 \bar{j}_2 \rangle_{S_1} & \langle \bar{j}_1, \eta_1 K_3 \bar{j}_2 \rangle_{S_1} \\ \langle \bar{j}_1, -\eta_1 K_3 \bar{j}_2 \rangle_{S_1} & \left\langle \bar{j}_1, -\eta_1^2 \left(\frac{L_3}{\eta_2^2} \right) \bar{j}_2 \right\rangle_{S_1} \end{pmatrix} \quad (9b)$$

$$\bar{Z}_{21} = \begin{pmatrix} \langle \bar{j}_2, -L_2 \bar{j}_1 \rangle_{S_2} & \langle \bar{j}_2, \eta_1 K_2 \bar{j}_1 \rangle_{S_2} \\ \langle \bar{j}_2, -\eta_1 K_2 \bar{j}_1 \rangle_{S_2} & \left\langle \bar{j}_2, -\eta_1^2 \left(\frac{L_2}{\eta_2^2} \right) \bar{j}_1 \right\rangle_{S_2} \end{pmatrix} \quad (9c)$$

$$\bar{Z}_{22} = \begin{pmatrix} \langle \bar{j}_2, (L_3 + L_4) \bar{j}_2 \rangle_{S_2} & \langle \bar{j}_2, -\eta_1 (K_3 + K_4) \bar{j}_2 \rangle_{S_2} \\ \langle \bar{j}_2, \eta_1 (K_3 + K_4) \bar{j}_2 \rangle_{S_2} & \left\langle \bar{j}_2, \eta_1^2 \left(\frac{L_3}{\eta_2^2} + \frac{L_4}{\eta_3^2} \right) \bar{j}_2 \right\rangle_{S_2} \end{pmatrix} \quad (9d)$$

Notice that there are symmetrical relationships between some of the sixteen submatrices in Eqs. (9a-d), for example, it can be proven that (Appendix A):

$$\langle \bar{j}_1, \eta_1 K_3 \bar{j}_2 \rangle_{S_1} = -\langle \bar{j}_2, -\eta_1 K_2 \bar{j}_1 \rangle_{S_2} \quad (10)$$

This leads to computing and storing elements of only nine out of the above sixteen submatrices, to gain significant reduction in CPU time and computer memory requirements for the MOM part.

To obtain the formulations for a PEC interior scatterer S_2 , we delete P rows and P columns from the total matrix $\bar{\bar{Z}}$ starting from index $(2N+P+1)$ up to index $(2N+2P)$. Moreover, P rows should be deleted from vectors \bar{I} and \bar{V} starting from index $(2N+P+1)$ up to index $(2N+2P)$. This leads to a new matrix $\bar{\bar{Z}}$ that has an order of $(2N+P) \times (2N+P)$.

The objective of the above analysis is to solve Eq. 8 for the unknown current coefficients I^1 and I^2 . Upon substituting the obtained coefficients into Eq. 7a and b, the surface electric and magnetic currents can be obtained on the exterior scatterer S_1 and on the interior scatterer S_2 . Thus complex vector electric fields in the near zone can be obtained [13].

As is well known, solving the system of equations in Eq. 8 by using the MOM requires computing, storing and multiplying a full dense matrix of order $2(N+P) \times 2(N+P)$ by a vector of order $2(N+P)$, which is computationally prohibitive for large-scale scatterers even using supercomputers. This necessitates the need for the SDFMM, which makes these computations significantly tractable. Applying the SDFMM to Eq. 8, the matrix $\bar{\bar{Z}}$ is divided as $\bar{\bar{Z}}^{NF} + \bar{\bar{Z}}^{FF}$, where $\bar{\bar{Z}}^{NF}$ represents the dense near field part, and $\bar{\bar{Z}}^{FF}$ represents the sparse far field part of matrix $\bar{\bar{Z}}$. The entries of $\bar{\bar{Z}}^{NF}$ are computed directly, and then the matrix vector multiply is conducted following the conventional MOM method. The entries of $\bar{\bar{Z}}^{FF}$ are not directly computed nor stored as before, but the matrix vector multiply is conducted in one step using the inhomogeneous plane wave expansions. The interaction decomposition into near field, NF , and far field, FF , is assumed according to the distance, in wavelength, between those interacting elements on the scatterer. We should mention that the computational complexity of the SDFMM for the CPU time and the memory requirement are of $O(K)$ per iteration of the iterative solver, while they are of order $O(K^2)$ for the MOM method, where K is the total number of the surface current unknowns and it is equal to $2(N+P)$ for the buried penetrable object case [1]-[6].

III. NUMERICAL RESULTS

In this work, the exterior closed scatterer S_1 is excited with an incident Gaussian beam that induces stronger surface currents on a localized area of the surface and weaker currents away from that area [14], [15]. As a result, these surface currents become negligible on the back of the closed surface S_1 , which makes it possible to approximate S_1 by an open rough surface with dimensions $L \times L$. In all results here, L is chosen to be $8\lambda_0$ where λ_0 is the free space wavelength. In reality, the rough surface is an infinite surface, but by using a carefully tapered incident Gaussian beam, the edge effect of the modeled finite surface can be eliminated. The random rough surface, characterized by Gaussian statistics for the random heights and for the autocorrelation function, is generated using the computer random number generator [16] and will be described by the root mean square height σ and the correlation length l_c . The interior scatterer S_2 is representing the object, which is always buried at the center of the ground with burial depth measured from its center to the mean plane of the rough surface.

In Figs. 3a and b, the SDFMM computer code is compared with the sparse matrix canonical grid (SMCG) method published in [14]. The buried object is assumed to be a PEC sphere of radius $r = 0.3\lambda_0$ at burial depth $d = 0.6\lambda_0$, where the incident angle is $\vartheta^i = 20^\circ$ and the relative dielectric constant of the soil is $\epsilon_r = 2.0 - i0.2$. The HH-polarization results for the normalized radar cross section (RCS) is shown for the rough surface ($\sigma = 0.02\lambda_0$ and $l_c = 0.5\lambda_0$) without the buried PEC sphere in Fig. 3a and with the buried PEC sphere in Fig. 3b. The horizontal axis in Fig. 3 is the scatter angle θ (elevation angle in degrees) measured from the z -axis. For comparison purpose with [14], the rough surface is sampled at 8 points per λ_0 leading to a number of surface

current unknowns $N=12,160$ as in [14]. The PEC sphere is sampled at 20 points in the ϕ -direction ($0-360^\circ$) and 10 points in the ϑ -direction ($0-90^\circ$) leading to a number of surface current unknowns $P=480$. The normalized RCS is calculated by taking the average of those values produced using ten azimuth angles. The excellent agreement shown in these figures validates the SDFMM code for the buried PEC object under the rough ground.

Next, new analysis of a buried penetrable or PEC oblate spheroid of dimensions $a=0.3\lambda_0$ and $b=0.15\lambda_0$, as shown in Fig. 1, is presented. To assure accuracy, the rough surface is sampled at 12.5 points per λ_0 with a number of surface current unknowns equal to $N=29,800$. The spheroid is sampled at 20 points in the ϕ -direction and 8 points in the ϑ -direction corresponding to $P=360$ surface current unknowns.

In Figs. 4a-d, the magnitude of the complex vector electric field in the near zone is calculated at height $z=0.5\lambda_0$ above and below the nominal ground and at (x, y) grid points of resolution equal to $0.1\lambda_0$. The ground surface is rough with $\sigma = 0.04\lambda_0$, $l_c = 0.5\lambda_0$, and relative dielectric constant $\epsilon_r = 2.5 - i0.18$ which is a typical Bosnian dry (3.8% moisture) clay loam at 1GHz [17]. To enhance target detection, the near scattered electric field is computed and averaged over multiple views each pointed to the same ground point directly over the target. Ten incident Gaussian beams with same elevation angle $\vartheta^i = 10^\circ$ and different azimuth angles $\phi^i = 0, 36, 72, \dots, 324^\circ$, are simulated as suggested by Zhang, Tsang and Pak [14]. The rationale of averaging over these multiple views is that this process minimizes the clutter effect of the rough surface on individual scattered beams. In Figs. 4a and 4b, the magnitude of the near electric field

scattered from the rough surface only is plotted for the H- and V- incident polarization, respectively, while in Figs. 4c and 4d, the near electric field transmitted into the same rough ground is plotted for the H- and V- polarizations, respectively. The shown results in Figs. 4a and 4b indicate the difference in the scattered electric fields upon changing the polarization of the incident Gaussian beam (H and V). Notice that in the H-polarization at oblique incident angle $\vartheta^i = 10^\circ$, the incident electric field always has a single component perpendicular to the plane of incidence (y-component for $\phi^i = 0, \pi$), while it has two components (x- and z- for $\phi^i = 0, \pi$) in the V-polarization. Accordingly, the local reflection and transmission coefficients on each point on the rough surface are different, giving rise to the difference observed between Figs. 4a and 4b and between Figs. 4c and 4d. As seen in Figs. 4a and 4b, the scattered electric fields for each polarization appear to have ten lobes, one every 36° in azimuth direction (as seen clearly at the 0.062 V/m contour level). For a perfectly flat surface, the scattered pattern would have ten-folded symmetry, but for the rough surface, the random variations distort this pattern. As expected, the transmitted electric fields shown in Figs. 4c and 4d are larger than the scattered electric fields shown in Figs. 4a and 4b due to the relatively small value of the dielectric constant ($\epsilon_r = 2.5 - i0.18$). Moreover, the results show that the electric fields transmitted into the ground are less distorted than the scattered electric fields. This is because the local transmitted angle is smaller than the local incident angle.

The near electric field scattered from the oblate PEC spheroid buried at depth $d = 0.3\lambda_o$ is calculated for the H- and V-incident polarizations. The results obtained are only slightly different from those presented in Figs. 4a and 4b, therefore they are not presented here. The near electric field scattered due just to the object, obtained by subtracting the electric fields scattered from the

surface only (Figs. 4a and 4b) from the electric fields scattered from the surface with the buried object, is shown in Figs. 5a and 5b for H- and V-polarization, respectively. The results indicate that the signature of the buried oblate PEC spheroid is smaller than the ground scattered field, since the target is small compared with the wavelength, and is a polarization dependent, as expected, with almost four times larger magnitude V-polarization than H-polarization.

The oblate penetrable spheroid with the relative dielectric constant of TNT $\epsilon_r=2.9-i0.07$, is buried under the same rough surface as before at the same depth $d=0.3\lambda_0$. The near electric scattered fields due just to the oblate penetrable spheroid are shown in Figs. 6a and 6b for H- and V- polarization, respectively. These results indicate that the signature of the oblate penetrable spheroid is not as polarization dependent as that of the oblate PEC spheroid shown in Figs. 5a and 5b. Since the Gaussian beams on opposite sides of the object are oppositely polarized, electric fields scattered from the centered object tend to cancel. It is interestingly to note that the rough ground surface perturbs this destructive interference. The average scattered electric field from the PEC object illuminated by V-polarized beams (Fig. 5b) is greater than for the H-polarized beams (Fig. 5a), due to constructive interference of the z-components. If a progressive phase of $2\pi/10$ is successively applied to each of the scattered electric fields to focus the field at the center, stronger but almost equal signatures of the PEC object are observed in both polarizations providing a stronger but less discriminative detecting technique.

The iterative solver TFQMR [18] has been used with tolerance 1% for all obtained results. Five levels in the fast multilevel multipole part of the SDFMM with finest block size equal to $0.5025\lambda_0$ were employed here. For the oblate penetrable spheroid case, the SDFMM code

needed two CPU hours, using one processor on Compaq GS140 EV6 machine, for the iterative solver to converge for each incident beam and additional thirty-five minutes for the set up time. Thus the total required CPU time, for ten multiple views, is 20.6 CPU hours. The required computer memory is 800MB. For the oblate PEC buried spheroid, the SDFMM code needed 1.6 CPU hours per view for the iterative solver and almost the same set up time as before leading to total time of 16.5 CPU hours in this case. The required computer memory in later case was 600MB. For the rough surface only, the SDFMM code needed 0.9 CPU hours for the iterative solver, half an hour for set up time, and 500MB computer memory.

IV. CONCLUSIONS

In this paper, a study has been presented for 3-D electromagnetic scattering from shallow objects buried under the random rough surface. Our analysis is not limited to any particular incident waves, transmitter position, near or far field zone, object's geometry, nor to the Gaussian statistics characterization of the rough surface. There is a limitation, however, on the rough surface root mean square height and on the burial depth of the object due to the constraint of the SDFMM to quasi-planar structures. In the current application, this is not a major limitation since the surface root mean square height and the burial depth of the object are comparable to the free space wavelength as a typical case for antipersonnel mine detection application. The ultimate objective of our research is to distinguish between the target signal scattered from the buried object and the clutter signal scattered from the rough ground. As our numerical results show, only a slight difference can be observed between these two signals because of the small size and dielectric contrast of the buried target with respect to the surrounding soil. Even though subtracting the two signals could clearly lead to the target signal, it is not a practical detection

method. With the insight gained from the numerical simulations, we aim to derive methods to remove the clutter signal to achieve target detection with minimum "false alarms". Ongoing research using these numerical simulations tests advanced and robust signal processing techniques and provides a basis to optimize transmitter/receiver positions for best target detection. Interestingly, our results show that the scattered electric fields above the ground undergoes strong distortion compared with the electric fields transmitted into the ground. Moreover, employing multiple views of the target using ten simultaneous incident Gaussian beams has lead to strong polarization dependency only for the PEC object, which can be used in target discrimination.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. W. Chew and Prof. E. Michielssen at UIUC for allowing them the use and modification of the SDFMM computer code for the current application. This work benefited from the allocation of time at the Northeastern University Advanced Scientific Computation Center (NU-ASCC). This research was sponsored by the Army Research Office Demining MURI grant # DAA 0-55-97-0013 and in part by the Engineering Research Centers Program of the NSF under award number EEC-9986821.

Appendix A

The LHS of Eq. 10 can be written as

$$\begin{aligned}\langle \bar{j}_1, \eta_1 K_3 \bar{j}_2 \rangle_{S_1} &= \eta_1 \int_{S_1} \bar{j}_1(\bar{r}) \cdot \left(\int_{S_2} \bar{j}_2(\bar{r}') \times \nabla \Phi_2(|\bar{r} - \bar{r}'|) ds' \right) ds \\ &= -\eta_1 \int_{S_1} \int_{S_2} \bar{j}_2(\bar{r}) \cdot (\bar{j}_1(\bar{r}') \times \nabla \Phi_2(|\bar{r} - \bar{r}'|)) ds' ds\end{aligned}\quad (A1)$$

where \bar{r}' is the vector position on S_2 and \bar{r} is the vector position on S_1 . From the RHS of Eq.

10, with using $\nabla \Phi = -\nabla' \Phi$, we get

$$\begin{aligned}-\langle \bar{j}_2, -\eta_1 K_2 \bar{j}_1 \rangle_{S_2} &= \eta_1 \int_{S_2} \bar{j}_2(\bar{r}) \cdot \left(\int_{S_1} \bar{j}_1(\bar{r}') \times \nabla \Phi_2(|\bar{r} - \bar{r}'|) ds' \right) ds \\ &= -\eta_1 \int_{S_2} \int_{S_1} \bar{j}_2(\bar{r}) \cdot (\bar{j}_1(\bar{r}') \times \nabla' \Phi_2(|\bar{r} - \bar{r}'|)) ds' ds\end{aligned}\quad (A2)$$

where \bar{r}' is the vector position on S_1 and \bar{r} is the vector position on S_2 . Upon renaming \bar{r}' by \bar{r} , and vice versa, and exchanging the order of integrations in Eq. A2, we get

$$\begin{aligned}-\langle \bar{j}_2, -\eta_1 K_2 \bar{j}_1 \rangle_{S_2} &= -\eta_1 \int_{S_2} \int_{S_1} \bar{j}_2(\bar{r}') \cdot (\bar{j}_1(\bar{r}) \times \nabla \Phi_2(|\bar{r} - \bar{r}'|)) ds ds' \\ &= -\eta_1 \int_{S_1} \int_{S_2} \bar{j}_2(\bar{r}') \cdot (\bar{j}_1(\bar{r}) \times \nabla \Phi_2(|\bar{r} - \bar{r}'|)) ds' ds\end{aligned}\quad (A3)$$

Notice that Eq. A3 is equivalent to Eq. A1, which proves Eq. 10. Thus the only nine submatrices computed and stored here are: three submatrices in \bar{Z}_{11} , three submatrices in \bar{Z}_{22} , two submatrices in \bar{Z}_{12} , and one submatrix in \bar{Z}_{21} .

REFERENCES

- [1] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [2] V. Jandhyala, E. Michielssen, B. Shanker and W.C. Chew, "A combined steepest descent-fast multipole algorithm for the fast analysis of three-dimensional scattering by rough surfaces," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, no. 3, pp.738-748, May 1998.
- [3] V. Jandhyala, B. Shanker, E. Michielssen and W. C. Chew, "A fast algorithm for the analysis of scattering by dielectric rough surfaces," *J. Opt. Soc. Am. A*, vol. 15, no. 7, pp. 1877-1885, July 1998.
- [4] M. El-Shenawee, V. Jandhyala, E. Michielssen and W. C. Chew, "An enhanced steepest descent fast multipole method for the analysis of scattering from two dimensional multilayered rough surfaces," in *Proc. of the IEEE APS/URSI'98 Meeting*, Atlanta, June 1998, p. 182.
- [5] M. El-Shenawee, V. Jandhyala, E. Michielssen and W. C. Chew, "The steepest descent fast multipole method (SDFMM) for solving combined field integral equation pertinent to rough surface scattering," in *Proc. of the IEEE APS/URSI'99*, 1999, pp. 534-537.
- [6] W. C. Chew, M. El-Shenawee, V. Jandhyala and E. Michielssen, "Scattering at low grazing angles from large scale two dimensional random rough surfaces using the steepest descent fast multipole method (SDFMM)," in *Proc. of the XXVIth General Assembly*, 1999, p. 195.
- [7] L. Medgyesi-Mitschang, J. Putnam and M. Gedera, "Generalized method of moments for three-dimensional penetrable scatterers," *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1383-1398, April 1994.

- [8] S. M. Rao, D. R. Wilton and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. Antennas Propagat.*, vol. AP 30, no. 3, pp.409-418, May 1982.
- [9] A. Poggio and E. Miller, "*Integral equation solutions of three-dimensional scattering problems*," in *Computer Techniques for Electromagnetics*, R. Mittra, ed., 2nd ed., Hemisphere, New York, 1987, pp. 159-264.
- [10] J. Stratton, *Electromagnetic theory*, McGraw-Hill, New York, ch. 8, pp.434-481.
- [11] J. R. Mautz and R. F. Harrington, "H-field, E-field, and combined field solutions for conducting bodies of revolutions," *Arch. Elek. Ubertrag.*, vol. 32, pp.157-164, 1978.
- [12] P. L. Huddleston, L. N. Medgyesi-Mitschang and J. M. Putnam, "Combined field integral equation formulation for scattering from dielectrically coated conducting bodies," *IEEE Trans. Antennas Propagat.*, vol. Ap-34, pp. 510-520, 1986.
- [13] C. A. Balanis, *Advanced engineering electromagnetics*, John Wiley & Sons Inc, ch. 6, pp.254-309.
- [14] G. Zhang, L. Tsang, and K. Pak, "Angular correlation function and scattering coefficient of electromagnetic waves scattered by a buried object under a two-dimensional rough surface," *J. Opt. Soc. Am A.*, pp. 2995-3002, December 1998.
- [15] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: Enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.
- [16] N. Garcia and E. Stoll, "Monte Carlo calculation for electromagnetic-wave scattering from random rough surfaces," *Phy. Rev. Let.*, vol. 52, no. 20, pp. 1798-1801, May 1984.

- [17] J. Curtis, "Dielectric properties of soils; various sites in Bosnia," US Army Corp. of Eng., Waterways Experim., Station Data Rep., 1996.
- [18] R. W. Freund, "A Transpose-free quasi-minimal residual algorithm for non-hermitian linear systems," *SIAM J. Sci. Comput.*, vol. 14, no. 2, pp. 470-482, March 1993.

List of Figures

Figure1. Cross section of 3-D object buried under the rough surface ground. The object is modeled as oblate spheroid as: $x = a \sin(\vartheta) \cos(\phi)$, $y = a \sin(\vartheta) \sin(\phi)$, and $z = b \cos(\vartheta)$.

Figure 2. (a) General penetrable 3-D scatterers R_2 and R_3 , (b) Equivalent problem exterior to S_1 , (c) Equivalent problem interior to S_1 and exterior to S_2 , (d) Equivalent problem interior to S_2 .

Figure 3a. Normalized RCS for rough surface of root mean square height $\sigma=0.02\lambda_0$ and correlation length $l_c=0.5\lambda_0$, at incident angle $\vartheta^i=20^\circ$. HH- polarization.

Figure 3b. Normalized RCS for a PEC sphere buried under rough surface of root mean square height $\sigma=0.02\lambda_0$ and correlation length $l_c=0.5\lambda_0$, with burial depth (from the center) $d=0.6\lambda_0$, and sphere radius $r = 0.3\lambda_0$, at incident angle $\vartheta^i=20^\circ$. HH- polarization.

Figure 4a. Near scattered electric field at $z = 0.5\lambda_0$ for H- incident polarization from rough surface of root mean square height $\sigma = 0.04\lambda_0$ and correlation length $l_c=0.5\lambda_0$, at incident angle $\vartheta^i=10^\circ$.

Figure 4b. Near scattered electric field at $z = 0.5\lambda_0$ for V- incident polarization from rough surface of root mean square height $\sigma = 0.04\lambda_0$ and correlation length $l_c=0.5\lambda_0$, at incident angle $\vartheta^i=10^\circ$.

Figure 4c. Near transmitted electric field at $z = -0.5\lambda_0$ for H- incident polarization from rough surface of root mean square height $\sigma = 0.04\lambda_0$ and correlation length $l_c=0.5\lambda_0$, at incident angle $\vartheta^i=10^\circ$.

Figure 4d. Near transmitted electric field at $z = -0.5\lambda_0$ for V- incident polarization from rough surface of root mean square height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$, at incident angle $\vartheta^i = 10^\circ$.

Figure 5a. Near scattered electric field at $z = 0.5\lambda_0$ for H- incident polarization due only to the PEC spheroid (obtained by subtraction).

Figure 5b. Near scattered electric field at $z = 0.5\lambda_0$ for V- incident polarization due only to the PEC spheroid (obtained by subtraction).

Figure 6a. Near scattered electric field at $z = 0.5\lambda_0$ for H- incident polarization due only to the penetrable spheroid (obtained by subtraction).

Figure 6b. Near scattered electric field at $z = 0.5\lambda_0$ for V- incident polarization due only to the penetrable spheroid (obtained by subtraction).

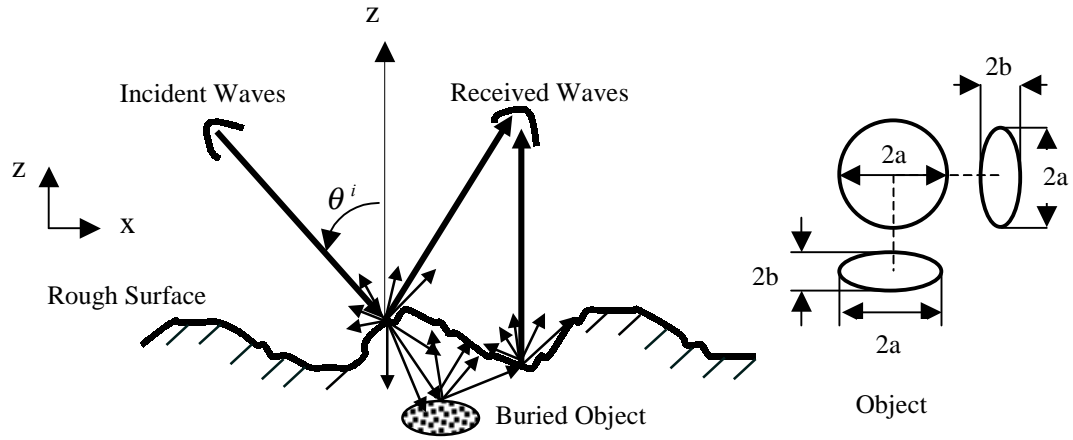
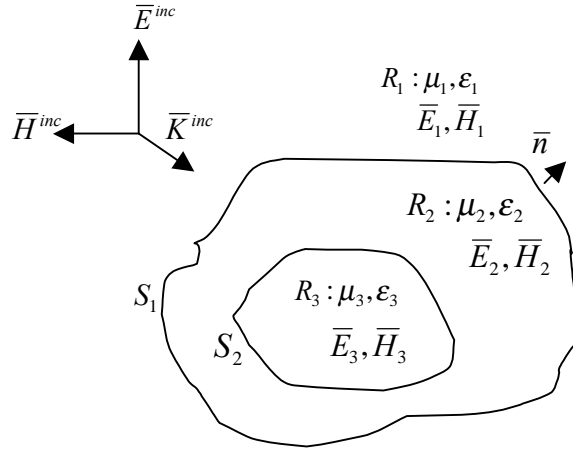
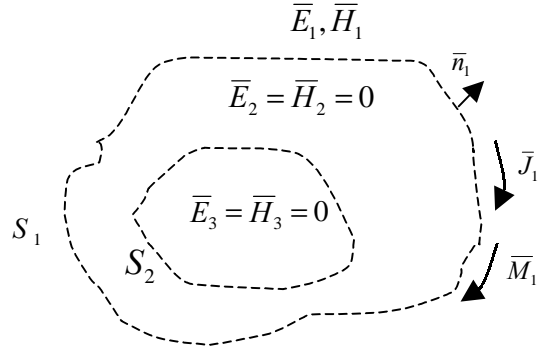


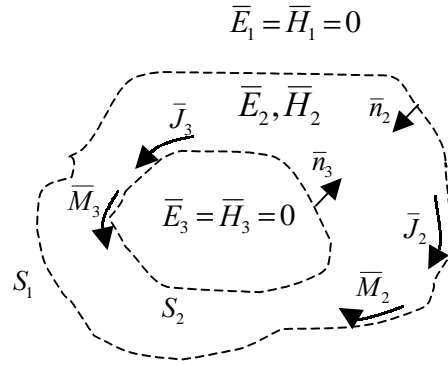
Figure1. Cross section of 3-D object buried under the rough surface ground. The object is modeled as oblate spheroid as: $x = a \sin(\vartheta) \cos(\phi)$, $y = a \sin(\vartheta) \sin(\phi)$, and $z = b \cos(\vartheta)$.



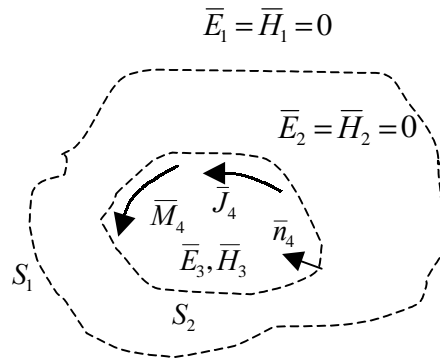
(a) General penetrable 3-D scatterers R_2 and R_3 .



(b) Equivalent problem exterior to S_1 .



(c) Equivalent problem interior to S_1 and exterior to S_2 .



(d) Equivalent problem interior to S_2

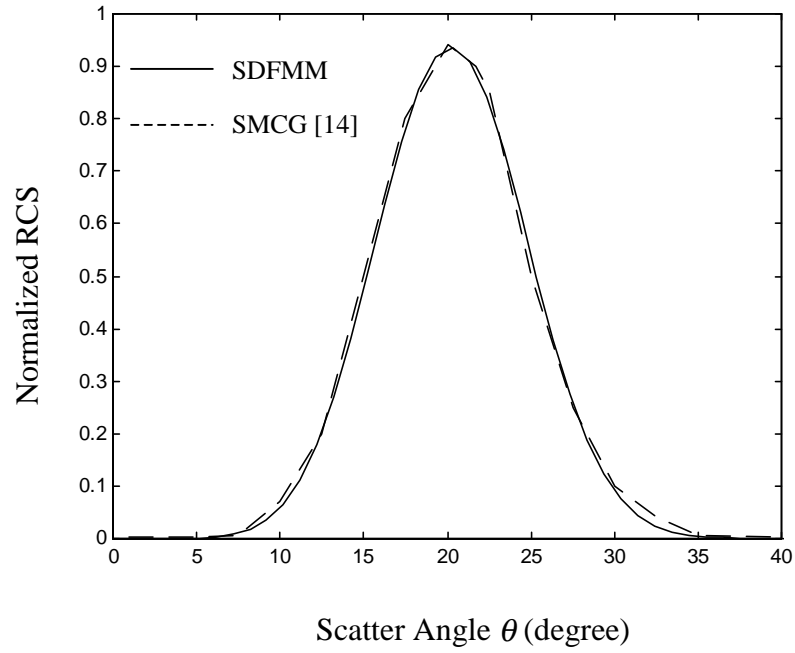


Figure 3a. Normalized RCS for a rough surface of root mean square height $\sigma=0.02\lambda_0$ and correlation length $l_c=0.5\lambda_0$, at incident angle $\vartheta^i=20^\circ$. HH polarization.

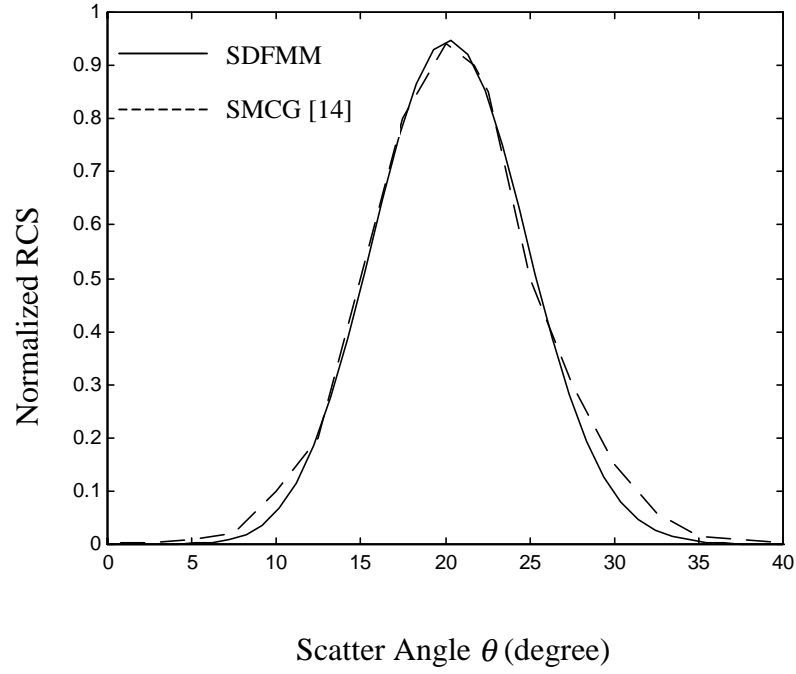


Figure 3b. Normalized RCS for a PEC sphere buried under a rough surface of root mean square height $\sigma=0.02\lambda_0$ and correlation length $l_c=0.5\lambda_0$, with burial depth (from the center) $d=0.6\lambda_0$, and sphere radius $r=0.3\lambda_0$, at incident angle $\vartheta^i=20^\circ$. HH polarization.

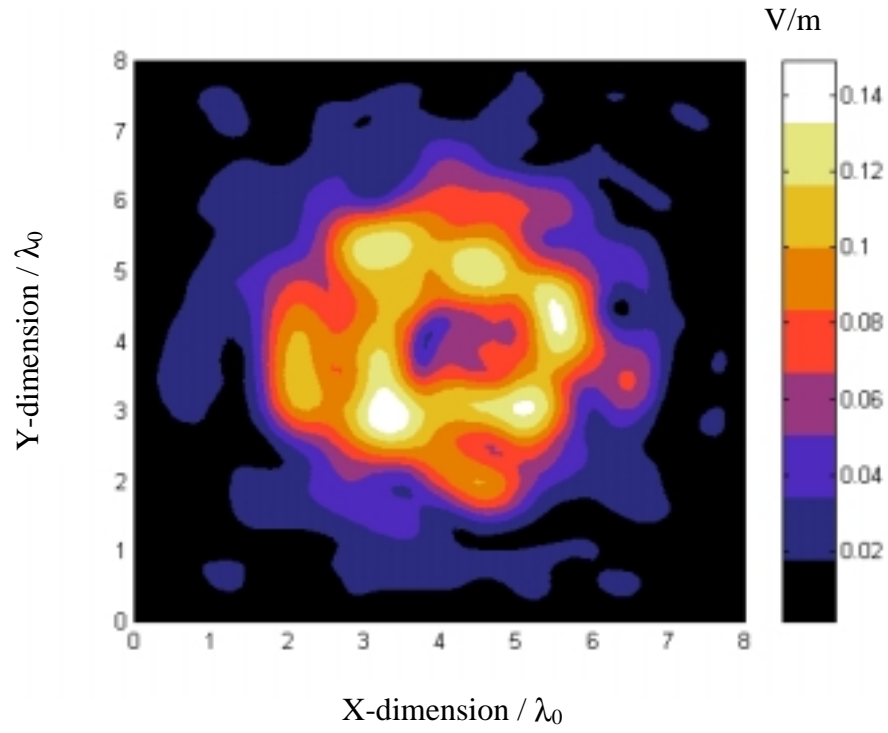


Figure 4a. Near scattered electric field at $z = 0.5\lambda_0$ for H- incident polarization from a rough surface of root mean square height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$, at incident angle $\vartheta^i = 10^\circ$.

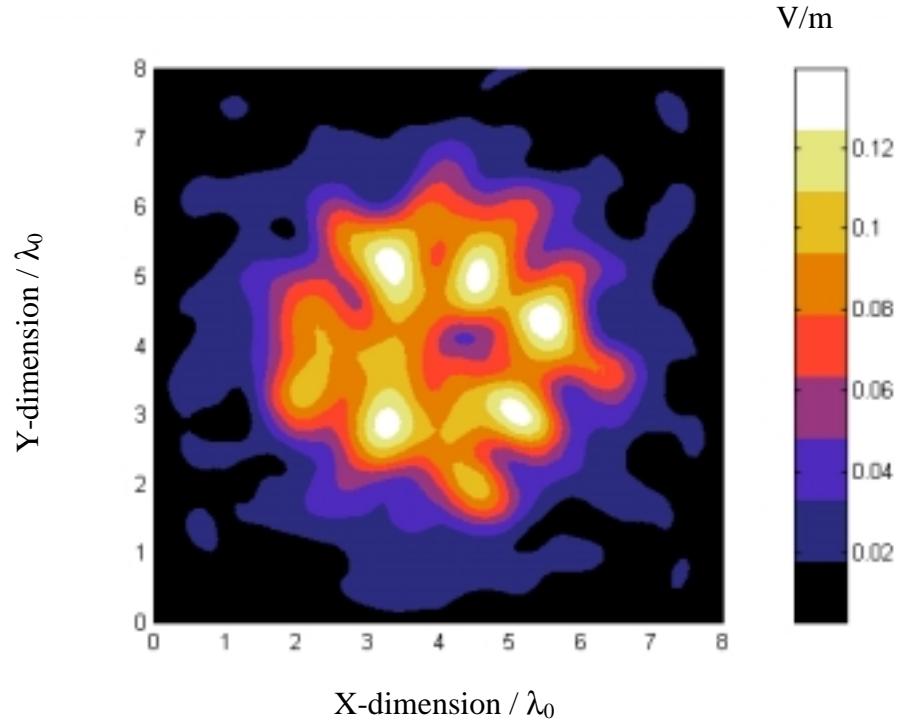


Figure 4b. Near scattered electric field at $z = 0.5 \lambda_0$ for V- incident polarization from a rough surface of root mean square height $\sigma = 0.04 \lambda_0$ and correlation length $l_c = 0.5 \lambda_0$, at incident angle $\vartheta^i = 10^\circ$.

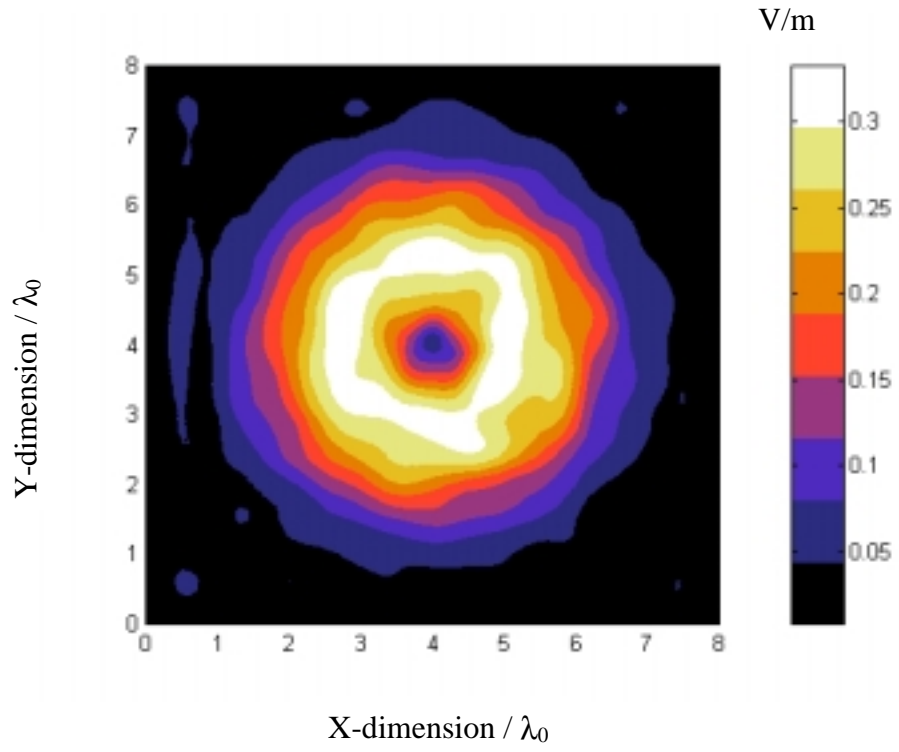


Figure 4c. Near transmitted electric field at $z = -0.5\lambda_0$ for H- incident polarization from a rough surface of root mean square height $\sigma = 0.04\lambda_0$ and correlation length $l_c = 0.5\lambda_0$, at incident angle $\vartheta^i = 10^\circ$.

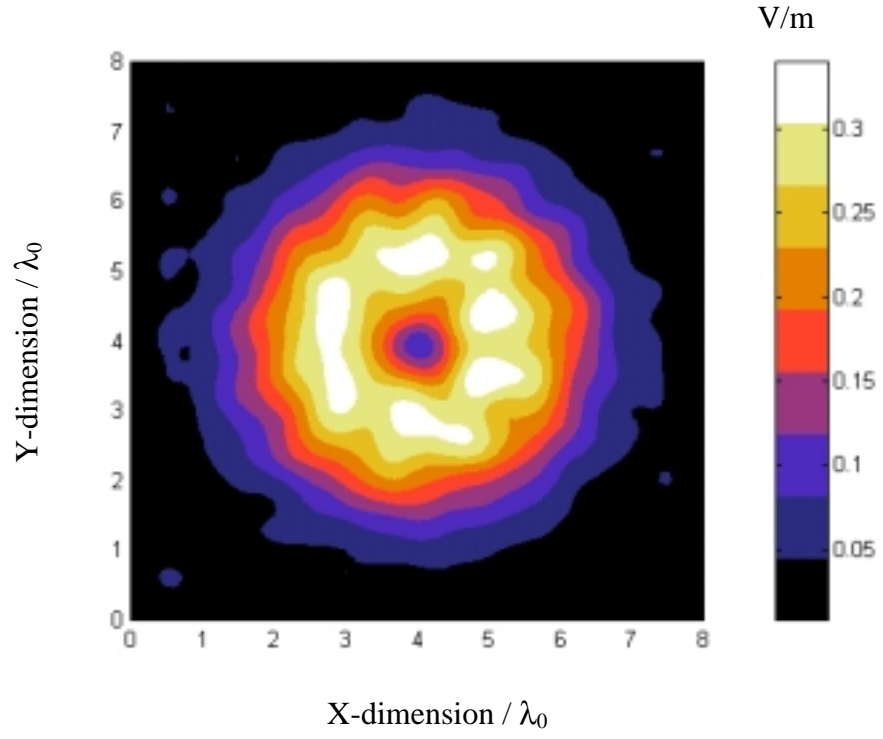


Figure 4d. Near transmitted electric field at $z = -0.5 \lambda_0$ for V- incident polarization from a rough surface of root mean square height $\sigma = 0.04 \lambda_0$ and correlation length $l_c = 0.5 \lambda_0$, at incident angle $\vartheta^i = 10^\circ$.

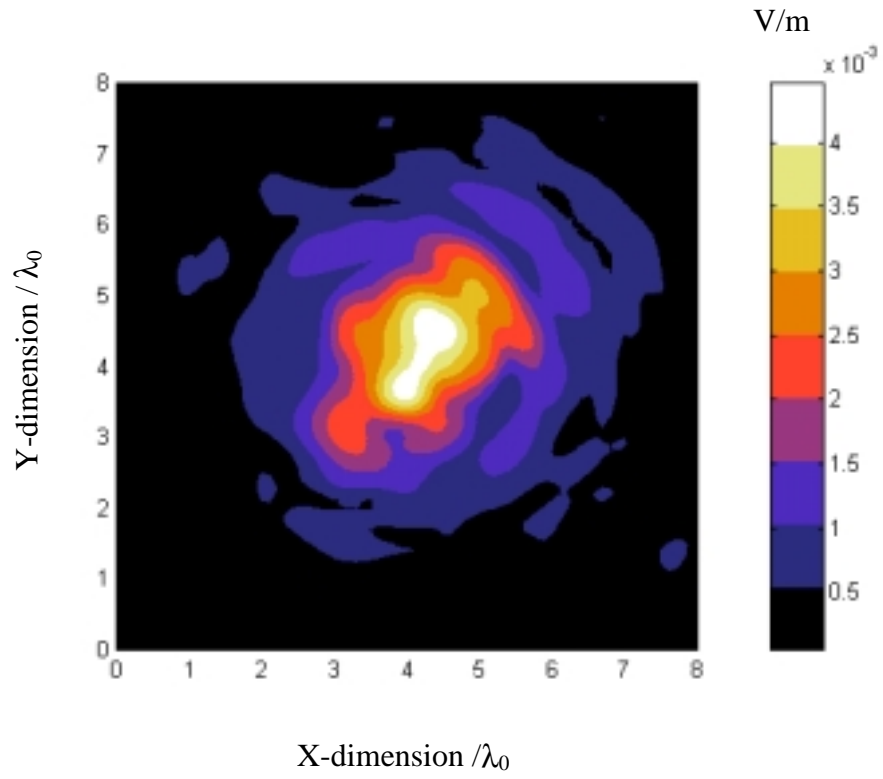


Figure 5a. Near scattered electric field at $z = 0.5\lambda_0$ for H- incident polarization due only to the PEC spheroid (obtained by subtraction).

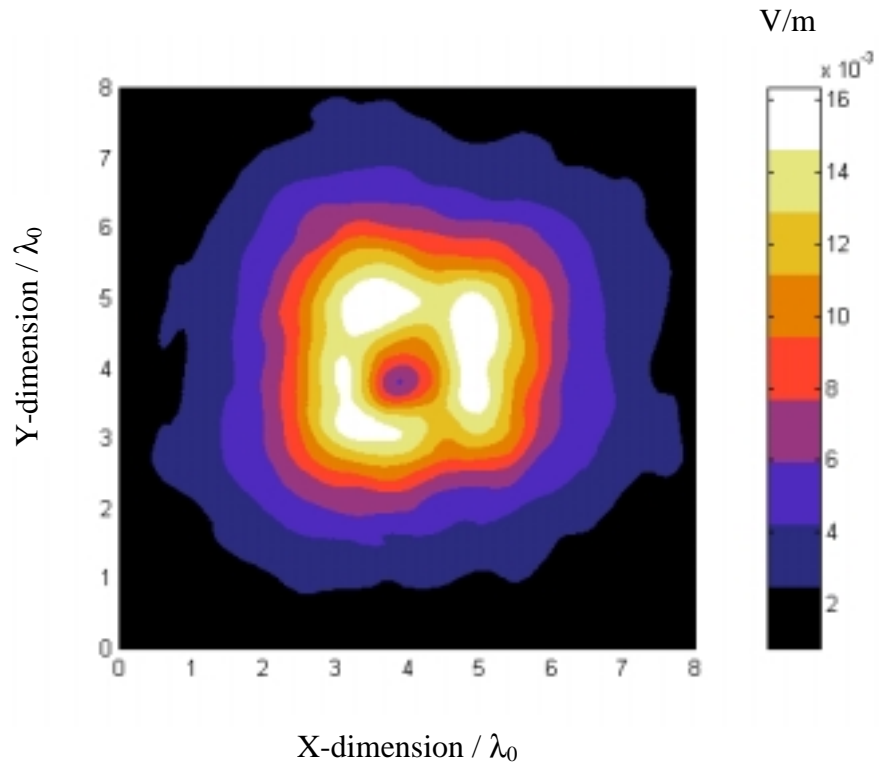


Figure 5b. Near scattered electric field at $z = 0.5\lambda_0$ for V- incident polarization due only to the PEC spheroid (obtained by subtraction).

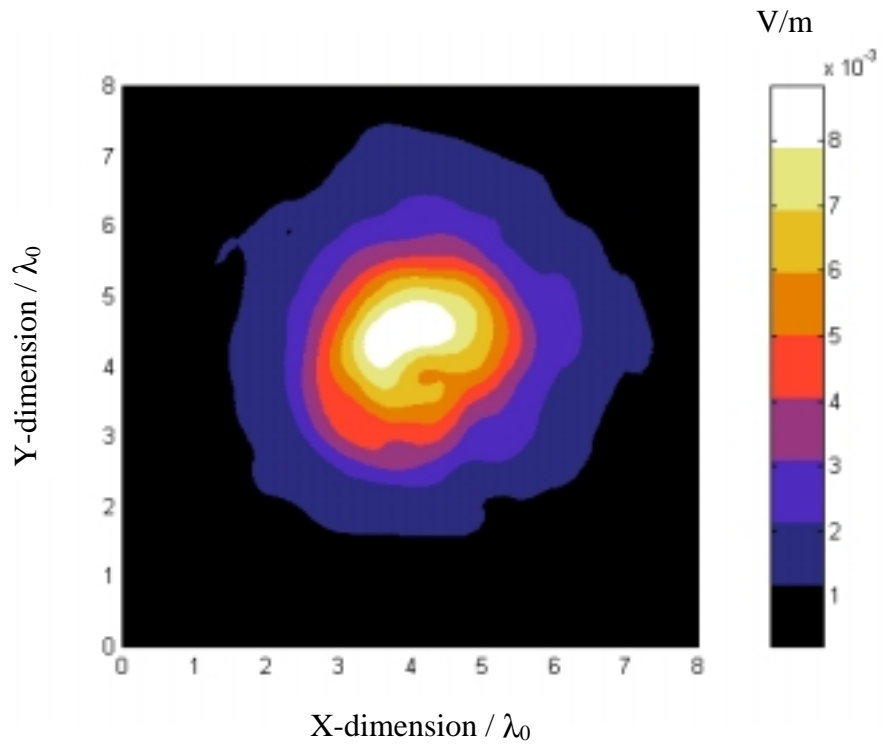


Figure 6a. Near scattered electric field at $z = 0.5\lambda_0$ for H- incident polarization due only to the penetrable spheroid (obtained by subtraction).

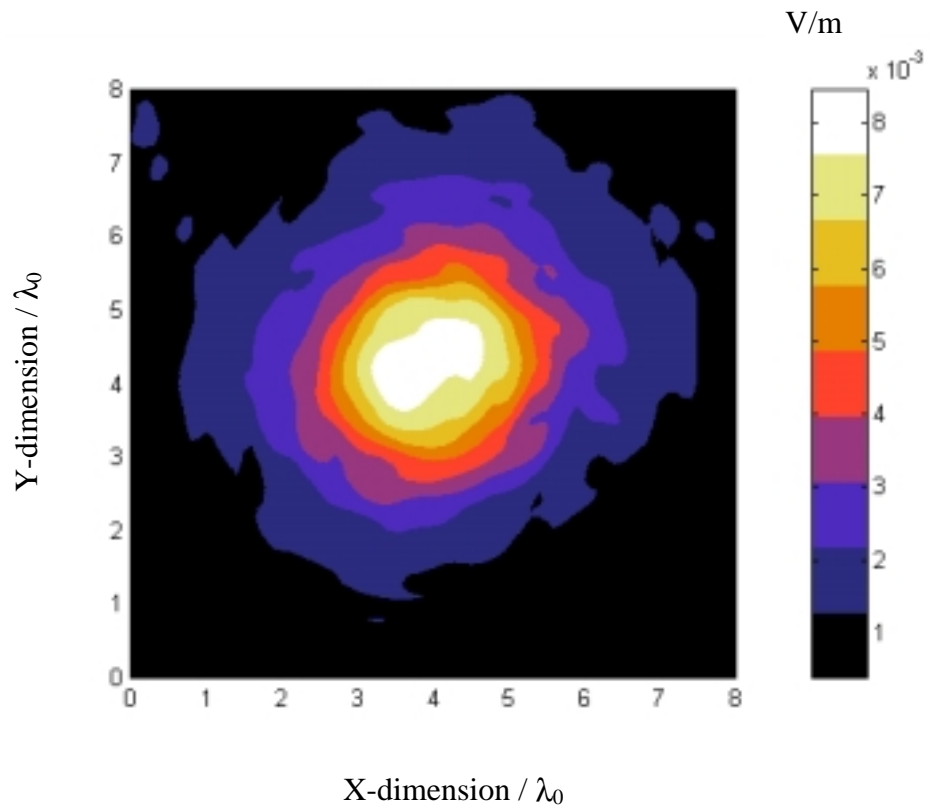


Figure 6b. Near scattered electric field at $z = 0.5 \lambda_0$ for V- incident polarization due only to the penetrable spheroid (obtained by subtraction).

The Steepest Descent Fast Multipole Method for Scattering from Two Penetrable Shallow Objects Buried Under Two-Dimensional Randomly Rough Surface

Magda El-Shenawee

Department of Electrical Engineering
3217 Bell Engineering Center
University of Arkansas
Fayetteville, Arkansas 72701
Tel: (501)-575-6582, Fax: (501)-575-7967
magda@uark.edu

ABSTRACT

New formulations are presented to analyze the near electric field scattered from two penetrable shallow objects buried under two-dimensional randomly rough surface. These two objects could have different materials, shapes and/or orientations; in addition their separation distance may range from a fraction of a wavelength to several wavelengths. The fast algorithm, Steepest Descent Fast Multipole Method (SDFMM), is used to compute the unknown electric and magnetic surface currents on the rough air-ground interface and on both the buried objects. Parametric investigation is presented to study the effect of the objects proximity, orientations, materials, shapes, the polarization of the incident waves, and the roughness of the ground on the scattered fields. A significant interference is observed between the two objects when they are separated by less than one free space wavelength. Moreover, even when the clutter due to the rough ground is removed, the return from the second object, can be dominating causing a possible false alarm in detecting the target. However, when all sources of clutter are removed, by removing the return from both the rough ground and the second object, the signature of the target can be clearly observed and analyzed. Our results show that the observed distortion in the target signature significantly increases with the increase of the ground roughness.

Keywords: Rough surface scattering, computational electromagnetics, buried objects.

INTRODUCTION

Several research works, in the frequency and/or in the time domain, have been published in the area of sensing objects buried under and/or above the Earth's surface. Some of the published work is related to objects buried in random media to simulate the clutter (no presence of a rough or planar interface) [1], in other work, the air-ground interface has been completely ignored justified by the very nearby position of the transmitting/receiving antennas to the ground [2], and in other several publications the roughness of the Earth's surface has been completely ignored by assuming a planar half space, where the objects are either buried under the interface [3]-[9], partially immersed in a semi-infinite dielectric medium [10] or located above the half space [11]. Interesting radar images for targets buried under planar and or rough ground are published in [12]-[15]. It is known that the roughness of the ground is a major source of clutter for target detection, however, due to its complicated analysis, several researchers have assumed that the interface is rough only in one-dimension (i.e., 2-D scattering problem) [16]-[23] where there was either a PEC (perfectly electric conducting) or a penetrable (dielectric) cylindrical target buried under the rough interface. In reality, the Earth's surface is rough in two-dimensions, which added more complexity to the analysis of the problem as in [24] where the buried target is a PEC sphere, in [25]-[26] where the object is located above the interface, and in [27]-[28] where the buried target is a penetrable and/or PEC spheroid. Recent experimental work is conducted [29] to discriminate landmines from various clutter items using EMI (electromagnetic induction). Even though, the author in [29] has concentrated his investigations on using the EMI, however the concept of discrimination between targets (not only landmines) and clutter items is the motivation of this work, which can benefit several electromagnetic subsurface sensing applications. For example, but not limited to, detection of cancerous tumors in the human breast,

plastic anti-personnel or metallic anti-tank mines, or the location of hazardous environmental waste, etc. In reality, these targets may be located next to other objects, either a second target or a clutter item. In this work, the emphasis will be on sensing two dielectric objects buried under a two-dimensional random rough ground, see Fig. 1. To the best of our knowledge, this application has not been analyzed before due its computational complexity that lead to excessive computational requirements. Therefore the Steepest Descent Fast Multipole Method (SDFMM), originally was developed at the University of Illinois at Urbana-Champaign [30]-[33], is used in this work to make these computations tractable. The new mathematical formulations are derived in Section II, numerical results are presented in Section III and concluding remarks are given in Section IV. More details are given in Appendices A and B.

II. FORMULATION

The electromagnetic waves are used to excite electric and magnetic surface currents on an exterior surface (e.g. rough ground) and consequently, due to wave penetration into the surface, currents are excited on the two interior surfaces (e.g. buried objects). The equivalence theorem is applied to express the electric and the magnetic fields as functions of surface electric and magnetic currents. There are four different regions involved in the current problem as R_1 , R_2 , R_3 and R_4 with permittivity and permeability given by ϵ_1 and μ_1 , ϵ_2 and μ_2 , ϵ_3 and μ_3 , and ϵ_4 and μ_4 , respectively, as shown in Fig.2a-e. Two interior scatterers of different materials R_3 and R_4 are immersed in the arbitrary penetrable 3-D scatterer R_2 , which is immersed in region R_1 as shown in Fig. 2a. The region R_2 is bounded by surfaces S_1 (exterior), and S_2 and S_3 (interiors) with normal vectors \bar{n}_1 and \bar{n}_2 on S_1 , \bar{n}_3 and \bar{n}_4 on S_2 , and \bar{n}_5 and \bar{n}_6 on S_3 , see Fig. 2a-e.

While regions R_3 and R_4 are bounded by surfaces S_2 and S_3 , respectively, as shown in Figs.

2d-e. The normal vectors (\bar{n}_1) , $(\bar{n}_2, \bar{n}_3$ and $\bar{n}_4)$, (\bar{n}_5) and (\bar{n}_6) are assumed to point into regions

R_1, R_2, R_3 , and R_4 respectively. The electric and magnetic fields in regions R_1, R_2, R_3 and R_4

are (\bar{E}_1, \bar{H}_1) , (\bar{E}_2, \bar{H}_2) , (\bar{E}_3, \bar{H}_3) and (\bar{E}_4, \bar{H}_4) which are due to equivalent currents (\bar{J}_1, \bar{M}_1) ,

$(\bar{J}_2, \bar{M}_2, \bar{J}_3, \bar{M}_3, \bar{J}_5$ and $\bar{M}_5)$, (\bar{J}_4, \bar{M}_4) and (\bar{J}_6, \bar{M}_6) , respectively, as shown in Figs. 2b-2e.

Assuming time dependence as $\exp(i\omega t)$, the electric and magnetic fields at an arbitrary point \bar{r}

in region R_1 are [27], [34]:

$$\theta_1(\bar{r})\bar{E}_1(\bar{r}) = \bar{E}^{inc}(\bar{r}) - \int_{S_1} (i\omega\mu_1(\bar{n}_1 \times \bar{H}_1)\Phi_1 - (\bar{n}_1 \times \bar{E}_1) \times \nabla' \Phi_1 - (\bar{n}_1 \cdot \bar{E}_1) \nabla' \Phi_1) ds' \quad (1a)$$

$$\theta_1(\bar{r})\bar{H}_1(\bar{r}) = \bar{H}^{inc}(\bar{r}) + \int_{S_1} (i\omega\epsilon_1(\bar{n}_1 \times \bar{E}_1)\Phi_1 + (\bar{n}_1 \times \bar{H}_1) \times \nabla' \Phi_1 + (\bar{n}_1 \cdot \bar{H}_1) \nabla' \Phi_1) ds' \quad (1b)$$

Similarly, the electric and magnetic fields in region R_2 are:

$$\begin{aligned} \theta_2(\bar{r})\bar{E}_2(\bar{r}) = & - \int_{S_1} (i\omega\mu_2(\bar{n}_2 \times \bar{H}_2)\Phi_2 - (\bar{n}_2 \times \bar{E}_2) \times \nabla' \Phi_2 - (\bar{n}_2 \cdot \bar{E}_2) \nabla' \Phi_2) ds' \\ & - \int_{S_2} (i\omega\mu_2(\bar{n}_3 \times \bar{H}_2)\Phi_2 - (\bar{n}_3 \times \bar{E}_2) \times \nabla' \Phi_2 - (\bar{n}_3 \cdot \bar{E}_2) \nabla' \Phi_2) ds' \\ & - \int_{S_3} (i\omega\mu_2(\bar{n}_5 \times \bar{H}_2)\Phi_2 - (\bar{n}_5 \times \bar{E}_2) \times \nabla' \Phi_2 - (\bar{n}_5 \cdot \bar{E}_2) \nabla' \Phi_2) ds' \end{aligned} \quad (1c)$$

$$\begin{aligned} \theta_2(\bar{r})\bar{H}_2(\bar{r}) = & \int_{S_1} (i\omega\epsilon_2(\bar{n}_2 \times \bar{E}_2)\Phi_2 + (\bar{n}_2 \times \bar{H}_2) \times \nabla' \Phi_2 + (\bar{n}_2 \cdot \bar{H}_2) \nabla' \Phi_2) ds' \\ & + \int_{S_2} (i\omega\epsilon_2(\bar{n}_3 \times \bar{E}_2)\Phi_2 + (\bar{n}_3 \times \bar{H}_2) \times \nabla' \Phi_2 + (\bar{n}_3 \cdot \bar{H}_2) \nabla' \Phi_2) ds' \\ & + \int_{S_3} (i\omega\epsilon_2(\bar{n}_5 \times \bar{E}_2)\Phi_2 + (\bar{n}_5 \times \bar{H}_2) \times \nabla' \Phi_2 + (\bar{n}_5 \cdot \bar{H}_2) \nabla' \Phi_2) ds' \end{aligned} \quad (1d)$$

Similarly, the electric and magnetic fields in region R_3 , are:

$$\theta_3(\bar{r})\bar{E}_3(\bar{r}) = - \int_{S_2} (i\omega\mu_3(\bar{n}_4 \times \bar{H}_3)\Phi_3 - (\bar{n}_4 \times \bar{E}_3) \times \nabla' \Phi_3 - (\bar{n}_4 \cdot \bar{E}_3) \nabla' \Phi_3) ds' \quad (1e)$$

$$\theta_3(\bar{r})\bar{H}_3(\bar{r}) = \int_{S_2} (i\omega\epsilon_3(\bar{n}_4 \times \bar{E}_3)\Phi_3 + (\bar{n}_4 \times \bar{H}_3) \times \nabla' \Phi_3 + (\bar{n}_4 \cdot \bar{H}_3) \nabla' \Phi_3) ds' \quad (1f)$$

Finally, the electric and magnetic fields in region R_4 , are:

$$\theta_4(\bar{r})\bar{E}_4(\bar{r}) = - \int_{S_3} (i\omega\mu_4(\bar{n}_4 \times \bar{H}_4)\Phi_4 - (\bar{n}_4 \times \bar{E}_4) \times \nabla' \Phi_4 - (\bar{n}_4 \cdot \bar{E}_4) \nabla' \Phi_4) ds' \quad (1g)$$

$$\theta_4(\bar{r})\bar{H}_4(\bar{r}) = \int_{S_3} (i\omega\epsilon_4(\bar{n}_6 \times \bar{E}_4)\Phi_4 + (\bar{n}_6 \times \bar{H}_4) \times \nabla' \Phi_4 + (\bar{n}_6 \cdot \bar{H}_4) \nabla' \Phi_4) ds' \quad (1h)$$

The 3-D scalar Green's function is given by $\Phi_i = \exp(-ik_i|\bar{r} - \bar{r}'|)/4\pi|\bar{r} - \bar{r}'|$, where \bar{r} is the field point, \bar{r}' is the source point, and $k_i = \omega\sqrt{\epsilon_i\mu_i}$ is the wave number in each region, $i=1, 2, 3$ and

4. The symbol $\theta_i(\bar{r})$ is the Heaviside function that gives the jump condition at surface S [34].

Applying the boundary conditions for the tangential electric and magnetic field components on surfaces S_1 , S_2 , and S_3 lead to continuity of equivalent surface currents as $(\bar{J}_1 = -\bar{J}_2$ and $\bar{M}_1 = -\bar{M}_2)$, $(\bar{J}_3 = -\bar{J}_4$ and $\bar{M}_3 = -\bar{M}_4)$ and $(\bar{J}_5 = -\bar{J}_6$ and $\bar{M}_5 = -\bar{M}_6)$, respectively.

Equating the tangential component of the electric and magnetic fields on surface S_1 , we get:

$$\bar{E}^{inc}(\bar{r})\big|_{\text{tang.}} = [(L_1 + L_2)\bar{J}_1 - (K_1 + K_2)\bar{M}_1 - L_3\bar{J}_3 + K_3\bar{M}_3 - L_4\bar{J}_5 + K_4\bar{M}_5]_{\text{tang.}} \quad (2)$$

$$\bar{H}^{inc}(\bar{r})\big|_{\text{tang.}} = \left[(K_1 + K_2)\bar{J}_1 + \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{M}_1 - K_3\bar{J}_3 - \frac{L_3}{\eta_2^2} \bar{M}_3 - K_4\bar{J}_5 - \frac{L_4}{\eta_2^2} \bar{M}_5 \right]_{\text{tang.}} \quad (3)$$

on surface S_2

$$0 = [-L_2\bar{J}_1 + K_2\bar{M}_1 + (L_3 + L_5)\bar{J}_3 - (K_3 + K_5)\bar{M}_3 + L_4\bar{J}_5 - K_4\bar{M}_5]_{\text{tang.}} \quad (4)$$

$$0 = \left[-K_2\bar{J}_1 - \frac{L_2}{\eta_2^2} \bar{M}_1 + (K_3 + K_5)\bar{J}_3 + \left(\frac{L_3}{\eta_2^2} + \frac{L_5}{\eta_3^2} \right) \bar{M}_3 + K_4\bar{J}_5 + \frac{L_4}{\eta_2^2} \bar{M}_5 \right]_{\text{tang.}} \quad (5)$$

on surface S_3

$$0 = \left[-L_2 \bar{J}_1 + K_2 \bar{M}_1 + L_3 \bar{J}_3 - K_3 \bar{M}_3 + (L_4 + L_6) \bar{J}_5 - (K_4 + K_6) \bar{M}_5 \right]_{\text{tang.}} \quad (6)$$

$$0 = \left[-K_2 \bar{J}_1 + \frac{L_2}{\eta_2^2} \bar{M}_1 + K_3 \bar{J}_3 + \frac{L_3}{\eta_2^2} \bar{M}_3 + (K_4 + K_6) \bar{J}_5 + \left(\frac{L_4}{\eta_2^2} + \frac{L_6}{\eta_4^2} \right) \bar{M}_5 \right]_{\text{tang.}} \quad (7)$$

in which the intrinsic impedance is $\eta_i = \sqrt{\mu_i / \epsilon_i}$, in region R_i with $i = 1, 2, \dots, 4$ and L_j and K_j , $j = 1, 2, \dots, 6$ are the integro-differential operators as given in Appendix A. The surfaces S_1, S_2 and S_3 are discretized into triangular patches where the unknown equivalent electric and magnetic currents in Eqs. 2-7 are approximated by using the RWG vector basis functions $\bar{j}(\bar{r})$ [35] as:

$$\bar{J}_i(\bar{r}) = \sum_{n=1}^{N_k} I_{in} \bar{j}_{kn}(\bar{r}), \quad \bar{M}_i(\bar{r}) = \eta_i \sum_{n=1}^{N_k} I_{(i+1)n} \bar{j}_{kn}(\bar{r}), \quad \bar{r} \in S_k, \quad \text{for } i = 1, 3 \text{ and } 5 \quad (8)$$

in which $k = 1$ for $i = 1$, $k = 2$ for $i = 3$ and $k = 3$ for $i = 5$. Upon substituting Eq. 8 into Eq. 2-7, normalizing the H -field equations by η_1 , and applying Galerkin's method for testing the E - and H -field equations, the linear system of equations is obtained:

$$\bar{\bar{Z}} \bar{I} = \bar{V} \Rightarrow \begin{pmatrix} \bar{Z}_{11} & \bar{Z}_{12} & \bar{Z}_{13} \\ \bar{Z}_{21} & \bar{Z}_{22} & \bar{Z}_{23} \\ \bar{Z}_{31} & \bar{Z}_{32} & \bar{Z}_{33} \end{pmatrix} \begin{pmatrix} I^1 \\ I^2 \\ I^3 \end{pmatrix} = \begin{pmatrix} \bar{V}_1 \\ 0 \\ 0 \end{pmatrix} \quad (9)$$

The submatrix \bar{Z}_{11} has order of $2N_1 \times 2N_1$ and represents interactions between elements only on surface S_1 , the submatrix \bar{Z}_{22} has order of $2N_2 \times 2N_2$ and represents interactions between elements only on surface S_2 , and the submatrix \bar{Z}_{33} has order of $2N_3 \times 2N_3$ and represents interactions between elements only on surface S_3 . The other submatrices in Eq. 9 represent interactions between different surfaces, e.g., \bar{Z}_{12} represents interactions between elements on surface S_1 and elements on surface S_2 , etc. The details of these submatrices are given in

Appendix B. This implies that the total matrix $\bar{\bar{Z}}$ has order of $2(N_1 + N_2 + N_3) \times 2(N_1 + N_2 + N_3)$ and it accounts for all interactions between different surfaces. The tested tangential incident electric field \bar{E}^{inc} and the tested normalized magnetic field $\eta_1 \bar{H}^{inc}$ on the exterior surface S_1 are represented by vector \bar{V}_1 . Solving the system of equations in Eq. 9 by using the MoM (Method of Moments) requires computing and storing the full dense matrix $\bar{\bar{Z}}$ and then multiplying it by the vector \bar{V} , which leads to a very computational expensive process. Therefore, the SDFMM is used to solve Eq. 9 leading to reduce the computational complexity for the CPU time and the computer memory to be only of $O(K)$ per iteration instead of $O(K^2)$ using the MoM [30]-[33], where $K = 2(N_1 + N_2 + N_3)$ is the total number of the current unknowns.

I. NUMERICAL RESULTS

In all results, the incident wave is assumed to be a Gaussian beam [36] that is carefully tapered to eliminate the effects of the artificial edges of the ground surface that has dimensions of $8\lambda_0 \times 8\lambda_0$ [27]. The half beam width of the Gaussian beam is assumed to be $1.6\lambda_0$ which illuminates the ground with plane waves over a spot of diameter $3.2\lambda_0$ centered at $x = 4\lambda_0, y = 4\lambda_0$ (same as ground center). The beam is incident normally towards the ground, with either horizontal or vertical polarization, i.e. the incident electric field will be in the y -direction or in the x -direction, respectively, see Fig. 1. Point receivers are located at height $0.5\lambda_0$ above the mean ground with resolution equal to $0.04\lambda_0$. The rough ground characterized with Gaussian random heights and Gaussian autocorrelation function such that the roughness will be described by the standard deviation of the height σ and the correlation length l_c , with assuming

zero mean height. The relative dielectric constant of the ground is assumed to be $\epsilon_r = 2.5 - j0.18$ [37] while the dielectric constant of each object will be stated for each in this section. The used discretization distance on the ground surface is $0.08\lambda_0$ leading to 60,000 electric and magnetic surface current unknowns, while using almost the same discretization rate for the two objects generates approximately 600 unknowns for each one.

In this section, images for the two penetrable objects that are buried under the ground will be presented. These images are based on the magnitude of the scattered electric fields which are received in the near zone above the ground [38], see Fig. 1. In Figs. 3 and 4, the ground is assumed to be flat in order to demonstrate the clutter due only to the second object (without the clutter due to presence of the rough ground). The scattered fields due to just the two objects are obtained by subtracting the ground returns from the total scattered fields in complex vector representation. Excellent agreement between the results of the SDFMM and those of the MoM is obtained which is used to validate the current computer code. Our parametric investigation includes studying the effect on the scattered fields due to proximity between the two objects, their materials, shapes and orientations, incidence polarization and ground roughness. The images shown in Figs. 3a-c are for two oblate spheroids both have dimensions $a = 0.15\lambda_0, b = 0.3\lambda_0$, material of $\epsilon_r = 2.9 - j0.072$ and buried at $z = -0.3\lambda_0$. They are located diagonally at $x = 4.5\lambda_0, y = 3.5\lambda_0$ and $x = 3.5\lambda_0, y = 4.5\lambda_0$ in Fig. 3a, at $x = 5\lambda_0, y = 3\lambda_0$ and $x = 3\lambda_0, y = 5\lambda_0$ in Fig. 3b, and at $x = 5.5\lambda_0, y = 2.5\lambda_0$ and $x = 2.5\lambda_0, y = 5.5\lambda_0$ in Fig. 3c. This implies that the separation distance from the center of the first object to the center of the second object is $D_x = D_y = 1\lambda_0, 2\lambda_0$, and $3\lambda_0$ in Fig. 3a, 3b and 3c, respectively, where D_x and D_y are the distance in the x - and y -direction. A significant interference between the two objects is observed in Fig. 3a where the separation distance is equal to one free space wavelength

($D_x = D_y = 1\lambda_0$). This coupling between the two objects decreases with the increase of the separation distance as shown in Fig. 3b and 3c. This implies that if these two objects are separated by three wavelengths of free space or more, then each object can be detected separately from the other one, e.g. by narrowing the width of the incident beam. Moreover, the results in Figs. 3a-c show the degradation of the signature of both objects as their positions are moved away from the center of the ground where the incident beam is centered, which agrees with our finding in [28].

Next, we change the shape of the objects, which consequently affects the scattered electric fields as shown in Figs. 4a-c. In these figures, the shape of one object is assumed to be the same spheroid as used in Fig. 3 while the second object is changed to be a sphere of radius $a = 0.15\lambda_0$ in Fig. 4a, an ellipsoid of dimensions $a = 0.15\lambda_0, b = 0.3\lambda_0, c = 0.15\lambda_0$ in Fig. 4b and a vertical cylinder of radius $a = 0.15\lambda_0$ and height $h = 0.15\lambda_0$ in Fig. 4c. The two objects in Fig. 4a-c are located similar to those in Fig. 3a where the separation distance between them (from center to center) is also $D_x = D_y = 1\lambda_0$. The scattered fields from just both the spheroid and the sphere, both the spheroid and the ellipsoid and both the spheroid and the cylinder are shown in Fig. 4a, 4b and 4c, respectively. The results show that the signature of the spheroid (located in the upper left quadrant) is much larger than the signature of the sphere, ellipsoid or the vertical cylinder (located in the lower right quadrant). However, the signature of the spheroid with the presence of the ellipsoid has slightly larger value than it with the presence of either the sphere or the vertical cylinder. Moreover, because the two objects have different shapes and sizes, the interference between them is clearly not symmetric, contrary to the results of Figs. 3a-c. In Fig. 3 and 4, the two objects are assumed to have the same material ($\epsilon_r = 2.9 - j0.072$) and they are buried under a flat ground. To simulate a more realistic case, different material for the second object will be

assumed with also incorporating the ground roughness in the model. In this case, the spheroid is assumed to have the same material and dimensions as in Fig. 3 and it will be located at $x = 4.5\lambda_0$, $y = 3.5\lambda_0$ and $z = -0.3\lambda_0$ (lower right quadrant). The second object is assumed to be a horizontal cylinder of radius $a = 0.15\lambda_0$, height $h = 0.9\lambda_0$ with its axis tilted at angle 30° with the x -direction (see Fig. 1). The material of this horizontal cylinder is assumed to be lossless with $\epsilon_r = 4$ and its center is located at $x = 4.01\lambda_0$, $y = 4.375\lambda_0$ and $z = -0.3\lambda_0$ (upper right quadrant) which implies that the separation distance will be $D_x = 0.49\lambda_0$ and $D_y = 0.875\lambda_0$. The rms height of the random rough ground is assumed to be $\sigma = 0.04\lambda_0$ with the correlation length of $l_c = 0.5\lambda_0$. The results in Fig. 5a and 5b are for the incident H- and V-polarization, respectively. These results show that the signature of the horizontal cylinder (upper left quadrant) is dominating the image with maximum amplitude of almost three times larger than that of the spheroid (lower right quadrant). This is due to the larger size and dielectric constant of the cylinder than those of the spheroid. Moreover, the tilted position of the horizontal cylinder could have contributed to this observation as well. Polarization dependency of the scattered electric fields is clearly observed upon comparing Fig. 5a (horizontal polarization) with Fig. 5b (vertical polarization), with slight increase in the maximum amplitude in the later case. Moreover, the results in Figs. 5a and 5b indicate that the presence of the cylinder (clutter item) can easily cause a false alarm in detecting the spheroid (target) in both polarizations. In order to analyze the image of the spheroid alone, all sources of clutter should be removed, i.e. removing the return from both the cylinder and the rough ground. The output, for incident H-polarization, is presented in Fig. 5c where a clear image of just the spheroid is shown. Similar results for the case of incident V-polarization is observed. Then, for the incident H-polarization, the roughness

of the ground is increased to be $\sigma = 0.1\lambda_0$ with the same correlation length as before where the image of the just spheroid is shown in Fig. 5d. All data in Fig. 5d are the same as in Fig. 5a except that the burial depth is increased to be $z = -0.4\lambda_0$ in this case. Comparing Fig. 5d with Fig. 5c, it is clear that increasing the ground roughness has increased the distortion in the image. The decrease in the amplitude in Fig. 5d compared with that in Fig. 5c could mainly be due to the larger burial depth of the spheroid in the later case that causes more attenuation in the transmitting waves. The same spheroid and cylinder used in Figs. 5a are buried under a flat ground instead of the rough ground as shown in Fig. 6a where the image of the spheroid alone is obtained also by removing the return from both the cylinder and the flat ground. Moreover, the same spheroid used in Fig. 5a is buried as a single object (no nearby buried cylinder) under the flat ground as shown in Fig. 6b where the image of the spheroid alone is obtained. Comparing Fig. 6a with Fig. 6b show the distortion due to only the presence of the nearby cylinder, while comparing Fig. 5c and 5d with Fig. 6a show the distortion due to the ground roughness. However, comparing Fig. 5c and 5d with Fig. 6b show the distortion due to the presence of both the rough ground and the nearby cylinder. Moreover, the image in Fig. 5c is slightly different from that in Fig. 6a due to the small roughness in this case but the image in Fig. 5d is more distorted due to the increase in ground roughness. In all previous results, the clutters are removed by subtraction using complex vector representation, but when the subtraction is processed using only the magnitude of the fields (i.e. ignoring the phase), the images of the spheroid looked more distorted as the ground roughness increases.

In all results, the SDFMM code required 850MB computer memory, 25 CPU minutes to fill the near field matrix, 23 CPU minutes to calculate the near scattered fields with $0.04\lambda_0$, and 4.6 CPU hours to obtain relative error equal to 10^{-5} for the TFQMR iterative solver.

CONCLUSIONS

In this work, we analyzed the 3-D scattering problem of the two dielectric shallow objects buried under the two-dimensional random rough ground. The fast algorithm, SDFMM, is used to solve the obtained linear system of equations for the electric and magnetic surface current unknown coefficients. Our parametric investigation results show the significant effect of the separation distance between the two objects and the ground roughness on the target signature. The numerical results show clear distortion due to the ground roughness and, more interestingly, they show the possible false alarm in detecting the target due to the presence of the second object (clutter item). The SDFMM has efficiently made the current 3-D computations tractable in an accurate and fast manner.

ACKNOWLEDGEMENTS

This research was sponsored from the Northeastern University's Demining MURI grant # DAA 0-55-97-0013 and in part by the College of Engineering at the University of Arkansas. The author would like to thank Prof. W. Chew, Prof. E. Michielssen at UIUC for allowing the use and modification of the SDFMM computer code for the current application. This work benefited from the allocation of time at the Northeastern University Advanced Scientific Computation Center (NU-ASCC).

Appendix A

With representing the surface electric and magnetic currents \bar{J} and \bar{M} on S_1 , S_2 and S_3 by the vector \bar{X} , the integro-differential operators L_j and $K_j, j=1, 2, \dots, 6$, are [27], [34]:

$$L_{1,2}\bar{X} = \int_{S_1} \left\{ i\omega\mu_{1,2}\Phi_{1,2}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{1,2}}\nabla\nabla' \cdot \bar{X}(\bar{r}')\Phi_{1,2} \right\} ds', \quad K_{1,2}\bar{X} = \int_{S_1} \bar{X}(\bar{r}') \times \nabla\Phi_{1,2} ds' \quad (A1)$$

$$L_{3,5}\bar{X} = \int_{S_2} \left\{ i\omega\mu_{2,3}\Phi_{2,3}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{2,3}}\nabla\nabla' \cdot \bar{X}(\bar{r}')\Phi_{2,3} \right\} ds', \quad K_{3,5}\bar{X} = \int_{S_2} \bar{X}(\bar{r}') \times \nabla\Phi_{2,3} ds' \quad (A2)$$

$$L_{4,6}\bar{X} = \int_{S_3} \left\{ i\omega\mu_{2,4}\Phi_{2,4}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{2,4}}\nabla\nabla' \cdot \bar{X}(\bar{r}')\Phi_{2,4} \right\} ds', \quad K_{4,6}\bar{X} = \int_{S_3} \bar{X}(\bar{r}') \times \nabla\Phi_{2,4} ds' \quad (A3)$$

Appendix B

The elements of the submatrix \bar{Z}_{11} in (9) are given by

$$\bar{Z}_{11} = \begin{pmatrix} \langle \bar{j}_1, (L_1 + L_2)\bar{j}_1 \rangle_{S_1} & \langle \bar{j}_1, -\eta_1(K_1 + K_2)\bar{j}_1 \rangle_{S_1} \\ \langle \bar{j}_1, \eta_1(K_1 + K_2)\bar{j}_1 \rangle_{S_1} & \left\langle \bar{j}_1, \eta_1^2 \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{j}_1 \right\rangle_{S_1} \end{pmatrix} \quad (B1)$$

in which $\langle \bar{A}, \bar{B} \rangle_S$ denotes the complex inner product between vector functions \bar{A} and \bar{B} on a

surface S [27], [34]. Similarly, the elements of the submatrices \bar{Z}_{22} and \bar{Z}_{33} are given by

$$\bar{Z}_{22} = \begin{pmatrix} \langle \bar{j}_2, (L_3 + L_5)\bar{j}_2 \rangle_{S_2} & \langle \bar{j}_2, -\eta_1(K_3 + K_5)\bar{j}_2 \rangle_{S_2} \\ \langle \bar{j}_2, \eta_1(K_3 + K_5)\bar{j}_2 \rangle_{S_2} & \left\langle \bar{j}_2, \eta_1^2 \left(\frac{L_3}{\eta_2^2} + \frac{L_5}{\eta_3^2} \right) \bar{j}_2 \right\rangle_{S_2} \end{pmatrix} \quad (B2)$$

$$\bar{Z}_{33} = \begin{pmatrix} \langle \bar{j}_3, (L_4 + L_6)\bar{j}_3 \rangle_{S_3} & \langle \bar{j}_3, -\eta_1(K_4 + K_6)\bar{j}_3 \rangle_{S_3} \\ \langle \bar{j}_3, \eta_1(K_4 + K_6)\bar{j}_3 \rangle_{S_3} & \left\langle \bar{j}_3, \eta_1^2 \left(\frac{L_4}{\eta_2^2} + \frac{L_6}{\eta_4^2} \right) \bar{j}_3 \right\rangle_{S_3} \end{pmatrix} \quad (B3)$$

While submatrices \bar{Z}_{12} , \bar{Z}_{13} , \bar{Z}_{21} , \bar{Z}_{31} , \bar{Z}_{23} and \bar{Z}_{32} are given by

$$\bar{Z}_{12} = \begin{pmatrix} \langle \bar{j}_1, -L_3 \bar{j}_2 \rangle_{s_1} & \langle \bar{j}_1, \eta_1 K_3 \bar{j}_2 \rangle_{s_1} \\ \langle \bar{j}_1, -\eta_1 K_3 \bar{j}_2 \rangle_{s_1} & \left\langle \bar{j}_1, -\eta_1^2 \left(\frac{L_3}{\eta_2^2} \right) \bar{j}_2 \right\rangle_{s_1} \end{pmatrix} \quad (\text{B4})$$

$$\bar{Z}_{21} = \begin{pmatrix} \langle \bar{j}_2, -L_2 \bar{j}_1 \rangle_{s_2} & \langle \bar{j}_2, \eta_1 K_2 \bar{j}_1 \rangle_{s_2} \\ \langle \bar{j}_2, -\eta_1 K_2 \bar{j}_1 \rangle_{s_2} & \left\langle \bar{j}_2, -\eta_1^2 \left(\frac{L_2}{\eta_2^2} \right) \bar{j}_1 \right\rangle_{s_2} \end{pmatrix} \quad (\text{B5})$$

$$\bar{Z}_{13} = \begin{pmatrix} \langle \bar{j}_1, -L_4 \bar{j}_3 \rangle_{s_1} & \langle \bar{j}_1, \eta_1 K_4 \bar{j}_3 \rangle_{s_1} \\ \langle \bar{j}_1, -\eta_1 K_4 \bar{j}_3 \rangle_{s_1} & \left\langle \bar{j}_1, -\eta_1^2 \left(\frac{L_4}{\eta_2^2} \right) \bar{j}_3 \right\rangle_{s_1} \end{pmatrix} \quad (\text{B6})$$

$$\bar{Z}_{31} = \begin{pmatrix} \langle \bar{j}_3, -L_2 \bar{j}_1 \rangle_{s_3} & \langle \bar{j}_3, \eta_1 K_2 \bar{j}_1 \rangle_{s_3} \\ \langle \bar{j}_3, -\eta_1 K_2 \bar{j}_1 \rangle_{s_3} & \left\langle \bar{j}_3, -\eta_1^2 \left(\frac{L_2}{\eta_2^2} \right) \bar{j}_1 \right\rangle_{s_3} \end{pmatrix} \quad (\text{B7})$$

$$\bar{Z}_{23} = \begin{pmatrix} \langle \bar{j}_2, L_4 \bar{j}_3 \rangle_{s_2} & \langle \bar{j}_2, -\eta_1 K_4 \bar{j}_3 \rangle_{s_2} \\ \langle \bar{j}_2, \eta_1 K_4 \bar{j}_3 \rangle_{s_2} & \left\langle \bar{j}_2, \eta_1^2 \frac{L_4}{\eta_2^2} \bar{j}_3 \right\rangle_{s_2} \end{pmatrix} \quad (\text{B8})$$

$$\bar{Z}_{32} = \begin{pmatrix} \langle \bar{j}_3, L_3 \bar{j}_2 \rangle_{s_3} & \langle \bar{j}_3, -\eta_1 K_3 \bar{j}_2 \rangle_{s_3} \\ \langle \bar{j}_3, \eta_1 K_3 \bar{j}_2 \rangle_{s_3} & \left\langle \bar{j}_3, \eta_1^2 \frac{L_3}{\eta_2^2} \bar{j}_2 \right\rangle_{s_3} \end{pmatrix} \quad (\text{B9})$$

in which there are nice symmetrical relationships that leads to computing and storing elements of only eighteen out of the thirty six submatrices to gain significant reduction in CPU time and computer memory requirements for the dense matrix $\bar{\bar{Z}}$ [27].

REFERENCES

- [1] G. Zhang and L. Tsang, "Application of angular correlation function of clutter scattering and correlation imaging in target detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, no. 5, pp. 1485-1493, September 1998.
- [2] K. O'Neil, "Discrimination of UXO in soil using broadband polarimetric GPR backscatter," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 2, pp. 356-367, September 2001.
- [3] N. Geng, A. Sullivan and L. Carin, "Multilevel fast-multipole algorithm for scattering from conducting targets above or embedded in a lossy half space," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 4, pp. 1561-1573, July 2000.
- [4] N. Geng, D. R. Jackson and L. Carin, "On the resonances of a dielectric BOR buried in a dispersive layered medium," *IEEE Trans. Antennas Propagat.*, vol 47, no. 8, pp. 1305-1313, August 1999.
- [5] N. Dasgupta, N. Geng, T. Dogaru and L. Carin, "On the extended-Born technique for scattering from buried dielectric targets," *IEEE Trans. Antennas Propagat.*, vol 47, no. 11, pp. 1306-1315, September 2000.
- [6] N. Geng and L. Carin, "Wide-band electromagnetic scattering from a dielectric BOR buried in a layered lossy dispersive medium," *IEEE Trans. Antennas Propagat.*, vol 47, no. 4, pp. 610-619, April 1999.
- [7] Z. Xiong and A. C. Tripp, "3-D electromagnetic modeling for near-surface targets using integral equations," *Geophysics*, vol. 62, no. 4, pp. 1097-1106, August 1997.
- [8] G. S. Smith and L. E. Rickard Petersson, "On the use of evanescent electromagnetic waves in the detection and identification of objects buried in lossy soil," *IEEE Trans. Antennas Propagat.*, vol 48, no. 9, pp. 1295-1300, September 2000.
- [9] L. Gürel and U. Oğuz, "Three-dimensional FDTD modeling of a ground-penetrating radar," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 4, pp. 1513-1521, July 2000.

- [10] R. T. Ling and P. Y. Ufimtsev, "Scattering of electromagnetic waves by a metallic object partially immersed in a semi-infinite dielectric medium," *IEEE Trans. Antennas Propagat.*, vol 49, no. 2, pp. 223-233, February 2001.
- [11] J. T. Johnson, "A study of the four-path model for scattering from an object above a halfspace," to appear, *Microwave Opt. Tech. Letters*, July 2001.
- [12] A. Sullivan, R. Damarla, N. Geng, Y. Dong and L. Carin, "Ultrawide-band synthetic aperture radar for detection of unexploded ordnance: Modeling and measurements," *IEEE Trans. Antennas Propagat.*, vol 48, no. 9, pp. 1306-1315, September 2000.
- [13] Y. Xu, R. M. Narayanan, X. Xu and J. O. Curtis, "Polarimetric processing of coherent random noise radar data for buried object detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 3, pp. 467-478, March 2001.
- [14] L. Carin, R. Kapoor and C. E. Baaum, "Polarimetric SAR imaging of buried landmines," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, no. 6, pp. 1985-1988, November 1998.
- [15] S. Nag and L. Peters, Jr., "Radar images of penetrable targets generated from ramp profile functions," *IEEE Trans. Antennas Propagat.*, vol 49, no. 1, pp. 32-40, January 2001.
- [16] T. Dogaru and L. Carin, "Time-domain sensing of targets buried under a rough air-ground interface," *IEEE Trans. Antennas Propagat.*, vol 46, no. 3, pp. 360-372, March 1998.
- [17] T. Dogaru, L. Collins and L. Carin, "Optimal time-domain detection of a deterministic target buried under a randomly rough interface," *IEEE Trans. Antennas Propagat.*, vol 49, no. 3, pp. 313-326, March 2001.
- [18] L. Tsang, G. Zhang and K. Pak, "Detection of a buried object under a single random rough surface with angular correlation function in EM wave scattering," *Microwave and Optical Technology Letters*, vol. 11, no. 6, pp. 300-304, April 1996.
- [19] K. O'Neill, "Broadband bistatic coherent and incoherent detection of buried objects beneath randomly rough surfaces," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 2, pp. 891-898, March 2000.

- [20] G. Zhang, L. Tsang and Y. Kuga, "Studies of the angular correlation function of scattering by random rough surfaces with and without a buried object," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, no. 2, pp. 444-453, March 1997.
- [21] G. Zhang, L. Tsang and Y. Kuga, "The angular correlation function of wave scattering by a buried object embedded in random discrete scatterers under a random rough surface," *Microwave and Optical Technology Letters*, vol. 14, no. 3, pp. 144-151, February 1997.
- [22] A. Madrazo and M. Nieto-Vesperinas, "Scattering of light and other electromagnetic waves from a body buried beneath a highly rough random surface," *J. Opt. Soc. Am. A*, vol. 14, no. 8, pp. 1859-1866, August 1997.
- [23] K. O'Neill, R. F. Lussky Jr. and K. D. Paulsen, "Scattering from a metallic object embedded near the randomly rough surface of a lossy dielectric," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, no. 2, pp. 367-376, March 1996.
- [24] G. Zhang, L. Tsang and K. Pak, "Angular correlation function and scattering coefficient of electromagnetic waves scattered by a buried object under a two-dimensional rough surface," *J. Opt. Soc. Am. A*, vol. 15, no. 12, pp. 2995-3002, December 1998.
- [25] J. T. Johnson and R. J. Burkholder, "Coupled canonical grid/discrete dipole approach for computing scattering from objects above or below a rough interface," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 6, pp. 1214-1220, June 2001.
- [26] J. T. Johnson, "A numerical study of scattering from an object above a rough surface," submitted to the *IEEE Trans. Antennas Propagat.*, 2001.
- [27] M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, "3-D subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces: Use of the steepest descent fast multipole method (SDFMM)," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 6, pp. 1174-1182, June 2001.

- [28] M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo simulations of electromagnetic wave scattering from random rough surface with 3-D penetrable buried object: Mine detection application using the SDFMM," *J. Opt. Soc. Am. A*, to appear in August 2001.
- [29] I. J. Won, D. A. Keiswetter and T. H. Bell, "Electromagnetic induction spectroscopy for clearing landmines," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 4, pp. 703-709, April 2001.
- [30] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [31] V. Jandhyala, E. Michielssen, B. Shanker and W.C. Chew, "A combined steepest descent-fast multipole algorithm for the fast analysis of three-dimensional scattering by rough surfaces," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, no. 3, pp. 738-748, May 1998.
- [32] V. Jandhyala, B. Shanker, E. Michielssen, and W. C. Chew, "A fast algorithm for the analysis of scattering by dielectric rough surfaces," *J. Opt. Soc. Am. A*, vol. 15, no. 7, pp. 1877-1885, July 1998.
- [33] M. El-Shenawee, V. Jandhyala, E. Michielssen and W. C. Chew, "The steepest descent fast multipole method (SDFMM) for solving combined field integral equation pertinent to rough surface scattering," *Proc. of the IEEE APS/URSI '99 conf.*, Orlando, Florida, pp. 534-537, July 1999.
- [34] L. Medgyesi-Mitschang, J. Putnam, and M. Gedera, "Generalized method of moments for three-dimensional penetrable scatterers," *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1383-1398, April 1994.
- [35] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Anten. and Prop.*, vol. AP 30, no. 3, pp.409-418, May 1982.
- [36] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.
- [37] J. Curtis, "Dielectric properties of soils; various sites in Bosnia," *US Army Corp. of Eng., Waterways Experim.*, Station Data Rep., 1996.
- [38] C. A. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons Inc, ch. 6, pp. 254-309.

List of Figures

Figure 1. Cross section of two penetrable objects buried under the rough surface ground.

Figure 2. (a) General penetrable 3-D scatterers R_2 and R_3 , (b) Equivalent problem exterior to S_1 , (c) Equivalent problem interior to S_1 and exterior to S_2 and S_3 , (d) Equivalent problem interior to S_2 , (e) Equivalent problem interior to S_3 .

Figure 3. Scattered electric field above a flat ground at $z = 0.5\lambda_0$ due to just the two penetrable spheroids with separating distance equal to (a) $D_x = D_y = 1\lambda_0$, (b) $D_x = D_y = 2\lambda_0$, (c) $D_x = D_y = 3\lambda_0$. H-polarization.

Figure 4. Scattered electric field above a flat ground at $z = 0.5\lambda_0$ due to just (a) the spheroid and the sphere, (b) the spheroid and the ellipsoid, (c) the spheroid and the vertical cylinder. The separation distance is $D_x = D_y = 1\lambda_0$. H-polarization.

Figure 5. Scattered electric field at $z = 0.5\lambda_0$ above the rough ground due to just (a) both the spheroid and the horizontal cylinder for $\sigma = 0.04\lambda_0$; H-polarization, (b) both the spheroid and the horizontal cylinder for same data of (a) but for V-polarization (c) the spheroid alone for same data of (a), (d) the spheroid alone for same data of (a) except $\sigma = 0.1\lambda_0$. All results have $D_x = 0.49\lambda_0$ and $D_y = 0.875\lambda_0$, and $l_c = 0.5\lambda_0$.

Figure 6. Scattered electric field above a flat ground at $z = 0.5\lambda_0$ due to just the spheroid alone (a) originally buried under the flat ground and nearby the cylinder as in Fig. 5a, (b) originally buried under the flat ground as a single object (no nearby cylinder). H-polarization.

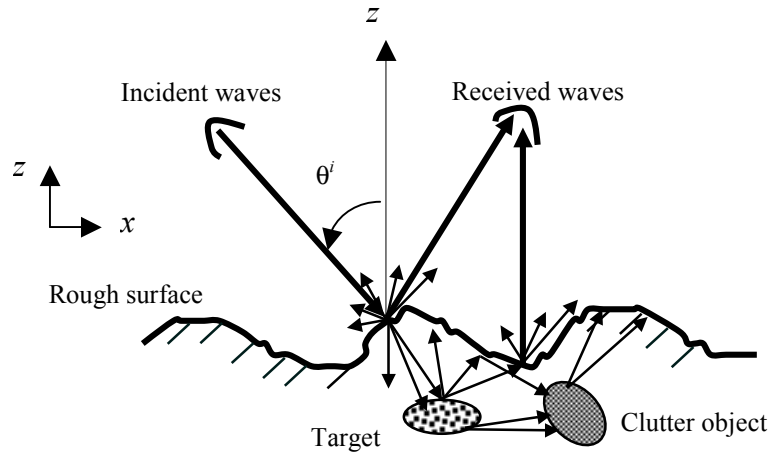


Fig. 1. Cross section of two penetrable objects buried under the rough surface ground.

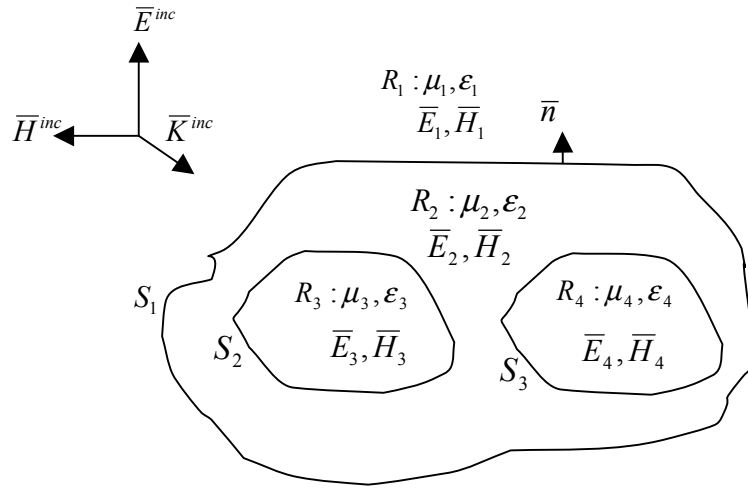


Fig. 2a. General penetrable 3-D scatterers R_2 and R_3 .

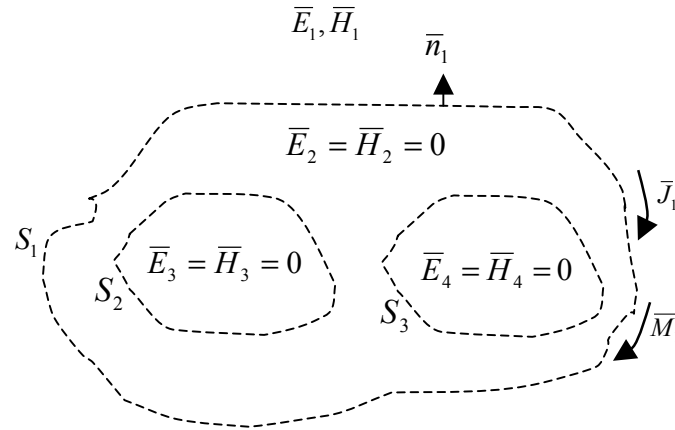


Fig. 2b. Equivalent problem exterior to S_1 .

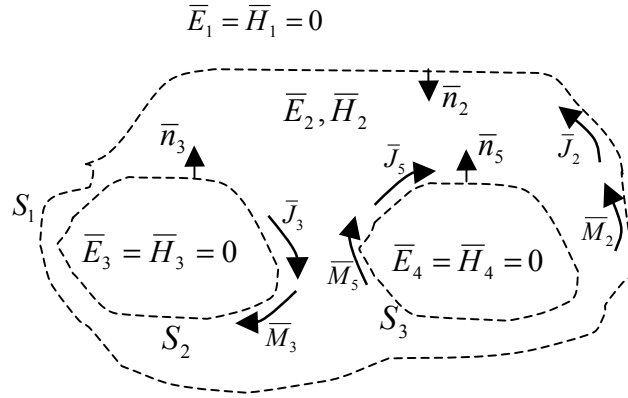


Fig. 2c. Equivalent problem interior to S_1 and exterior to S_2 and S_3 .

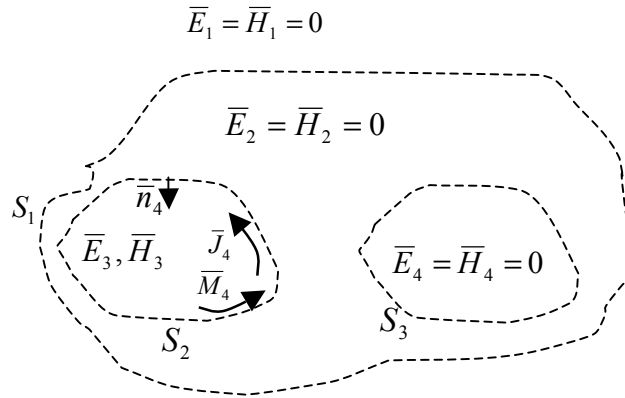


Fig. 2d. Equivalent problem interior to S_2 .

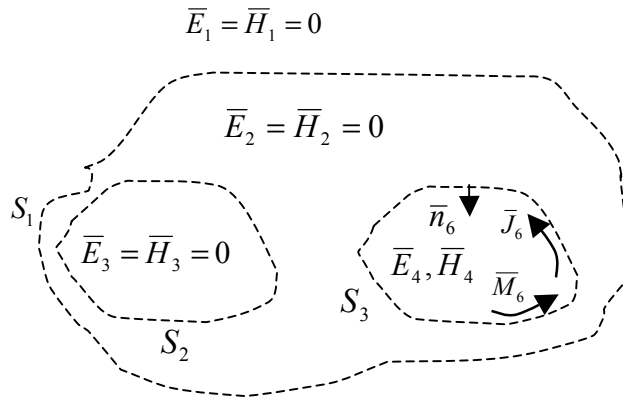
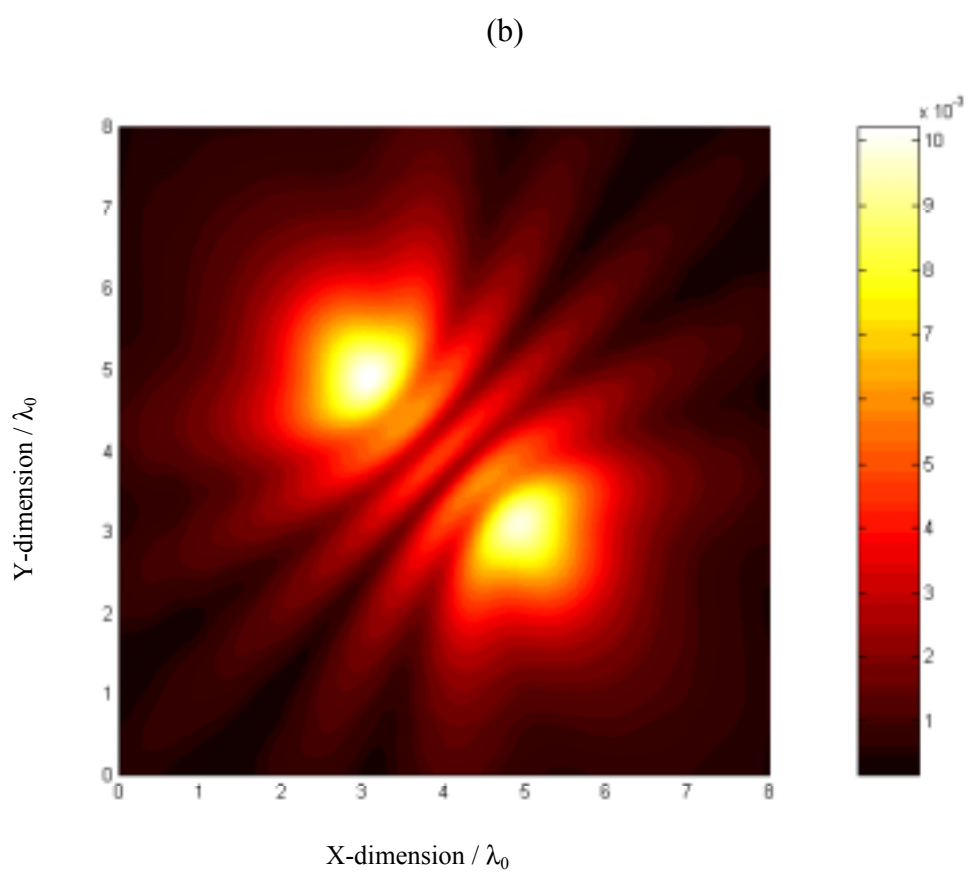
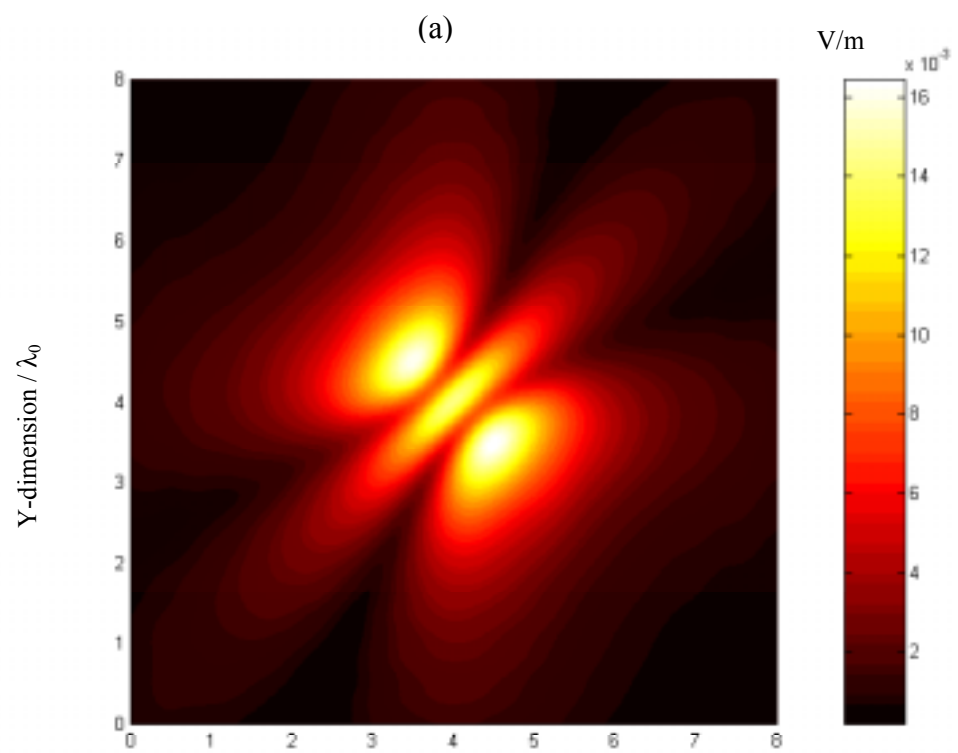


Fig. 2e Equivalent problem interior to S_3 .



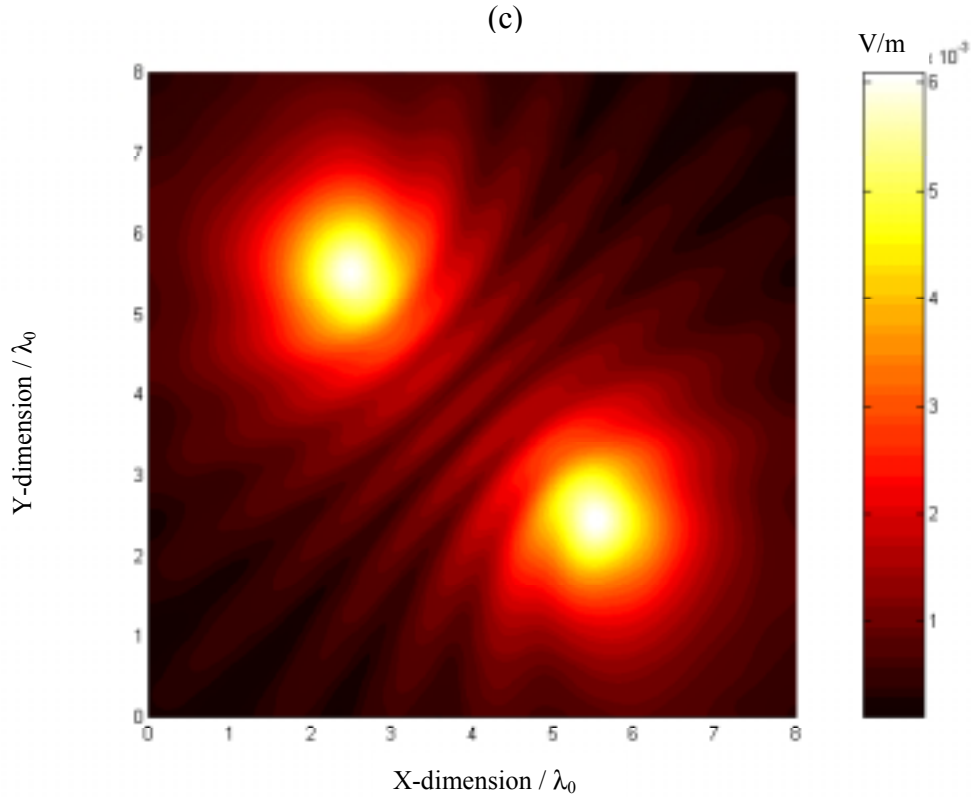
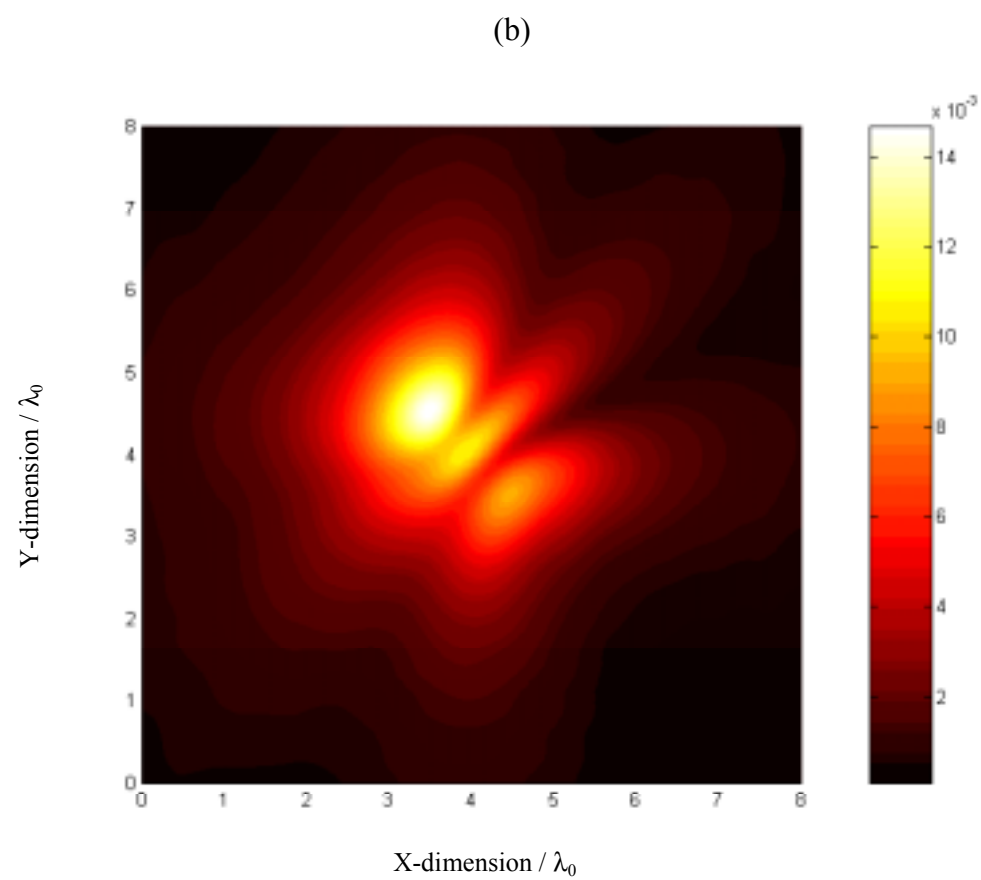
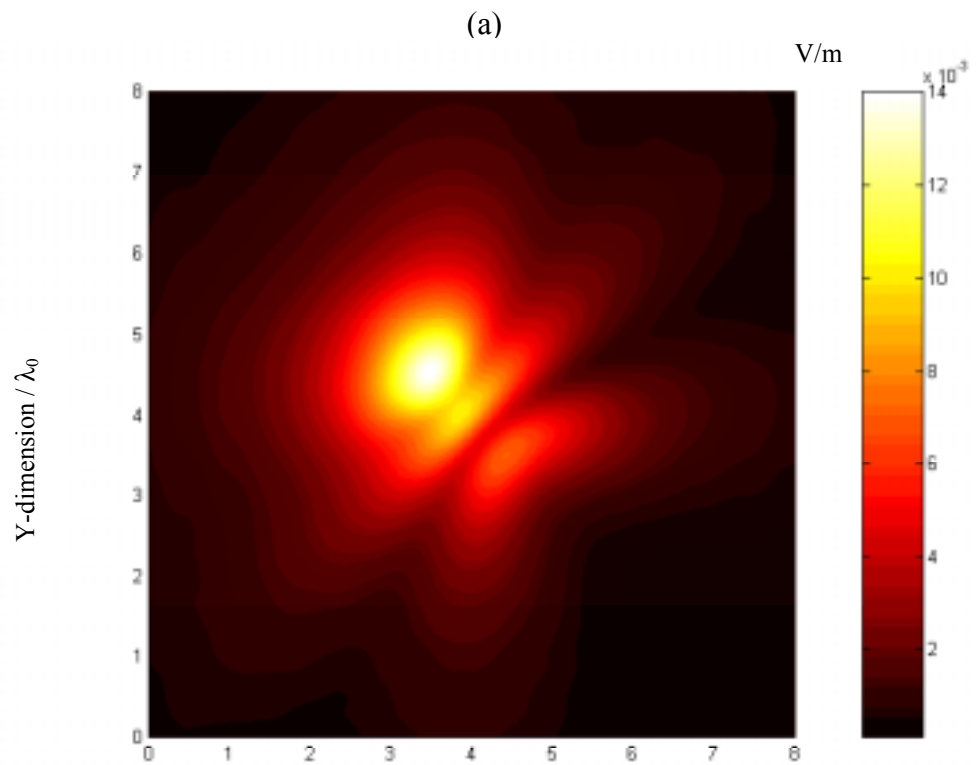


Fig. 3. Scattered electric field above a flat ground at $z = 0.5\lambda_0$ due to just the two penetrable spheroids with separating distance equal to (a) $D_x = D_y = 1\lambda_0$, (b) $D_x = D_y = 2\lambda_0$, (c) $D_x = D_y = 3\lambda_0$. H-polarization.



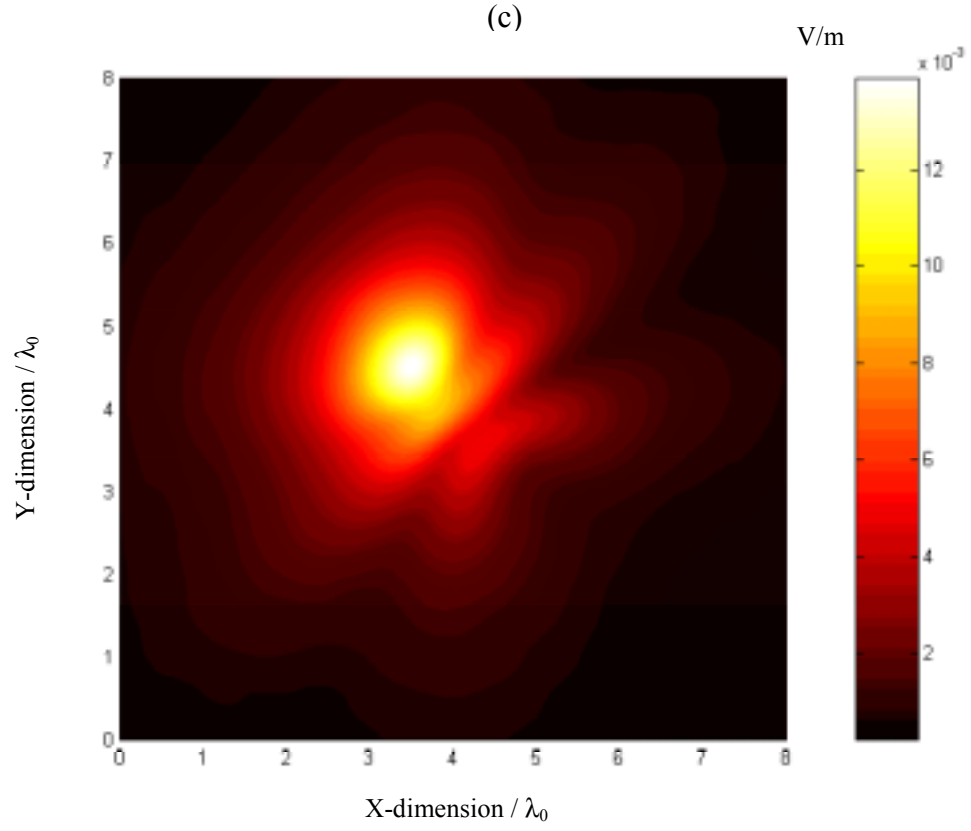
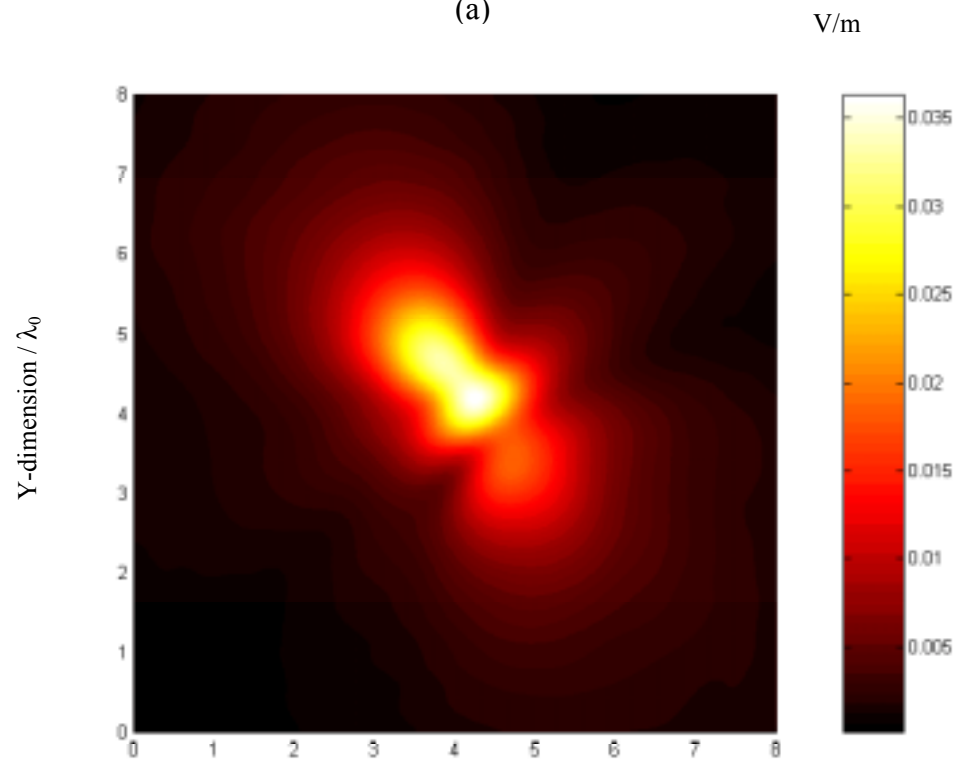
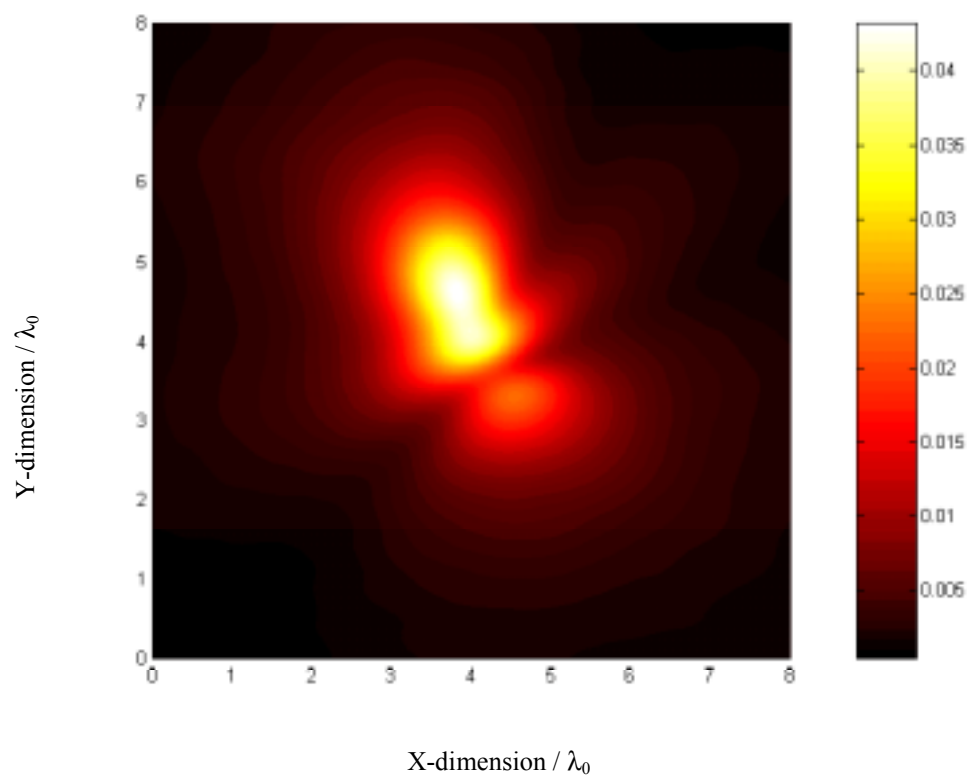


Fig. 4. Scattered electric field above a flat ground at $z = 0.5\lambda_0$ due to just (a) the spheroid and the sphere, (b) the spheroid and the ellipsoid, (c) The spheroid and the vertical cylinder. The separation distance is $D_x = D_y = 1\lambda_0$. H-polarization.

(a)



(b)



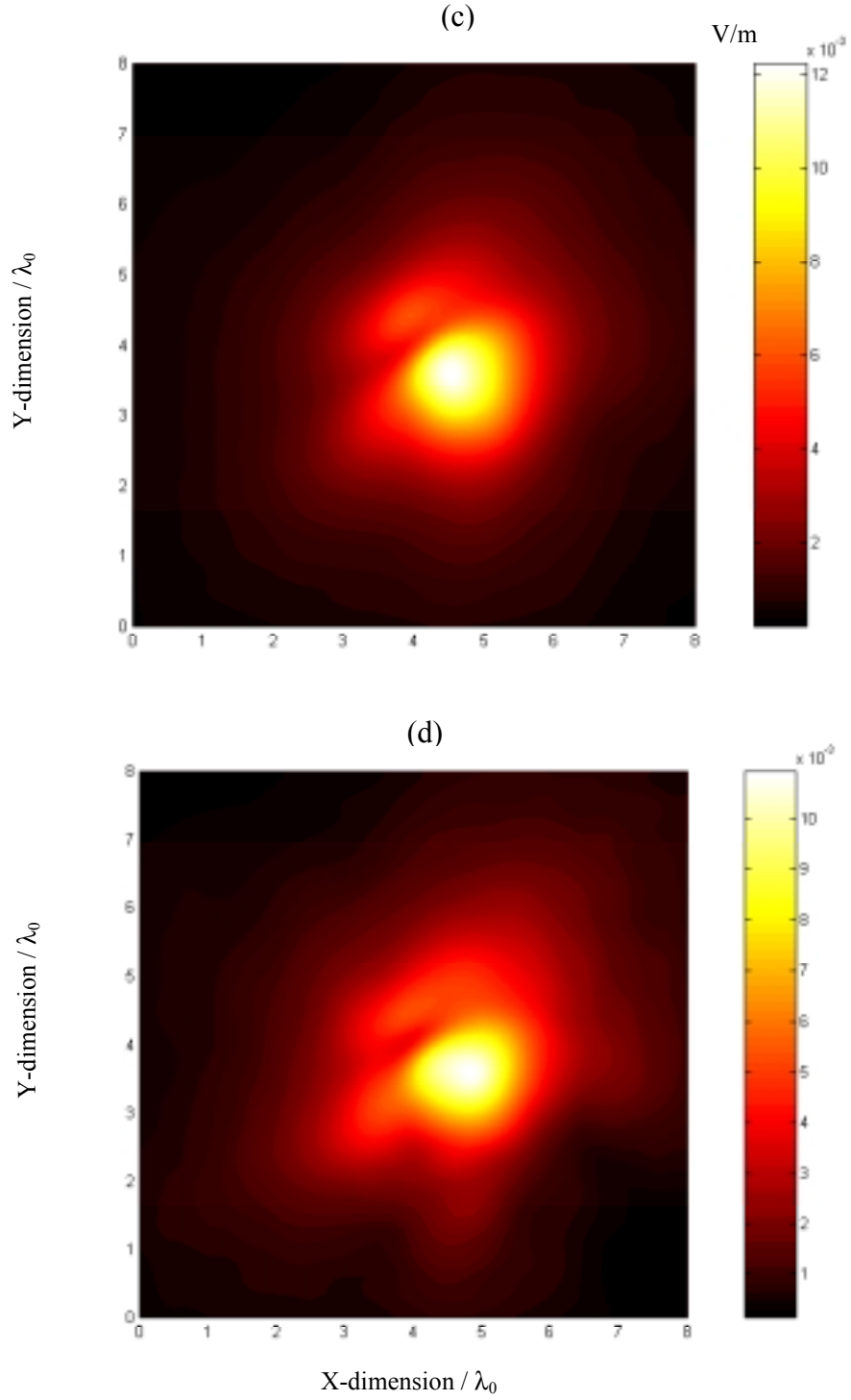


Fig. 5. Scattered electric field at $z = 0.5\lambda_0$ above the rough ground due to just (a) both the spheroid and the horizontal cylinder for $\sigma = 0.04\lambda_0$; H-polarization, (b) both the spheroid and the horizontal cylinder for same data of (a) but for V-polarization (c) the spheroid alone for same data of (a), (d) the spheroid alone for same data of (a) except $\sigma = 0.1\lambda_0$. All results have $D_x = 0.49\lambda_0$ and $D_y = 0.875\lambda_0$, and $l_c = 0.5\lambda_0$.

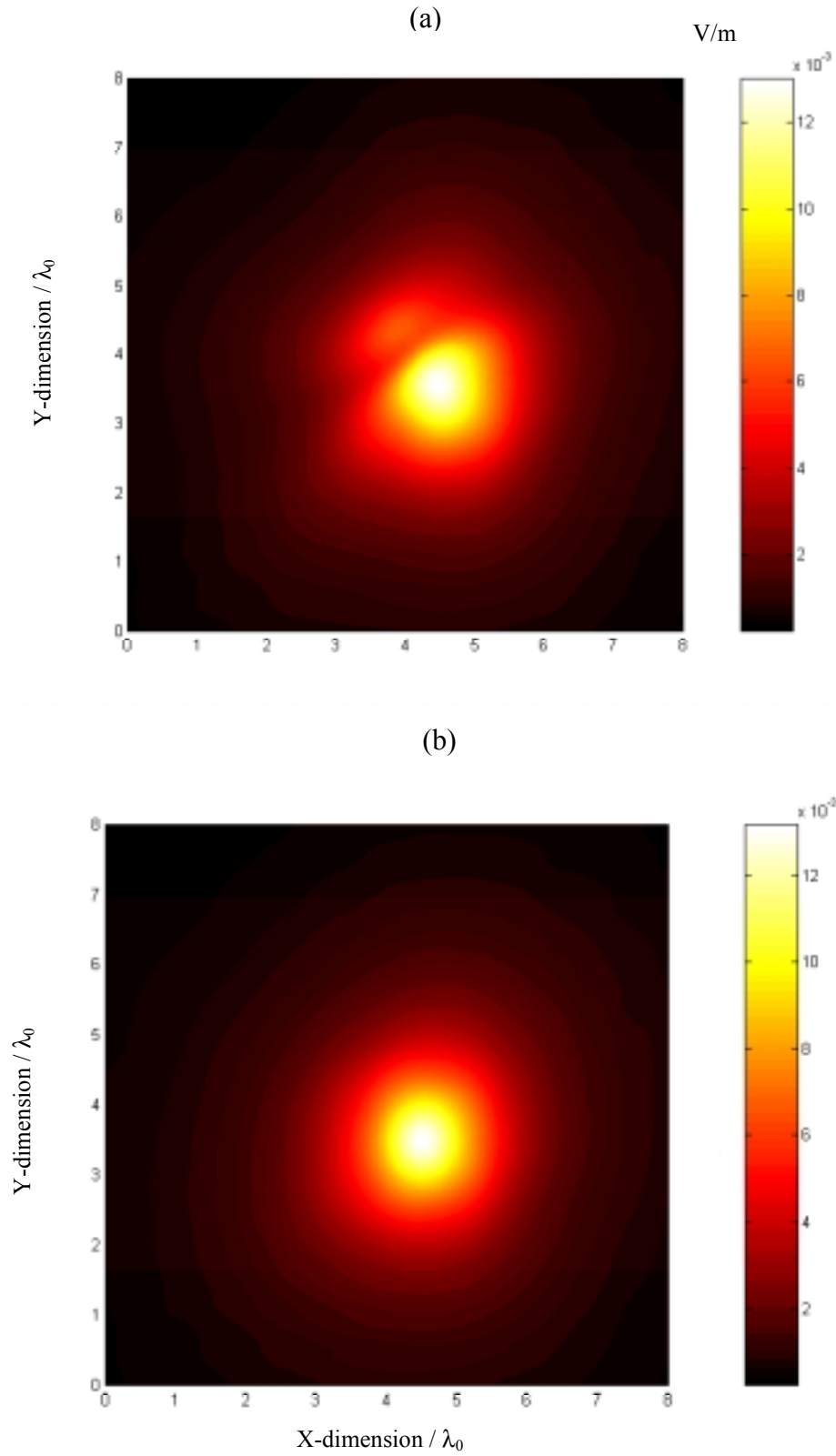


Fig. 6. Scattered electric field above a flat ground at $z = 0.5\lambda_0$ due to just the spheroid alone (a) originally buried under the flat ground and nearby the cylinder as in Fig. 5a, (b) originally buried under the flat ground as a single object (no nearby cylinder). H-polarization.

Subsurface Sensing of Targets Buried Beneath 2-D Multilayered Random Rough Surfaces: Use of the Steepest Descent Fast Multipole Method

Magda El-Shenawee

Department of Electrical Engineering
3217 Bell Engineering Center
University of Arkansas
Fayetteville, Arkansas 72701
Tel: (501)-575-6582, Fax: (501)-575-7967
magda@uark.edu

The scattering of electromagnetic waves from a penetrable shallow target buried beneath 2-D multilayered random rough surfaces will be presented. There are several applications for this work, e.g. detection of anti-personnel mines, anti-tank mines, water and/or gas pipes, location of underground water, etc. In reality, these targets are buried under the Earth's surface, which is a randomly rough interface and also is not simply composed of a single layer, but is a multilayered media. The closer the real environment is incorporated into the electromagnetic model, the more accurate and practical inferences can be gained from the numerical results. Without modeling the multiple ground layers, many targets *cannot be detected*. One unobtrusive way these buried targets can be detected is by bombarding the Earth's surface with electromagnetic waves, and comparing the scattered signature of the ground alone with that of the ground with the buried target.

A rigorous electromagnetic model based on the equivalence theorem and the method of moments (MoM) is developed to analyze this 3-D scattering problem. Three layers are considered in this work; air, dry-soil and wet-soil. The penetrable target is buried between the air/dry-soil interface and the dry/wet-soil interface. The Steepest Descent Fast Multipole Method (SDFMM) is implemented to significantly accelerate the computations of the unknown electric and magnetic surface currents. The effect of the lossy underground rough layer (wet soil) on the target signature will be investigated. Moreover, images based on the scattered electric fields for the buried target will be presented.

Invited paper.

Session organizer: Alex Maradudin, UCI.

- 1) Subsurface Sensing of Targets Buried Beneath 2-D Multilayered Random Rough Surfaces: Use of the Steepest Descent Fast Multipole Method.
- 2) Magda El-Shenawee
- 3) University of Arkansas
- 4) 3217 Bell Engineering Center
Fayetteville, AR 72701
Tel.: 501-575-6582
Fax: 501-575-7967
Email: magda@uark.edu

The Steepest Descent Fast Multipole Method for Clutter Statistics in Minefields Using Monte Carlo Simulations

Magda El-Shenawee¹ and Carey Rappaport²

¹Department of Electrical Engineering
University of Arkansas
Fayetteville, Arkansas 72701
Tel: (501)-575-6582, Fax: (501)-575-7967
magda@uark.edu

²Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
Tel: 617-373-2064, Fax: 617-373-8627
rappaport@neu.edu

A rigorous electromagnetic model has been developed to analyze the scattering mechanism of a target buried near a clutter-object under the two-dimensional random rough ground (3-D scattering problem). In realistic landmine fields, the anti-personnel (AP) nonmetallic mine is often buried nearby a rock, tree root, etc. The presence of a second object buried near the nonmetallic mine can easily obscure the target and/or cause a false alarm during the detection process.

The rigorous model is based on the classical electromagnetic equivalence theorem leading to producing six new integral equations. Using the Method of Moment (MoM), the new integral equations are transformed into a linear system of equations to be solved for the unknown electric and magnetic currents on the surface of three scatterers; rough ground, target and clutter-object. The MoM impedance matrix completely represents every interaction between these three scatterers. The Steepest Descent Fast Multipole Method (SDFMM) is used to tremendously accelerate the computations of the unknown MoM surface currents.

In previous work, we thoroughly investigated the effect of ground roughness on the signature of the target when it is buried alone under the ground. In this work, we will present numerical results for parametric investigations of the objects proximity, orientations, materials, and shapes. The results show that in certain situations, the target can be completely obscured due to the presence of the nearby clutter-object (e.g., tree root). In other cases a false indication of presence of a third buried object is observed. When the sources of clutter (e.g. the rough ground and the clutter object) are removed, by subtracting the return from both the rough ground and the clutter-object, the signature of the target can be clearly observed and analyzed. The numerical results show that the ground roughness along with the separation distance, between the target and the clutter-object, play a significant role on the probability of true or false alarm in the detection process.

Topics: #21 Rough Surface Scattering, #2 computational electromagnetics, or # 18 Subsurface Images and detection technology (GPR)

Oral presentation

1) The Steepest Descent Fast Multipole Method for Clutter Statistics in Minefields
Using Monte Carlo Simulations.

2) Presenting Author:

Magda El-Shenawee
University of Arkansas
3217 Bell Engineering Center
Fayetteville, AR 72701
Tel.: 501-575-6582
Fax: 501-575-7967
Email: magda@uark.edu

Co-author:

Carey Rappaport
Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
Tel: 617-373-2064, Fax: 617-373-8627
rappaport@neu.edu

Scattering from Non-Shallow Targets Buried Beneath Two-Dimensional Random Rough Surfaces Using the Multiple Interaction Model

Magda El-Shenawee

Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701
Tel: 501-575-6582, Fax: 501-575-7967
magda@uark.edu

The multiple interaction approach is used with the robust Steepest Descent Fast Multipole Method (SDFMM) to compute the signature of non-shallow penetrable scatterers buried beneath 2-D random rough surfaces. The most attractive feature of the proposed model is removing the quasi-planar structure constraint of the Steepest Descent Fast Multipole Method (SDFMM) when used in analyzing non-quasi-planar scatterers. The basic idea of the multiple interaction model is to decompose certain non-quasi-planar structures into two quasi-planar scatterers where the conventional SDFMM can be applied separately to each one. The interactions between the sub-quasi-planar scatterers are calculated directly using the electromagnetic vector potentials near-field expressions. Significant reductions in the CPU time and computer memory are achieved by using the SDFMM in the model. A variety of geometries are used to test the model and their numerical results are validated with the conventional MoM.

The results show that the buried object's signature is largely due to the first interaction mechanism (i.e. ground-object-ground). However, the contribution of each additional interaction is explicitly calculated using the model. Interestingly, the contributions from repeating this mechanism become insignificant especially for lossy background soil. This conclusion depends on the physical properties of the scatterer.

The multiple interaction model successfully demonstrates the exploitation of the SDFMM robustness when applied to the multilayered rough ground where the burial depth of the underground rough layer is on the order of a wavelength.

**Topics: #21 Rough Surface Scattering or #2 computational electromagnetics,
Oral presentation**

- 1) Scattering from Non-Shallow Targets Buried Beneath Two-Dimensional Random Rough Surfaces Using the Multiple Interaction Model
- 2) Magda El-Shenawee
- 3) University of Arkansas
- 4) 3217 Bell Engineering Center
Fayetteville, AR 72701
Tel.: 501-575-6582
Fax: 501-575-7967
Email: magda@uark.edu

Scattering from Multilayered Random Rough Surfaces Using the Steepest Descent Fast Multipole Method (SDFMM) and the Multiple Interaction Model

Magda El-Shenawee

Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701
Tel: 501-575-6582, Fax: 501-575-7967
magda@uark.edu

Abstract

Scattering of electromagnetic waves from multilayered random rough surfaces is crucial for subsurface sensing applications. A multiple interaction method of moments (MoM) model is used in this work to analyze scattering from two-dimensional multilayered random rough ground (3-D scattering problem) especially when the underground layer is deeply buried under the air/ground interface. The presented model removes a barrier and enables the application of the Steepest Descent Fast Multipole Method (SDFMM) to certain 3-D non-quasi-planar structures. The conventional SDFMM has been used to analyze electromagnetic wave scattering from quasi-planar structures where the scatterer's height is a fraction of a free-space wavelength. The presented model is based on a multiple interactions mechanism between the rough air/ground interface and the underground rough layer. The basic idea of the proposed model is to decompose the non-quasi-planar multilayered structure into two quasi-planar scatterers where the conventional SDFMM can be applied separately to each one. The interactions between the sub-quasi-planar scatterers are calculated using the electromagnetic vector potentials near-field expressions. This model is tested and validated with the MoM on a variety of geometries. The results show that the strongest signature of the buried scatterer is mainly due to the first multiple interaction mechanism (ground-object-ground) while the contributions from repeating this mechanism become insignificant even for lossless and/or slightly lossy soil.

Parametric Investigation of Ground Roughness on the Interference Between the AP-Mine and a Clutter-Object Buried Under Two-Dimensional Random Rough Surfaces

Magda El-Shenawee¹ and Carey Rappaport²

¹Department of Electrical Engineering
3217 Bell Engineering Center
University of Arkansas
Fayetteville, Arkansas 72701
Tel: (501)-575-6582, Fax: (501)-575-7967
magda@uark.edu

²Center for Subsurface Sensing and Imaging Systems
Department of Electrical and Computer Engineering
Northeastern University, 235 Forsyth BLD
Boston, MA 02115
Tel: (617)-373-2043, Fax: (617)-373-8627
rappaport@ece.neu.edu

In realistic landmine fields, the anti-personnel plastic mine is often buried nearby a clutter-object under the ground. The presence of a second object buried near the mine under a two-dimensional (2-D) rough ground can easily obscure the target and/or cause a false alarm. The separation distance between the AP mine and clutter-object plays a significant role on the probability of true or false alarm in this situation. A rigorous electromagnetic model has been developed to analyze the scattering mechanism between the target and the clutter-object, between the target and the rough ground, between the clutter-object and the rough ground and the multiple scattering between different spots on the rough ground itself. The new rigorous model is based on the classical electromagnetic equivalence theorem leading to producing six new integral equations. Using the Method of Moment (MoM), the new integral equations are transformed into a linear system of equations to be solved for the unknown electric and magnetic currents on the surface of three scatterers; (1) rough ground, (2) target and (3) clutter-object. The MoM impedance matrix completely represents every interaction between these three scatterers. The superior Steepest Descent Fast Multipole Method (SDFMM) is used to tremendously speed up the computations of the unknown MoM surface currents.

In previous work, we thoroughly investigated the significant effect of ground roughness on the signature of the target buried alone under the ground. Recently, we conducted a generalized parametric investigation on the objects proximity, orientations, materials, and shapes. In certain situations, the target is completely obscured due to the presence of the nearby clutter-object (e.g., tree root), and in other cases a false indication of the presence of a third buried object is observed. However, when all considered sources of clutter; (1) rough ground and (2) clutter object, are removed, by removing the return from both the rough ground and the clutter-object, the signature of the target can be clearly observed and analyzed.

In this work we will emphasize on investigating the association between the ground roughness parameters and the separation distance between the two objects and its influence on the possibility of false alarms.

Double-scatter cross sections for two-dimensional random rough surfaces that exhibit backscatter enhancement

Ezekiel Bahar

Department of Electrical Engineering, Center for Electro-Optics, University of Nebraska-Lincoln, Lincoln, Nebraska 68588-0511

Magda El-Shenawee

Center for Subsurface Sensing and Imaging Systems, Northeastern University, Boston, Massachusetts 02115-5000

Received May 9, 2000; revised manuscript received July 14, 2000; accepted July 19, 2000

Full-wave solutions are given for the single- and double-scatter radar cross sections for two-dimensional random rough surfaces. High-frequency approximations are used for the double-scatter cross sections in order to express them as numerically tractable four-dimensional integrals. The major contributions to the double-scatter cross sections are associated with the quasi-parallel and quasi-antiparallel double-scatter paths. They come from the neighborhoods of specular points. The enhancement of the backscatter cross sections, which is associated with the quasi-antiparallel double-scatter paths, is observed for both the like- and cross-polarized cross sections. © 2001 Optical Society of America

OCIS codes: 000.3860, 240.0240, 260.0260, 280.0280, 290.4210, 290.5880.

1. INTRODUCTION

In this work, full-wave solutions for the single- and double-scatter radar cross sections for two-dimensional random rough surfaces are given. The solutions are expressed as multidimensional integrals. The high-frequency, physical optics approximations are used to reduce 12-dimensional integral expressions for the double-scatter cross sections, and the large-radius-of-curvature approximation is used. The associated single-scatter cross sections are expressed in closed form. The incident waves are assumed to be plane waves.

Similar to the case for scattering from one-dimensional random rough surfaces,¹ the major contributions to the double-scatter cross sections are associated with two different pairs of propagation paths. They are the quasi-parallel double-scatter paths and the quasi-antiparallel double-scatter paths. The total incoherent double-scatter cross section is the sum of the incoherent quasi-parallel and quasi-antiparallel double-scatter cross sections.

In the high-frequency limit, the major contributions to the single- and double-scatter cross sections come from the neighborhoods of the specular points on the rough surface. Thus, upon integration with respect to the random-rough-surface slopes, the surface element scattering coefficients are evaluated at the specular points. The probability density functions for the slopes and the heights are assumed to be Gaussian. Shadow functions are included in the analysis. Physical interpretations of the analytical results are provided here.

The effects of changing the rough-surface parameters [such as mean square height and mean square slope (mss)] on the double-scatter cross sections are studied.

The level and the width of the enhancement in the backscatter direction depend on the mean square height and slope of the rough surface. The sharp backscatter enhancement, which is observed for both the like- and cross-polarized cross sections and for both normal and oblique incident angles, is associated with the quasi-antiparallel double-scatter paths. The medium below the rough interface is characterized by a complex dielectric coefficient. This work is an extension of earlier work that is restricted to scattering in the incident plane by one-dimensional rough surfaces.¹ Thus depolarization is not considered in the earlier work. It is relevant to work in remote sensing and to the design of targets with unusually large backscatter cross sections, such as traffic signs and decoys (reversed stealth).

2. FORMULATION OF THE PROBLEM

The full-wave solutions for the double-scatter electromagnetic far fields $G_d^f(\bar{r})$ from two-dimensional rough surfaces [$y = h(x_s, z_s)$] are given by¹

$$G_d^f(\bar{r}) = \left(\frac{k_0}{2\pi j} \right)^3 \frac{\exp(-jk_0 r)}{r} \int \frac{D_{2'}(\hat{\mathbf{n}}^f, \hat{\mathbf{n}}^i)}{n_y^f - n_y^i} \times \exp(jk_0 \hat{\mathbf{n}}^f \cdot \mathbf{r}_{s2'}) \times \exp[-jk_0 \hat{\mathbf{n}}^i \cdot (\mathbf{r}_{s2'} - \mathbf{r}_{s1'})] \frac{D_{1'}(\hat{\mathbf{n}}^f, \hat{\mathbf{n}}^i)}{-n_y^i + n_y^f} \times \exp(-jk_0 \hat{\mathbf{n}}^i \cdot \mathbf{r}_{s1'}) \frac{dn_y' dn_z'}{(1 - n_y'^2 - n_z'^2)^{1/2}} \times U(\mathbf{r}_{s1'}) U(\mathbf{r}_{s2'}) dx_{s1'} dz_{s1'} dx_{s2'} dz_{s2'} G^i(0), \quad (1)$$

in which the time-harmonic excitations $\exp(j\omega t)$ are assumed and the free-space wave number is $k_0 = \omega \sqrt{\epsilon_0 \mu_0}$. The integration in Eq. (1) is over the radar footprint $A = 4l_x l_z$ and components of \hat{n}' . The incident waves are in the direction of the unit vector \hat{n}^i , and the scattered waves are in the direction of the unit vector \hat{n}^f from the rough surface to the receiver at \mathbf{r} . Thus, in terms of the Cartesian unit vectors \hat{e}_i ($i = x, y, z$),

$$\hat{n}^i = n_x^i \hat{e}_x + n_y^i \hat{e}_y + n_z^i \hat{e}_z, \quad (2a)$$

$$\hat{n}^f = n_x^f \hat{e}_x + n_y^f \hat{e}_y + n_z^f \hat{e}_z, \quad (2b)$$

$$\mathbf{r} = x \hat{e}_x + y \hat{e}_y + z \hat{e}_z, \quad (2c)$$

$$\hat{n}' = n_x' \hat{e}_x + n_y' \hat{e}_y + n_z' \hat{e}_z. \quad (2d)$$

The scattering matrices at points 1' and 2' on the rough surface are $D_{1'}(\hat{n}', \hat{n}^i)$ and $D_{2'}(\hat{n}', \hat{n}^i)$ [see Fig. 1(a)]. The elements of the scattering matrices depend on the local slope of the rough surface.^{1,2} Moreover, they depend on the polarization of the incident and scattered waves and the media on both sides of the rough interface. The incident fields are assumed to be plane waves, and the receiver is located in the far field. The wave vectors of the scattered waves from the point 1' on the surface are in the direction of the unit vector \hat{n}' [see Fig. 1(a)]. The position vectors to points 1' and 2' on the rough surface are given by

$$\mathbf{r}_{s1'} = x_{s1'} \hat{e}_x + h(x_{s1'}, z_{s1'}) \hat{e}_y + z_{s1'} \hat{e}_z, \quad (3a)$$

$$\mathbf{r}_{s2'} = x_{s2'} \hat{e}_x + h(x_{s2'}, z_{s2'}) \hat{e}_y + z_{s2'} \hat{e}_z. \quad (3b)$$

At high frequencies, the shadow function $U(\mathbf{r}_{s1'})$ is equal to unity if point 1' is illuminated by the incident plane wave and visible at point 2', and it is zero otherwise; and $U(\mathbf{r}_{s2'})$ is equal to unity if point 2' is illuminated by a point source at 1' and visible from the receiver,³ and it is zero otherwise. The double-scatter cross section is obtained on multiplying Eq. (1) by its complex conjugate [see Fig. 1(b)]. The radar cross section for the two-dimensional rough surface is defined as

$$\sigma_T = \frac{4\pi r^2}{A} \left| \frac{G^f}{G^i} \right|^2, \quad (4)$$

$$\begin{aligned} \sigma_T = & \frac{k_0^6}{16A\pi^5} \\ & \times \int \frac{D_{2'}(\hat{n}', \hat{n}^i) D_{1'}(\hat{n}', \hat{n}^i) D_{2'}^*(\hat{n}', \hat{n}^i) D_{1'}^*(\hat{n}', \hat{n}^i)}{(n_y^f - n_y')(n_y^f - n_y'')(n_y^i + n_y')(n_y^i + n_y'')} \\ & \times U(\mathbf{r}_{s1'}) U(\mathbf{r}_{s1'')^*} U(\mathbf{r}_{s2'}) U(\mathbf{r}_{s2'')^*} \\ & \times \exp\{jk_0[n_x^f(x_{s2'} - x_{s2'') + n_y^f(h_{2'} - h_{2'')}) \\ & + n_z^f(z_{s2'} - z_{s2'')]\} \exp\{-jk_0[n_x^i(x_{s2'} - x_{s1'}) \\ & + n_y^i(h_{2'} - h_{1'}) + n_z^i(z_{s2'} - z_{s1'})]\} \\ & \times \exp\{jk_0[n_x''(x_{s2''} - x_{s1'') + n_y''(h_{2''} - h_{1'')}) \\ & + n_z''(z_{s2''} - z_{s1'')]\} \exp\{-jk_0[n_x^i(x_{s1'} - x_{s1'') \\ & + n_y^i(h_{1'} - h_{1'') + n_z^i(z_{s1'} - z_{s1'')]\} \\ & \times \frac{dn_y' dn_z'}{(1 - n_y'^2 - n_z'^2)^{1/2}} \frac{dn_y'' dn_z''}{(1 - n_y''^2 - n_z''^2)^{1/2}} \\ & \times dx_{s1'} dz_{s1'} dx_{s2'} dz_{s2'} dx_{s1''} dz_{s1''} dx_{s2''} dz_{s2''}. \end{aligned} \quad (5)$$

In Eq. (5) the superscript * denotes the complex conjugate. For the quasi-parallel double-scatter paths, between points 1' and 2' and points 1'' and 2'' [see Fig. 1(c)] ($n_\alpha', n_\alpha'' < 0$ and $n_\alpha', n_\alpha'' > 0$; $\alpha = x, z$), the following transformations of variables are used:

$$x_{d1} = x_{s1'} - x_{s1''}, \quad x_{d2} = x_{s2'} - x_{s2''}, \quad (6a)$$

$$z_{d1} = z_{s1'} - z_{s1''}, \quad z_{d2} = z_{s2'} - z_{s2''}, \quad (6b)$$

$$x_{a1} = (x_{s1'} + x_{s1'')/2, \quad x_{a2} = (x_{s2'} + x_{s2'')/2, \quad (7a)$$

$$z_{a1} = (z_{s1'} + z_{s1'')/2, \quad z_{a2} = (z_{s2'} + z_{s2'')/2. \quad (7b)$$

Thus, from Eqs. (5)–(7), one gets the following for the quasi-parallel case:

$$\begin{aligned} \sigma_T = & \frac{k_0^6}{16A\pi^5} \\ & \times \int \frac{D_{2'}(\hat{n}', \hat{n}^i) D_{2'}(\hat{n}', \hat{n}^i) D_{2''}^*(\hat{n}', \hat{n}^i) D_{1''}^*(\hat{n}', \hat{n}^i)}{(n_y^f - n_y')(n_y^f - n_y'')(n_y^i + n_y')(n_y^i + n_y'')} \\ & \times U(\mathbf{r}_{a1}) U(\mathbf{r}_{a2}) \exp\{jk_0[n_x^f x_{d2} \\ & - n_x'(0.5x_{d2} - 0.5x_{d1} + x_{a2} - x_{a1}) \\ & + n_x''(-0.5x_{d2} + 0.5x_{d1} + x_{a2} - x_{a1}) - n_x^i x_{d1}]\} \\ & \times \exp\{jk_0[n_z^f z_{d2} - n_z'(0.5z_{d2} - 0.5z_{d1} \\ & + z_{a2} - z_{a1}) + n_z''(-0.5z_{d2} + 0.5z_{d1} + z_{a2} - z_{a1}) \\ & - n_z^i z_{d1}]\} \exp\{jk_0[n_y^f(h_{2'} - h_{2'')}) \\ & - n_y'(h_{2'} - h_{1'}) + n_y''(h_{2''} - h_{1'') - n_y^i(h_{1'} - h_{1'')]\} \\ & \times \frac{dn_y' dn_z'}{(1 - n_y'^2 - n_z'^2)^{1/2}} \frac{dn_y'' dn_z''}{(1 - n_y''^2 - n_z''^2)^{1/2}} \\ & \times dx_{d1} dz_{d1} dx_{d2} dz_{d2} dx_{a1} dz_{a1} dx_{a2} dz_{a2}. \end{aligned} \quad (8)$$

For the quasi-parallel case [see Fig. 1(c)], the integrand in Eq. (8) becomes stationary as the points 1' and 1'' as

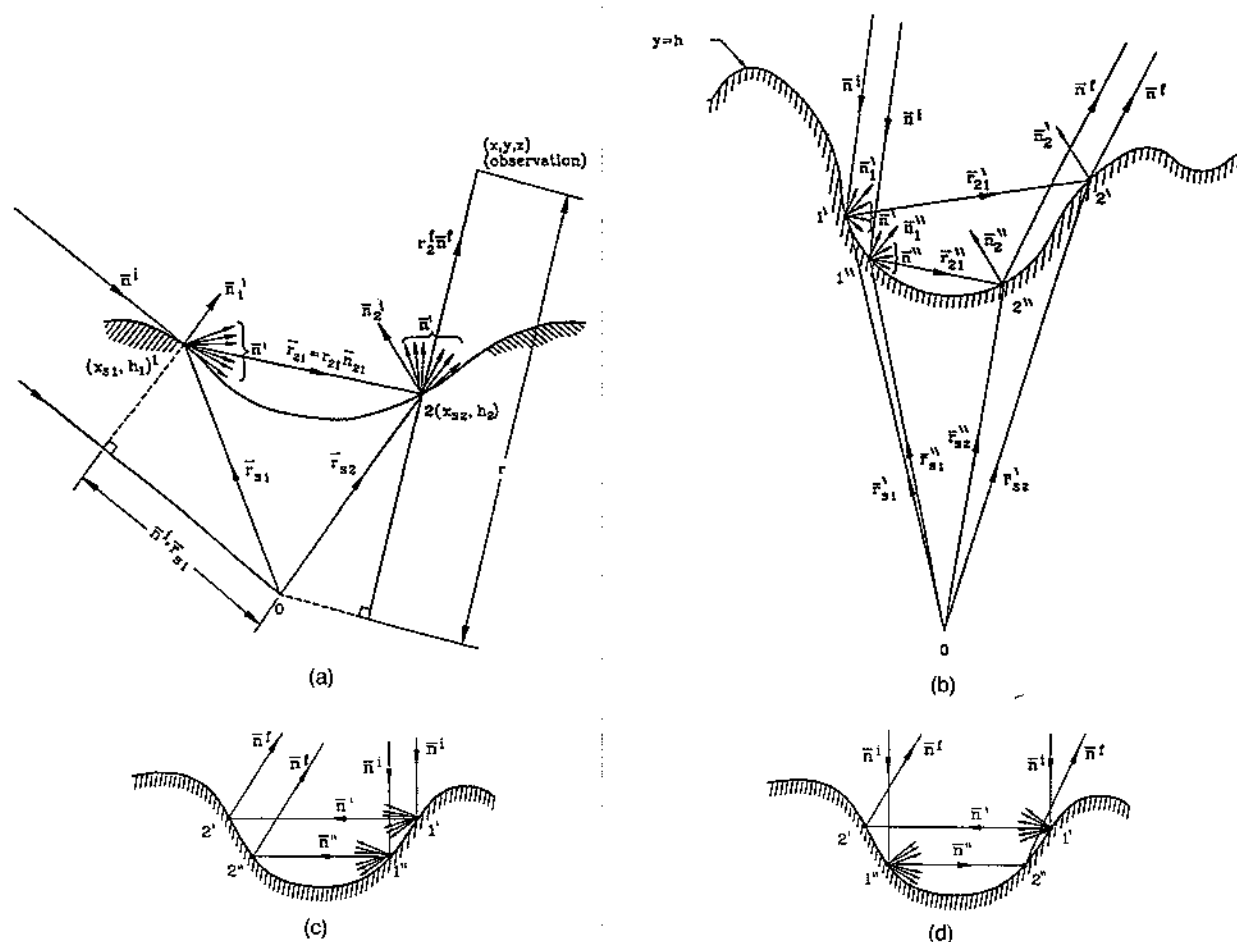


Fig. 1. (a) Double-scatter electromagnetic wave, (b) double-scatter intensity for two-dimensional random rough surfaces, (c) double-scatter quasi-parallel case, (d) double-scatter quasi-antiparallel case.

well as $2'$ and $2''$ approach each other. Thus, to simplify the 12-dimensional integral (8), we use the Taylor-series expansion for the heights at the pair of points $1', 1''$ and $2', 2''$ in terms of the heights and their derivatives at the midpoints (x_{a1}, z_{a1}) and (x_{a2}, z_{a2}) . Upon using Eqs. (6) and (7), we introduce the following approximations in Eq. (8):

$$h_{1'} = h_{a1} + (x_{d1}h_{xa1} + z_{d1}h_{za1})/2, \quad (9a)$$

in which

$$h_{a1} = h_y(x_{a1}, z_{a1}), \quad h_{xa1} = \partial h / \partial x, \quad (9b)$$

$$h_{za1} = \partial h / \partial z \text{ at } x = x_{a1} \text{ and } z = z_{a1}, \quad (9c)$$

$$h_{1''} = h_{a1} - (x_{d1}h_{xa1} + z_{d1}h_{za1})/2; \quad (9d)$$

thus

$$h_{1'} - h_{1''} = x_{d1}h_{xa1} + z_{d1}h_{za1}. \quad (9d)$$

Similar expressions are introduced in Eq. (8) for $h_{2'}$ and $h_{2''}$. The integration with respect to $x_{d1}x_{a1}$ is over the diamond-shaped area with corners at $(x_{d1} = 0, x_{a1} = l_x)$, $(x_{d1} = 2l_x, x_{a1} = 0)$, $(x_{d1} = 0, x_{a1} = -l_x)$, and $(x_{d1} = -2l_x, x_{a1} = 0)$. Similarly, diamond-shaped areas represent the limits of integration for $x_{d2}x_{a2}$, $z_{d1}z_{a1}$, and $z_{d2}z_{a2}$. Since $k_0l_x \gg k_0l_a \gg 1$ and $k_0l_z \gg k_0l_c \gg 1$ (l_c is the correlation length for the rough surface), the lim-

its of integration for the variables x_{d1} , x_{d2} , z_{d1} , and z_{d2} in Eq. (1) are assumed to be infinite.

The integral representation of the Dirac delta function is

$$\delta(h_x - h_{xs}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp[ix_d(h_x - h_{xs})] dx_d. \quad (10)$$

On integration with respect to x_{d1} , x_{d2} , z_{d1} , and z_{d2} , the 12-dimensional integral in Eq. (8) reduces to an eight-dimensional integral containing the following product of four Dirac delta functions:

$$\begin{aligned} & \delta(h_{xa1} - h_{xa1s}) \delta(h_{xa2} - h_{xa2s}) \delta(h_{za1} - h_{za1s}) \\ & \times \delta(h_{za2} - h_{za2s}) \\ & = \delta(\nabla h_{a1} - \nabla h_{a1s}) \delta(\nabla h_{a2} - \nabla h_{a2s}), \end{aligned} \quad (11)$$

in which ∇ is the symbol for gradient and the slopes at the specular points (x_{a1s}, z_{a1s}) , and (x_{a2s}, z_{a2s}) are given by

$$h_{xa1s} = - \left(-n_x^i + \frac{n_x' + n_x''}{2} \right) / \left(-n_y^i + \frac{n_y' + n_y''}{2} \right), \quad (12a)$$

$$h_{xa2s} = - \left(-n_x^f + \frac{n_x' + n_x''}{2} \right) / \left(-n_y^f + \frac{n_y' + n_y''}{2} \right), \quad (12b)$$

$$h_{za1s} = -\left(-n_z^i + \frac{n_z' + n_z''}{2}\right) / \left(-n_y^i + \frac{n_y' + n_y''}{2}\right), \quad (12c)$$

$$h_{za2s} = -\left(-n_z^f + \frac{n_z' + n_z''}{2}\right) / \left(-n_y^f + \frac{n_y' + n_y''}{2}\right). \quad (12d)$$

Thus, in the high-frequency (physical optics) limit, the major contributions to the double-scatter cross sections come from the neighborhoods of the specular points. Equation (8) reduces to

$$\sigma_T^{PQ} = \frac{k_0^2}{\pi A} \sum_{R,S=V,H} \int \frac{D_2^{PS}(\hat{n}', \hat{n}') D_1^{SQ}(\hat{n}', \hat{n}') D_2^{*PR}(\hat{n}', \hat{n}') D_1^{*RQ}(\hat{n}', \hat{n}')}{(n_y^f - n_y')(n_y^f - n_y'')(n_y^i - n_y')(n_y^i - n_y'')} U(\mathbf{r}_{a1}) U(\mathbf{r}_{a2}) \delta(\nabla h_{a1} - \nabla h_{a1s}) \\ \times \delta(\nabla h_{a2} - \nabla h_{a2s}) \exp[jk_0(\mathbf{r}_{a1} - \mathbf{r}_{a2}) \cdot (\hat{n}' - \hat{n}'')] \frac{dn_y' dn_z' dn_y'' dn_z'' dx_{a1} dx_{a2} dz_{a1} dz_{a2}}{n_x'' n_x' [n_y^f - (n_y' + n_y'')/2]^2 [n_y^i - (n_y' + n_y'')/2]^2}, \quad (13)$$

where \mathbf{r}_{a1} and \mathbf{r}_{a2} are position vectors to the midpoints [Eqs. (7)], i.e.,

$$\mathbf{r}_{ai} = x_{ai} \hat{\mathbf{e}}_x + h_{ai} \hat{\mathbf{e}}_y + z_{ai} \hat{\mathbf{e}}_z \quad (i = 1, 2). \quad (14)$$

The vectors \hat{n}' and \hat{n}'' are defined in Eqs. (2). Furthermore, for $i = 1, 2$,

$$\nabla h_{ai}(x, z) = h_{xai} \hat{\mathbf{e}}_x + h_{zai} \hat{\mathbf{e}}_z,$$

$$\nabla h_{ais}(x, z) = h_{xais} \hat{\mathbf{e}}_x + h_{zais} \hat{\mathbf{e}}_z. \quad (15)$$

The statistical average of σ_T over the random heights h_{ai} and slopes ∇h_{ai} is obtained on multiplying Eq. (13) by the probability density functions for the heights and the slopes at the midpoints and integrating over the heights and the slopes. Thus Eq. (13) reduces to

$$\langle \sigma_T^{PQ} \rangle \approx \frac{k_0^2}{\pi A} P_2(\hat{n}') P_2(\hat{n}'') \sum_{P,Q=V,H} \int \left[\frac{D_2^{PS}(\hat{n}', \hat{n}') D_1^{SQ}(\hat{n}', \hat{n}') D_2^{*PR}(\hat{n}', \hat{n}') D_1^{*RQ}(\hat{n}', \hat{n}')}{(n_y^f - n_y')(n_y^f - n_y'')(n_y^i - n_y')(n_y^i - n_y'')} \right] \\ \times \exp[jk_0(\rho_{a1} - \rho_{a2}) \cdot (\hat{n}' - \hat{n}'')] \exp[-k_0^2 \langle h^2 \rangle (n_y' - n_y'')^2] p(\nabla h_{a1s}, \nabla h_{a2s}) [1 - P_2(|n_y'|)] \\ \times [1 - P_2(|n_y''|)] \frac{dn_y' dn_z' dn_y'' dn_z'' dx_{a1} dx_{a2} dz_{a1} dz_{a2}}{n_x'' n_x' [n_y^f - (n_y' + n_y'')/2]^2 [n_y^i - (n_y' + n_y'')/2]^2}. \quad (16a)$$

For the derivation of Eq. (16), it has been assumed that the distance between the midpoints of the pairs of points 1', 1'' and 2', 2'' (namely, $|\mathbf{r}_{a1} - \mathbf{r}_{a2}|$) is larger than the correlation length l_c of the surface height autocorrelation function, such that

$$\langle \exp[ik_0(n_y' - n_y'')(h_{a1} - h_{a2})] \rangle \\ \approx \langle \exp[ik_0(n_y' - n_y'')h_{a1}] \rangle \langle \exp[-ik_0(n_y' - n_y'')h_{a2}] \rangle \\ = |\chi^2[k_0(n_y' - n_y'')]|. \quad (16b)$$

When the distance between the midpoints is small compared with l_c , the elements of the four scattering matri-

ces D^{PQ} become vanishingly small for the quasi-parallel case ($\hat{n}' \approx \hat{n}''$), since $[(\hat{n}' + \hat{n}'') \cdot \nabla h_{a1}]$, $[(\hat{n}' + \hat{n}'') \cdot \nabla h_{a2}] \rightarrow 0$. Furthermore, in Eq. (16), the surface element scattering coefficients D^{PQ} are evaluated at the specular points, P_2 is the shadow function,³ and $P_2(\hat{n}')$ and $P_2(\hat{n}'')$ are the probabilities that the rough surface is not in the shadow for the incident and scattered waves, respectively, while $[1 - P_2(|n_y|)]$ is the probability that a double-scatter event will occur. Furthermore, for $i = 1, 2$,

$$\rho_{ai} = \mathbf{r}_{ai} - h_{ai} \mathbf{e}_y \quad (17a)$$

and $p(\nabla h_{a1s}, \nabla h_{a2s})$ is the probability density function of the slopes at specular points. In Eq. (16) a Gaussian probability density function with mean square height $\langle h^2 \rangle$ has been assumed for the height of the random rough surface. To integrate with respect to the horizontal components of the midpoint position vectors, we make the following transformations:

$$\rho_{a1} - \rho_{a2} \equiv \rho_{ad}, \quad (\rho_{a1} + \rho_{a2})/2 \equiv \rho_{aa}. \quad (17b)$$

The range of the integration variables x_{aa} and z_{aa} is $-l_x < x_{aa} < l_x$ and $-l_z < z_{aa} < l_z$. The integration with respect to x_{aa} and z_{aa} yields the radar footprint $A = 4l_x l_z$, since the integrand is independent of ρ_{aa} . Furthermore, the integration with respect to the components of ρ_{ad} (x_{ad} and z_{ad}) yields a product of two sinc functions,

since the range of the variables x_{ad} and z_{ad} is $-L_m \leq x_{ad} < L_m$ and $-L_m < z_{ad} < L_m$, respectively. Thus

$$\int_{-L_m}^{L_m} \int_{-L_m}^{L_m} \exp[jk_0 \rho_{ad} \cdot (\hat{n}' - \hat{n}'')] dx_{ad} dz_{ad} \\ = 2L_m \text{sinc}[k_0 L_m (n_x' - n_x'')] 2L_m \text{sinc}[k_0 L_m (n_z' - n_z'')], \quad (18)$$

in which L_m is the mean width of a typical depression on the rough surface.^{1,4,5} The above analytical procedures reduce the 12-dimensional integral (5) (not including averaging over the random heights and slopes at two differ-

ent points on the rough surface) into a tractable four-dimensional integral over the variables n_y' , n_y'' , n_z' , and n_z'' .

Upon changing the integration variables n_y' , n_y'' , n_z' , and n_z'' to the spherical coordinate variables θ and ϕ , one gets

$$\langle \sigma_{dp}^{PQ} \rangle = \frac{(2k_0 L_m)^2}{\pi} P_2(\hat{\mathbf{n}}^i) P_2(\hat{\mathbf{n}}^f) \sum_{R,S=V,H} \int \left[\frac{[D_{2'}^{PS}(\hat{\mathbf{n}}^f, \hat{\mathbf{n}}^i) D_{1'}^{SQ}(\hat{\mathbf{n}}^f, \hat{\mathbf{n}}^i) D_{2''}^{*PR}(\hat{\mathbf{n}}^f, \hat{\mathbf{n}}^i) D_{1''}^{*RQ}(\hat{\mathbf{n}}^f, \hat{\mathbf{n}}^i)]}{(n_y^f - n_y')(n_y^f - n_y'')(-n_y^i + n_y')(-n_y^i + n_y'')} \right] \\ \times \frac{p(h_{x1s} h_{x2s} h_{z1s} h_{z2s})}{[n_y^f - (n_y' + n_y'')/2]^2 [-n_y^i + (n_y' + n_y'')/2]^2} [1 - P_2(|n_y^i|)] [1 - P_2(|n_y^f|)] \\ \times \sin[k_0 L_m (n_x' - n_x'')] \text{sinc}[k_0 L_m (n_z' - n_z'')] \exp[-(h^2) k_0^2 (n_y' - n_y'')^2] \sin \theta' \sin \theta'' d\theta' d\theta'' d\phi' d\phi''. \quad (19)$$

It is useful to determine *a priori* the regions (defined by the four variables n_y' , n_y'' , n_z' , and n_z'') that contribute most significantly to the like-polarized ($P = Q = V$ or H) and cross-polarized ($P \neq Q = V$ or H) bistatic double-scatter cross sections. In view of the two sinc functions and the characteristic function $\chi[k_0(n_y' - n_y'')h]$ in the integral expression for the double-scatter cross sections [Eq. (19)], the major contribution to the integral is from the region $\hat{\mathbf{n}}^i \rightarrow \hat{\mathbf{n}}^f$ (quasi parallel). Furthermore, because of the shadow function $(1 - P_2)$, when these quasi-parallel paths are also quasi horizontal ($n_y' \approx 0$, $n_y'' \approx 0$), the value of the integrand in Eq. (19) peaks. To facilitate the numerical integration of relation (16), in view of the integrable singularity associated with $n_x' n_x''$ in the denominator of relation (16), the following transformations to polar coordinates are implemented for both $\hat{\mathbf{n}}^i$ and $\hat{\mathbf{n}}^f$:

$$n_x = \sin \theta \cos \phi, \quad n_y = \cos \theta, \quad n_z = \sin \theta \sin \phi, \\ 0 < \theta < \pi, \quad \phi_0 < \phi < 2\pi + \phi_0, \quad (20)$$

where ϕ_0 is an arbitrary constant. The product of the Jacobians of these transformations is

$$J' J'' = \sin^2 \theta' \cos \phi' \sin^2 \theta'' \cos \phi''. \quad (21a)$$

The ranges of the variables are $0 < \theta'$, $\theta'' < \pi$ and $\phi_0 < \phi'$, $\phi'' < 2\pi + \phi_0$. For the quasi-parallel case, $\phi_0 = 0$ is a suitable choice. Thus

$$dn_y' dn_y'' dn_z' dn_z'' / n_x' n_x'' = \sin \theta' \sin \theta'' d\theta' d\theta'' d\phi' d\phi''. \quad (21b)$$

Since the scattering coefficients D^{PQ} vanish for grazing incident and scatter angles, there are no singularities that are due to the product appearing below the scattering coefficients D^{PQ} in Eq. (19). For $n_y^i \rightarrow 0$, $n_y^f \rightarrow 0$ grazing incident and scatter angles, the shadow functions $P_2(\hat{\mathbf{n}}^i)$ and $P_2(\hat{\mathbf{n}}^f)$ vanish. Furthermore, when $n_y' \rightarrow 0$ and $n_y'' \rightarrow 0$, the probability for a double-scatter event $(1 - P_2) \rightarrow 1$. Thus Eq. (19) is not singular when the product appearing below the probability density function for the slopes (at the specular points) vanishes. It takes 1–2 CPU hours to evaluate the four-dimensional integrals in Eq. (19) for all four cross sections σ^{PQ} on a Silicon Graphics Computer model 4D/380VGX in the Center for Electro-Optics at the University of Nebraska-Lincoln. To

make sure that the integration subroutine does not miss the very narrow region around $\theta'' \approx \theta' \approx \pi/2$ and $\phi'' \approx \phi'$, where the integrand peaks for the quasi-parallel case, and to reduce the regions of integration in Eq. (19), we subdivide the limits of integration ($0 \leq \theta'$, $\theta'' \leq \pi$ and $0 < \phi'$, $\phi'' \leq 2\pi$) as follows for the quasi-parallel

case:

$$0 \leq \theta'' \leq \theta', \quad \theta' \leq \theta'' \leq \pi, \\ 0 \leq \phi'' \leq \phi', \quad \phi' \leq \phi'' \leq 2\pi. \quad (22a)$$

It can be shown that if the medium below the rough interface [$y < h(x, z)$; μ_1 , ϵ_1] is finitely conducting, the complex value of the integral $\langle \sigma_{dp}^{PQ} \rangle$ corresponding to $0 \leq \theta'' \leq \theta'$ is equal to the complex conjugate of the value of the integral $\langle \sigma_{dp}^{PQ} \rangle$ corresponding to $\theta' \leq \theta'' \leq \pi$. Thus it is necessary to evaluate only the real part of the integral [Eq. (19)] over half the total region of integration in Eq. (19), since

$$\langle \sigma_{dp}^{PQ} \rangle = \langle \sigma_1^{PQ} \rangle + \langle \sigma_2^{PQ} \rangle = 2 \text{Re} \langle \sigma_1^{PQ} \rangle. \quad (22b)$$

As noted above, this scheme not only increases the accuracy of the numerical integration of $\langle \sigma_{dp}^{PQ} \rangle$ (since the most significant part of the integral is never missed) but also reduces the amount of computation needed to evaluate the integrals. However, no attempt has been made to parallelize the computer codes in order to decrease the computation time for parallel-processing supercomputers.

For the quasi-antiparallel double-scatter paths between points 1' and 2' and points 1'' and 2'' [see Fig. 1(d)], the following transformations of variables are used in Eq. (5):

$$x_{d1} = x_{s1'} - x_{s2''}, \quad x_{d2} = x_{s2'} - x_{s1''}, \quad (23a)$$

$$z_{d1} = z_{s1'} - z_{s2''}, \quad z_{d2} = z_{s2'} - z_{s1''}, \quad (23b)$$

$$x_{a1} = (x_{s1'} + x_{s2''})/2, \quad x_{a2} = (x_{s2'} + x_{s1''})/2, \quad (23c)$$

$$z_{a1} = (z_{s1'} + z_{s2''})/2, \quad z_{a2} = (z_{s2'} + z_{s1''})/2. \quad (23d)$$

Furthermore, in addition to Eqs. (9a) and (9b) (unchanged), Eqs. (9c) and (9d) are replaced by

$$h_{2''} = h_{a2} - (x_{d2} h_{xa2} + z_{d2} h_{za2})/2; \quad (24a)$$

thus

$$h_{1'} - h_{2''} = x_{d1} h_{xa1} + z_{d1} h_{za1}. \quad (24b)$$

Equations (17) (unchanged) are also used to reduce the 12-dimensional integral into a four-dimensional integral. Thus, on following the same procedures as those for the quasi-parallel case, one gets the following integral expres-

sion for the (high-frequency) quasi-antiparallel ($n'_a < 0$, $n''_a > 0$ or $n'_a > 0$, $n''_a < 0$) double-scatter cross sections:

$$\begin{aligned} \langle \sigma_{da}^{PQ} \rangle = & \frac{(2k_0 L_m)^2}{\pi} P_2(\hat{\mathbf{n}}^i) P_2(\hat{\mathbf{n}}^f) \sum_{R,S=V,H} \int \left[\frac{D_{2'}^{PS}(\hat{\mathbf{n}}^f, \hat{\mathbf{n}}^i) D_{1'}^{SQ}(\hat{\mathbf{n}}^f, \hat{\mathbf{n}}^i) D_{2''}^{*PR}(\hat{\mathbf{n}}^f, \hat{\mathbf{n}}^i) D_{1''}^{*RQ}(\hat{\mathbf{n}}^f, \hat{\mathbf{n}}^i)}{(n_y^f - n_y^i)(n_y^f - n_y^i)(-n_y^i + n_y^i)(-n_y^i + n_y^i)} \right] \\ & \times p(h_{xa1s}, h_{xa2s}, h_{za1s}, h_{za2s}) 16/(n_y^f - n_y^i + n_y^i - n_y^i) (n_y^f - n_y^i - n_y^i + n_y^i)^2 [1 - P_2(|n_y^i|)] [1 - P_2(|n_y^i|)] \\ & \times \text{sinc}[k_0 L_m(n_x^f + n_x^i - n_x^i - n_x^i)] \text{sinc}[k_0 L_m(n_z^f + n_z^i - n_z^i - n_z^i)] \exp[-(h^2)k_0^2(n_y^f - n_y^i - n_y^i + n_y^i)^2] \\ & \times \sin \theta' \sin \theta'' d\theta' d\theta'' d\phi' d\phi''. \end{aligned} \quad (25a)$$

For the quasi-antiparallel case, the arguments of the sinc functions and the characteristic function vanish (and the integrand peaks) when $\hat{\mathbf{n}}' = -\hat{\mathbf{n}}''$ only for backscatter ($\hat{\mathbf{n}}^f = -\hat{\mathbf{n}}^i$). In Eq. (25a) the vectors $\hat{\mathbf{n}}'$ and $\hat{\mathbf{n}}''$ are expressed in terms of their polar coordinates as in Eqs. (20), except that for the quasi-antiparallel case, $\hat{\mathbf{n}}' = -\hat{\mathbf{n}}''$ ($\theta'' = \pi - \theta'$, $\phi'' = \phi' + \pi$). In relations (22a), it is convenient to choose $\phi_0 = -\pi$ for ϕ' and $\phi_0 = 0$ for ϕ'' for the quasi-antiparallel case, such that $-\pi \leq \phi' < \pi$ and $0 \leq \phi'' \leq 2\pi$. In this case, the limits of integration are subdivided as follows:

$$\begin{aligned} 0 \leq \theta'' \leq \pi - \theta', \quad \pi - \theta' \leq \theta'' \leq \pi, \\ 0 \leq \phi'' \leq \phi' + \pi, \quad \phi' + \pi \leq \phi'' \leq 2\pi. \end{aligned} \quad (25b)$$

The slopes at the specular points for the quasi-antiparallel case are given by

$$h_{xa1s} = -(n_x^f - n_x^i + n_x^i - n_x^i)/(n_y^f - n_y^i + n_y^i - n_y^i), \quad (26a)$$

$$h_{xa2s} = -(n_x^f - n_x^i + n_x^i - n_x^i)/(n_y^f - n_y^i + n_y^i - n_y^i), \quad (26b)$$

$$h_{za1s} = -(n_z^f - n_z^i + n_z^i - n_z^i)/(n_y^f - n_y^i + n_y^i - n_y^i), \quad (26c)$$

$$h_{za2s} = -(n_z^f - n_z^i + n_z^i - n_z^i)/(n_y^f - n_y^i + n_y^i - n_y^i), \quad (26d)$$

The sharp enhancement in the backscatter direction ($-\hat{\mathbf{n}}^i = \hat{\mathbf{n}}^f$) is associated with the quasi-antiparallel ($\hat{\mathbf{n}}^f \approx -\hat{\mathbf{n}}''$) double-scatter cross section [Eq. (25a)].^{1,4,6,7} Note the differences in the expressions for the slopes at the specular points for the quasi-parallel and quasi-antiparallel cases (12) and (26), respectively. Equations (12) and (26) can be readily interpreted. The vectors normal to the surface at the specular points 1', 1'', 2', and 2'' are $\hat{\mathbf{n}}_{s1'}$, $\hat{\mathbf{n}}_{s1''}$, $\hat{\mathbf{n}}_{s2'}$, and $\hat{\mathbf{n}}_{s2''}$, respectively, where

$$\begin{aligned} \hat{\mathbf{n}}_{s1'} &= \hat{\mathbf{n}}' - \hat{\mathbf{n}}^i, & \hat{\mathbf{n}}_{s1''} &= \hat{\mathbf{n}}'' - \hat{\mathbf{n}}^i, \\ \hat{\mathbf{n}}_{s2'} &= \hat{\mathbf{n}}^f - \hat{\mathbf{n}}^i, & \hat{\mathbf{n}}_{s2''} &= \hat{\mathbf{n}}^f - \hat{\mathbf{n}}^i. \end{aligned} \quad (27)$$

For the quasi-parallel case $\hat{\mathbf{n}}' \approx \hat{\mathbf{n}}''$, the vectors $\hat{\mathbf{n}}_{s1'}$ and $\hat{\mathbf{n}}_{s2'}$ are also quasi parallel to the vectors $\hat{\mathbf{n}}_{s1''}$ and $\hat{\mathbf{n}}_{s2''}$, respectively, and the slopes at the midpoints between 1', 1'' and 2', 2'' [see Fig. 1(c)] are given by the vectors

$$\hat{\mathbf{n}}_{sdp1} = (\hat{\mathbf{n}}_{s1'} + \hat{\mathbf{n}}_{s1''})/2, \quad \hat{\mathbf{n}}_{sdp2} = (\hat{\mathbf{n}}_{s2'} + \hat{\mathbf{n}}_{s2''})/2, \quad (28)$$

corresponding to the slopes in Eqs. (12). Similarly, for the antiparallel case $\hat{\mathbf{n}}' \approx -\hat{\mathbf{n}}''$, the slopes at the midpoints between 1', 2'' and 1'', 2' [see Fig. 1(d)] are given by the components of the vectors

$$\hat{\mathbf{n}}_{sdp1} = (\hat{\mathbf{n}}_{s1'} + \hat{\mathbf{n}}_{s2''})/2, \quad \hat{\mathbf{n}}_{sdp2} = (\hat{\mathbf{n}}_{s1''} + \hat{\mathbf{n}}_{s2'})/2, \quad (29)$$

corresponding to the slopes in Eqs. (26). Furthermore, the integrand in Eq. (19) peaks when the pairs of vectors $\hat{\mathbf{n}}_{s1'}$, $\hat{\mathbf{n}}_{s1''}$ and $\hat{\mathbf{n}}_{s2'}$, $\hat{\mathbf{n}}_{s2''}$ are parallel to each other, namely,

$$\hat{\mathbf{n}}_{sdp} = \hat{\mathbf{n}}_{s1'} - \hat{\mathbf{n}}_{s1''} = \hat{\mathbf{n}}_{s2'} - \hat{\mathbf{n}}_{s2''} = \hat{\mathbf{n}}' - \hat{\mathbf{n}}'' = 0. \quad (30)$$

The integrand in Eq. (25a) peaks when the pairs of vectors $\hat{\mathbf{n}}_{s1'}$, $\hat{\mathbf{n}}_{s2''}$, and $\hat{\mathbf{n}}_{s1''}$, $\hat{\mathbf{n}}_{s2'}$, are parallel, namely,

$$\begin{aligned} \hat{\mathbf{n}}_{sda} &= \hat{\mathbf{n}}_{s1'} - \hat{\mathbf{n}}_{s2''} = \hat{\mathbf{n}}_{s1''} - \hat{\mathbf{n}}_{s2'} \\ &= \hat{\mathbf{n}}' + \hat{\mathbf{n}}'' - \hat{\mathbf{n}}^f - \hat{\mathbf{n}}^i \rightarrow 0. \end{aligned} \quad (31)$$

Note that for Eq. (30) to be satisfied, the paths must be parallel; however, for Eq. (31) to be satisfied, the paths must be antiparallel for backscatter only. This distinction between the quasi-parallel case and the quasi-antiparallel case explains why the enhanced backscatter is associated only with the quasi-antiparallel case.

For backscatter at normal incidence, the major contributions to the double-scatter cross sections are associated with the quasi-horizontal paths between points 1 and 2 [see Fig. 1(d)], and the slopes at the stationary points are approximately $\pm 45^\circ$. The integrands in the final expressions for the double-scatter cross sections [Eqs. (19) and (25)] peak for a very small range of angles at which $\hat{\mathbf{n}}'$ and $\hat{\mathbf{n}}''$ are quasi horizontal ($n_y^f \approx 0$ and $n_y^i \approx 0$), as a result of the presence of $1 - \bar{P}_2$ in the integrand. The characteristic functions

$$\chi[k_0(n_y^f - n_y^i)] = \exp[-(h^2)k_0^2(n_y^f - n_y^i)^2/2],$$

$$\begin{aligned} \chi[k_0(n_y^f + n_y^i - n_y^i - n_y^i)] \\ = \exp[-(h^2)k_0^2(n_y^f + n_y^i - n_y^i - n_y^i)^2/2] \end{aligned}$$

in Eqs. (19) and (25) peak for $(n_y^f - n_y^i) \rightarrow 0$ (quasi parallel) and $(n_y^f + n_y^i - n_y^i - n_y^i) \rightarrow 0$ (quasi antiparallel and backscatter), respectively. For the illustrative examples, Eqs. (19) and (25) are used to evaluate the bi-static double-scatter cross sections. The corresponding high-frequency single-scatter cross sections are

$$\begin{aligned} \langle \sigma_s^{PQ} \rangle &= 4\pi |D^{PQ}(\hat{\mathbf{n}}^f, \hat{\mathbf{n}}^i)|^2 p(h_{xs}, h_{zs})/(n_y^f - n_y^i)^4, \\ P &= Q, \quad \langle \sigma_s^{PQ} \rangle = 0, \quad P \neq Q, \end{aligned} \quad (32)$$

since the incident waves are not depolarized at the specular points in the plane of incidence.

3. ILLUSTRATIVE EXAMPLES

The incoherent single- and double-scatter cross sections (quasi parallel+quasi antiparallel), as well as the total (single+double) cross sections, for two-dimensional rough surfaces are plotted in Figs. 2-10 as functions of the scatter angles $\theta^f \cos \phi^f$ (where $\phi^f = 0, \pi$). The two-dimensional rough surface is assumed to be coated with gold permittivity $\epsilon_r = -9.888312 - j1.051766$ at $\lambda = 0.633 \mu\text{m}$. The incident angles are $\theta^i = 10^\circ$, $\phi^i = 0$. The Rayleigh roughness parameter is $\beta = 4k_0^2 \langle h^2 \rangle$. In Figs. 2-4, the single-scatter cross sections are plotted as functions of the scatter angles $\theta^f \cos \phi^f$, $\phi^f = 0$ and π . In Fig. 2 the polarizations of the incident and scattered waves are vertical (parallel). The Rayleigh roughness parameter $\beta = 4k_0^2 \langle h^2 \rangle$ varies from 10 to 394, and the mean square slope is $mss = 0.5$. The single-scatter incoherent cross sections are insensitive to $\langle h^2 \rangle$, since, for the entire range of the parameters considered, $\beta \gg 1$. They are sensitive to $\langle h^2 \rangle$ when $\beta < 1$. They are proportional to $\langle h^2 \rangle$ for $\beta \ll 1$ where the coherent scatter cross sections are not small. For $\beta \gg 1$ the coherent cross sections are negligible.

In Fig. 3 the polarizations of the incident and scattered waves are vertical (parallel), and the mss varies from 0.25 to 1.0 and $\beta = 394$. For the range of parameters consid-

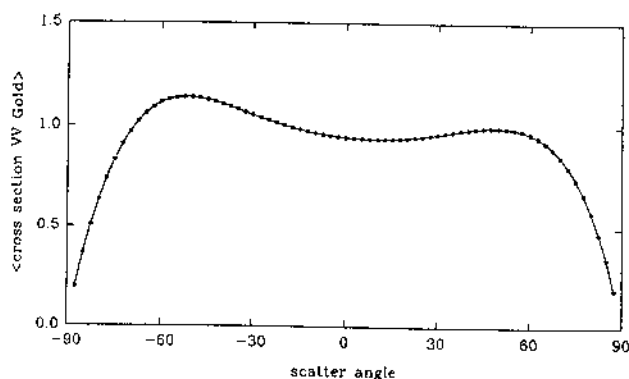


Fig. 2. Single-scatter vertical-to-vertical radar cross section. Mean square height (Rayleigh parameter β) is the variable parameter $10 \leq \beta \leq 394$. The mean square slope (mss) value is 0.5. The following parameters apply to Figs. 2-10: incident angle of 10° , relations permittivity of $-9.888312/j1.051766$, and wavelength of $0.633 \mu\text{m}$.

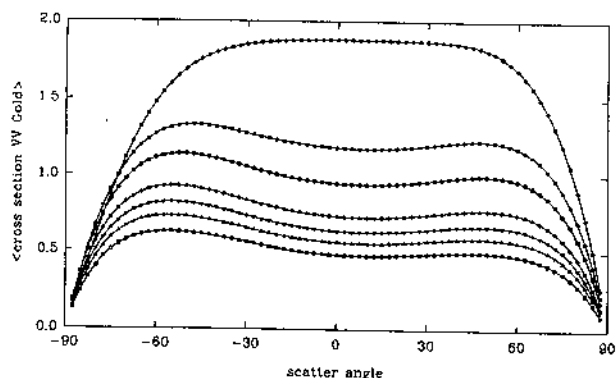


Fig. 3. Single-scatter vertical-to-vertical radar cross section. Mean square slope (mss) is the variable parameter, and the curves from top to bottom correspond to mss values of 0.25, 0.4, 0.5, 0.65, 0.75, 0.85, and 1.0. The value of β is 394.

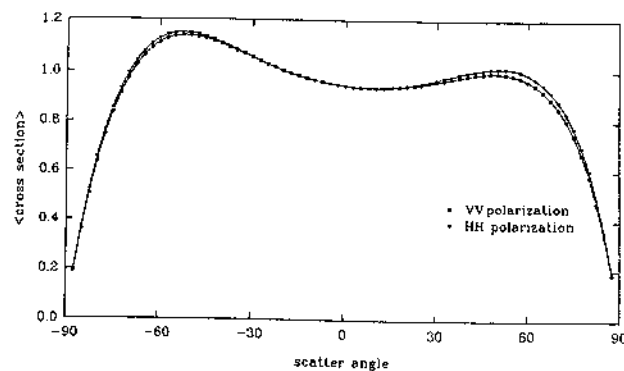


Fig. 4. Single-scatter like-polarized radar cross section for vertical to vertical (lower curve) and horizontal-to-horizontal (upper curve) polarization. The mss and β values are 0.5 and 394, respectively.

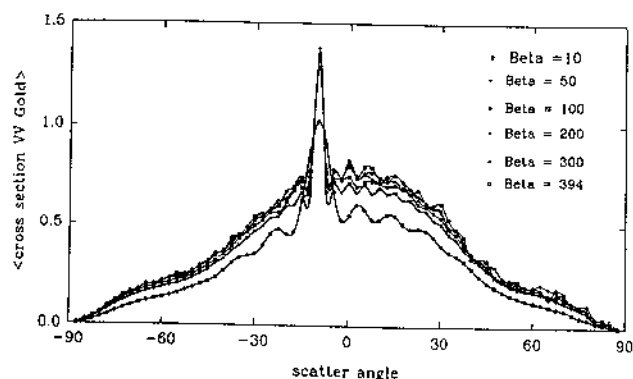


Fig. 5. Double-scatter radar cross section. Mean square height (Rayleigh parameter β) is the variable parameter (quasi-parallel and quasi-antiparallel contributions added). The polarization is vertical to vertical. The mss value is 0.5. Lowest curve, $\beta = 10$. Data saturate as β increases.

ered, the single-scatter cross sections decrease as the mss increases, while the double-scatter cross sections increase with increasing mss . As the mss becomes larger than unity, the values of both the single- and double-scatter cross sections saturate.

In Fig. 4 the single-scatter cross sections are plotted for both vertical (parallel) and horizontal (perpendicular) polarizations. The results indicate that for the given values of $\beta = 394$ and $mss = 0.5$, the single-scatter cross sections are practically the same for both like polarizations.

In Figs. 5-7 the double-scatter cross sections are plotted as functions of the scatter angles $\theta^f \cos \phi^f$, $\phi^f = 0$ or π . These three illustrations correspond to the cases considered in Figs. 2-4.

In Fig. 5 the polarizations of the incident and scatter waves are vertical (parallel). The roughness parameter β increases from 10 to 394, and $mss = 0.5$. The double-scatter cross sections (unlike the single-scatter cross sections) are sensitive to variations in β . The cross sections increase as β increases, and they practically merge for $\beta > 100$. For $\beta \gg 100$ higher-order multiple-scatter cross section are not negligible.

In Fig. 6 the polarizations of the incident and scattered waves are vertical (parallel). The roughness parameter $\beta = 394$, and the mss varies from 0.25 to 1. Both β and the mss are dimensionless quantities, since the slope is

dimensionless and $k_0 = 2\pi/\lambda$ is measured in units of inverse meters. The double-scatter cross sections are more sensitive to variations in mss than to those in β . Unlike the single-scatter cross sections, the double-scatter cross sections increase rather than decrease as mss increases, and they begin to merge for $mss > 0.75$. The level of the enhanced backscatter increases as both the roughness parameter β and mss increase. However, they merge for $\beta > 100$ and $mss > 0.75$. The angular width of the backscatter cross sections decreases as β increases; however, the angular width is practically insensitive to variations in the mss for the range of parameters considered.

In Fig. 7 all four like- and cross-polarized double-scatter cross sections are plotted as functions of the scatter angles $\theta^f \cos \phi^f$, $\phi^f = 0$. The level and the angular width of the backscatter enhancement for all four cross sections are practically the same, except that the values of the cross-polarized cross sections decrease somewhat faster as $\theta^f \rightarrow 90^\circ$ (grazing angles). For the illustrative examples, the first and second letters ($P = V, H$; $Q = V, H$) correspond to the polarizations of the scattered and incident waves, respectively.

In Figs. 8–10, the total (single+double) scatter cross sections are plotted as functions of the scatter angles. These three illustrations correspond to the cases considered in Figs. 2–4 and 5–7. In Fig. 8 the total cross sec-

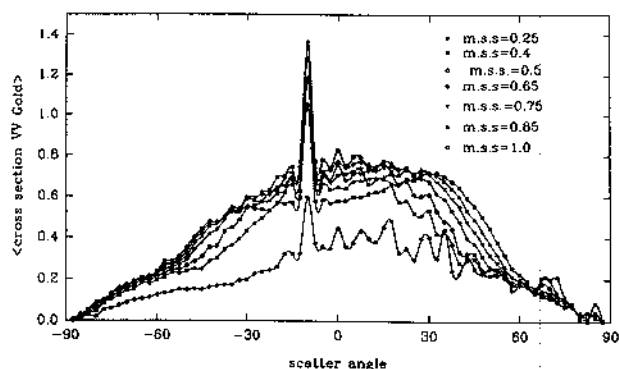


Fig. 6. Double-scatter radar cross section. The variable parameter is the mss (quasi-parallel and quasi-antiparallel contributions added). The polarization is vertical to vertical. The value of β is 394. Lowest curve, $mss = 0.25$. Data saturate as mss increases.

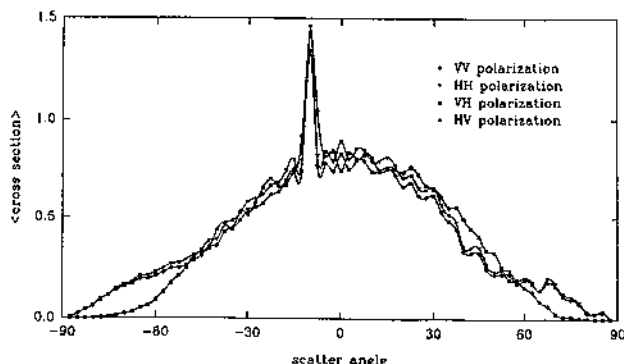


Fig. 7. Double-scatter radar cross sections for both like-polarized and both cross-polarized cases (quasi-parallel and quasi-antiparallel contributions added). The mss and β values are 0.5 and 394.105, respectively. Cross-polarized data are lower near grazing angles.

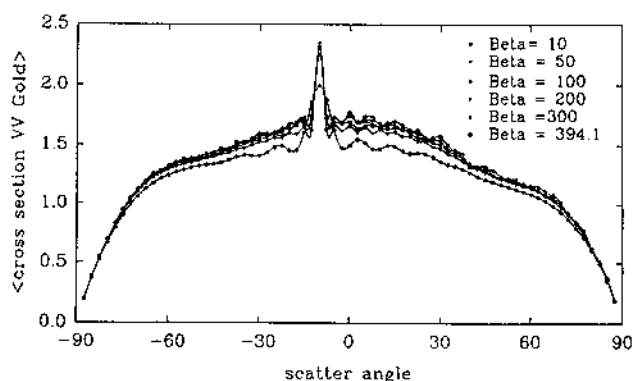


Fig. 8. Single- and double-scatter radar cross sections added. The variable parameter is β . The polarization is vertical to vertical. The mss value is 0.5. Lowest curve, $\beta = 10$. Data saturate as β increases.

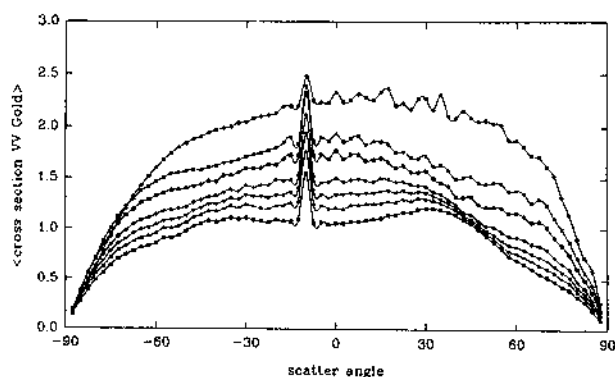


Fig. 9. Single- and double-scatter radar cross sections added. The variable parameter is the mss , and the curves from top to bottom correspond to mss values of 0.25, 0.4, 0.5, 0.65, 0.75, 0.85, and 1.0. The value of β is 394. The polarization is vertical to vertical.

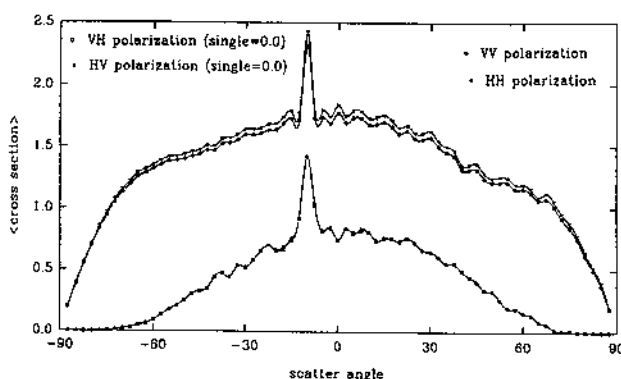


Fig. 10. Single- and double-scatter radar cross sections added for both like-polarized and both cross-polarized cases. The mss and β values are 0.5 and 394, respectively. Upper curves, HH followed by VV. Lowest curves, VH and HV merge.

tion, for the vertically polarized case, is plotted with β as the variable parameter, from $\beta = 10$ to $\beta = 394$, while $mss = 0.5$. The total scatter cross sections increase as β increases, and the results merge for $\beta > 50$. The angular width of the enhancement decreases as β increases. As noted above, the results saturate for $\beta \gg 1$.

In Fig. 9 the total cross sections for the vertically polarized case are plotted as functions of the scatter angles, with mss as the variable parameter, from $mss = 0.25$ to

$mss = 1.0$. For the cases considered, the total cross sections decrease as the mss increases because the contributions of the single-scatter cross sections dominate the results. The level of the backscatter enhancement increases as mss increases. However, the angular width of the enhancement does not vary significantly with changing mss . Also, as noted above, the results saturate for mss greater than unity.

In the last illustration, Fig. 10, the total scatter cross sections are plotted for all four (like and cross) polarizations. The fixed parameters are $\beta = 394$ and $mss = 0.5$. The like-polarized cross sections (VV and HH) are significantly larger than the corresponding cross-polarized cross sections (VH and HV). This is because the stationary-phase (high-frequency) approximations for the cross-polarized, single-scatter cross sections are negligible for scatter angles in the plane of incidence.^{1,4,5} At the specular points in the plane of incidence ($\phi^f = 0, \pi$), the scattered waves are not depolarized.⁸ For scattering out of the incident plane, the scattered waves do depolarize. This results in nonvanishing cross-polarized, double-scatter cross sections, even in the plane of incidence.⁹ There are two mechanisms for the generation of the cross-polarized, double-scatter cross sections. Schematically, these are given by

$$\begin{aligned} V \rightarrow V \rightarrow H, & \quad V \rightarrow H \rightarrow H, \\ H \rightarrow H \rightarrow V, & \quad H \rightarrow V \rightarrow V. \end{aligned}$$

In the above the first letter corresponds to the polarization of the incident wave, the second is the polarization of the single-scatter wave, and the third is the polarization of the double-scatter wave.

4. CONCLUSIONS

The results for the double-scatter radar cross sections exhibit sharp enhancements in the backscatter direction at normal and oblique incident angles. This sharp enhancement is associated with the quasi-antiparallel double-scatter path. The height and the width of the peak in the backscatter direction depend on the mean square height and slope of the two-dimensional random rough surface. The high-frequency approximations make the computations numerically tractable; however, the polarization dependence of the cross section is less obvious.^{1,4}

ACKNOWLEDGMENTS

This work was supported partially by the Department of Electrical Engineering and the Center for Electro-Optics at the University of Nebraska-Lincoln and the Supercomputer facility at Cornell University. The manuscript was prepared by M. Jane Craig.

Address correspondence to Ezekiel Bahar at the location on the title page or by e-mail, bahar@dragon.unl.edu.

REFERENCES AND NOTES

1. E. Bahar and M. El-Shenawee, "Enhanced backscatter from one dimensional random rough surfaces—stationary phase approximations to full wave solutions," *J. Opt. Soc. Am. A* **12**, 151–161 (1995).
2. E. Bahar and Y. Zhang, "A new unified full wave approach to evaluate the scatter cross sections of composite random rough surfaces," *IEEE Trans. Geosci. Remote Sens.* **34**, 973–980 (1996).
3. M. I. Sancer, "Shadow-corrected electromagnetics scattering from a randomly rough surface," *IEEE Trans. Antennas Propag.* **AP-17**, 577–585 (1969).
4. E. Bahar and M. El-Shenawee, "Vertically and horizontally polarized diffuse double scatter cross sections of one dimensional random rough surfaces that exhibit enhanced backscatter—full wave solutions," *J. Opt. Soc. Am. A* **11**, 2271–2285 (1994). An extensive list of references on backscatter enhancement is published here.
5. M. El-Shenawee and E. Bahar, "Double scatter cross sections for two dimensional random rough surfaces—high frequency approximation," presented at the 1995 IEEE AP-S International Symposium and UNC/URSI Radio Meeting, Newport Beach, Calif., June 18–23, 1995.
6. A. A. Maradudin and E. R. Méndez, "Enhanced backscattering of light from weakly rough, random metal surfaces," *Appl. Opt.* **32**, 3335–3343 (1993).
7. A. Ishimaru and J. S. Chen, "Scattering from very rough surfaces based on the modified second order Kirchhoff approximation with angular and propagation shadowing," *J. Acoust. Soc. Am.* **88**, 1877–1883 (1990).
8. E. Bahar, E. M. Herzinger, and M. A. Fitzwater, "Incoherent like and cross polarized cross sections of an anisotropic rough sea surface with swell," *J. Geophys. Res. [Oceans]* **94**, 2159–2169 (1989).
9. M. El-Shenawee and E. Bahar, "Double scatter radar cross sections for two dimensional random rough surfaces that exhibit backscatter enhancement," presented at the meeting of the Applied Computational Electromagnetic Society, Monterey, Calif., March 18–22, 1996.

Electromagnetic Scattering-Based Array Processing Methods for Near-Field Object Characterization*

Adnan Şahin and Eric L. Miller[†]
Center for Electromagnetics Research,
235 Forsyth Building, 360 Huntington Ave.
Northeastern University, Boston, MA 02115
Telephone: (617) 373-8386
Telefax : (617) 373-8627
E-mail: elmiller@cdsp.neu.edu

December 14, 1998

Abstract

We consider the use of high resolution array processing methods for detecting and localizing near-field extended targets for measurement geometries in which an array of electromagnetic receivers observes the fields scattered by the objects in response to a plane wave illumination. The algorithms presented here modify the conventional direction finding array processing techniques and use the spatial complexities of fields to determine both the range and the bearing of the targets in the region of interest. In contrast to previous source localization problems, we employ electromagnetic scattering models parameterized explicitly in terms of the target positions and which account for all multiple scattering effects.

These models play an integral role in allowing us to compute explicitly the analytical expressions for the Cramér-Rao bounds (CRB) for position estimates. The Cramér-Rao bound gives the lower bound for the estimates, and thus, specifies the lowest possible error variance that can be attained with an unbiased estimator. The theoretical Cramér-Rao bounds are then verified using Monte-Carlo simulated error variances.

Running title: EM-Based Near-Field Array Processing

*This work was supported in part DOE contract DE-FC07-95ID13395 and Army Research Office Demining MURI under Grant DAAG55-97-1-0013

[†]Communicating author

1 Introduction

Non-invasive detection and localization of objects in the near field of a receiver array have been of interest to many researchers in recent years. Some of the most promising application areas for this technology include landmine remediation, where relatively small metallic or plastic objects are located a few centimeters from the sensors, and hazardous waste remediation, where relatively large metallic objects (eg. steel metal drums) are located on the order of meters from the sensor array [1–3]. In this paper we consider a form of this problem shown in Fig. 1. A plane wave illuminates the region of interest assumed to be a homogeneous, possibly lossy medium containing one or more targets located in the near field of an array of receivers. The goal of the processing is the localization of objects with known structures.

The inherent array structure of the measurement geometry suggests that high resolution array processing techniques [4, 5] quite popular in the signal processing community would be well suited for the near-field detection problem. Typically, these techniques assume the targets are in the far field of the array so that the measured wavefronts are all planar in nature. The goal then is to determine the directions of arrival (DOA) of these wavefronts to characterize the bearing angles of the associated targets. In the case of near field target localization however, we must process scattered field data with distinctly non-planar wavefronts to determine both the range and the bearing angles of the objects. For far field range and bearing angle estimation problems, the matched field processing (MFP) approach has been successfully used for localization of typically point sources in ocean acoustics [6] and stratospherical electromagnetics [7]. MFP is an array processing technique that uses the spatial complexities of the fields to localize sources, and thus allows for estimation of both the range and the bearing of the objects.

In terms of near-field source localization, previous work has been concentrated on independently

radiating point sources [8–12] and the localization of a single extended object of unknown size and material [13]. The coherent source issue in near field localization is also dealt with in [9]. In [8,9], independent, spherical sound and noise sources are located using MUSIC based array processing methods. In [10] and [11], a similar near-field source localization scenario is considered, and performance bounds for coordinate estimates were derived for MUSIC and higher order ESPRIT based algorithms, respectively. Localization of distributed sources is reported in [14] where Valaee et al. introduced a MUSIC-based algorithm that parametrically localizes far field sources modeled with bell shaped distributions characterized by their peak position and 3 dB beamwidth. Finally, in [13] the authors employ a T-matrix type forward model as the basis for a non-linear least squares parameter estimation approach to determine the location, size, and material properties of a single object in the field of view of an array.

We present two MFP-based algorithms that localize multiple, extended scatterers placed in the near field of a receiver array. Furthermore, instead of independently radiating point sources, or sources with predetermined distribution characteristics, the radiators in this paper scatter the incident plane wave, and the physics of the multiple scattering is exactly accounted for using full scattering models. We have also determined geometries for which simpler forward scattering models can be utilized to reduce the computational complexity of the localization algorithm. At the end, we derive and verify the Cramer-Rao performance bounds for multiple, extended, near field objects.

We begin by describing a simple modification of the MUSIC algorithm in which the electromagnetic interactions between targets are ignored, and the problem is approached as if a number of spatially extended objects are independently scattering the incident electromagnetic field. While the resulting method has relatively low computational complexity, there are two issues in need of some analysis to characterize its performance. First, there is a signal coherence issue arising

from our frequency domain formulation of the problem. Second, it is necessary to determine in a quantitative manner conditions under which one may safely ignore the interactions among the scatterers. Both of these issues are addressed in Section 3.2.

In Section 3.3 we describe a second option for localizing multiple objects in which all multiple scattering effects are taken into consideration. While this approach is more computationally costly than the former, by modeling the non-trivial interactions of wavefields among scatterers, we improve our ability to resolve closely spaced scatterers. We demonstrate that this technique can easily localize the targets in geometries where the former approach fails to resolve the target positions.

Finally, we derive the Cramér-Rao bounds on variances of estimation errors for multiple object detection scenario. Our Cramér-Rao bound derivation adapts the results in [15], and accounts for near field observations. Analytical bounds of estimated object coordinates are then validated by running Monte-Carlo experiments for the estimator presented in this paper.

While our long term interests are in application of these ideas to the localization of buried objects (i.e. targets located in a halfspace), in this work, we concentrate on the simpler problem of localization when the objects are embedded in a homogeneous medium. It is our intent that the insight and experience gained from studying the homogeneous medium case will be of use when considering the more complicated half-space problem. Moreover, by considering this simpler physical problem, we are able to make extensive use of the T-matrix method [16–18] both in the localization algorithms as well as in the performance analysis. Indeed, the analytical structure of the T-matrix approach provides for the efficient computation of certain gradient information required for the method of Section 3.3 and for computing the Cramer-Rao bound in Section 4. Given the strong results in this paper, an interesting and non-trivial avenue of future research is the adaptation of the single object, half-space T-matrix approach of [19] to the multi-object

problem of interest here and the use of this forward model in an MFP-type inversion algorithm.

The remainder of the paper is organized as follows. In Section 2, we introduce the problem geometry and describe plane wave MUSIC. Section 3 is devoted to an exposition of the MFP-based algorithms. The Cramér-Rao lower bound is derived in Section 4. Numerical examples are presented in Section 5, and in Section 6, conclusions will be drawn.

2 Background

The measurement scheme depicted in Fig. 1 is considered in this paper. The objects are located in a background for which the constant electrical characteristics (relative permittivity and conductivity) are assumed known. A transverse magnetic (TM) polarized plane wave, $E_i(\mathbf{r})$, impinges on the objects, inducing surface and volume currents which in turn radiate a scattered field, $E_s(\mathbf{r})$ ¹. In this work, we are interested only in the object localization problem, thus material properties, shapes and the number of objects are assumed known.

The scattered electric field from N objects is spatially sampled by a uniformly spaced, linear array with M isotropic receivers, $M > N$. When the multiple scattering effects are taken into consideration, the measured data at the sensor outputs can be written as:

$$\mathbf{y} = \mathbf{A}_{ms}\mathbf{x} + \mathbf{n}, \quad (1)$$

where $\mathbf{A}_{ms} = [\mathbf{E}_{s1}(\mathbf{r}_1, \dots, \mathbf{r}_N) \ \mathbf{E}_{s2}(\mathbf{r}_1, \dots, \mathbf{r}_N) \ \dots \ \mathbf{E}_{sN}(\mathbf{r}_1, \dots, \mathbf{r}_N)]$ and $\mathbf{E}_{si}(\mathbf{r}_1, \dots, \mathbf{r}_N)$ denotes the scattered field observed at the array due to the i th object, in the presence of all other $(N - 1)$ scatterers, $i = 1, 2, \dots, N$. For time domain applications, the vector \mathbf{x} contains the narrow band time variations. Since we do our analysis in the frequency domain and suppress $e^{j\omega t}$, $\mathbf{x} = [1 \ 1 \ \dots \ 1]^T$. The i th column of matrix \mathbf{A}_{ms} depends not only on the position of the i th object, but also that of $(N - 1)$ other objects. Therefore, we may replace $\mathbf{A}_{ms}\mathbf{x}$ in (1) with $M \times 1$ vector

¹All analysis is in the frequency domain, thus the $e^{j\omega t}$ time dependence will be dropped.

$\mathbf{B} = \mathbf{E}_s(\mathbf{r}_1, \mathbf{r}_2 \cdots \mathbf{r}_N) = \mathbf{E}_{s1}(\mathbf{r}_1, \dots, \mathbf{r}_N) + \mathbf{E}_{s2}(\mathbf{r}_1, \dots, \mathbf{r}_N) + \dots + \mathbf{E}_{sN}(\mathbf{r}_1, \dots, \mathbf{r}_N)$, which denotes the total scattered field at the receiver array.

The recursive T matrix algorithm [16–18] is used to calculate the exact scattered field \mathbf{E}_{si} due to the i th object in the presence of other objects. The algorithm reported in [18] is designed for the efficient solution of near field scattering problems with heterogeneous collection of metallic and dielectric objects. In addition, using the recursive T-matrix algorithm the solution to the scattering problem can be written in a closed form. This property is especially useful in obtaining analytical expressions for gradients and derivatives in multi-dimensional parameter search in Section 3.3 and in Cramér-Rao bound analysis in Section 4.

Electromagnetic interactions between the objects may be ignored in favor of reducing the computational complexity of the localization process. For the method presented in Section 3.2, we use the same data model as in (1), except that the matrix \mathbf{A}_{ms} is replaced by

$$\mathbf{A}_{ss} = [\mathbf{E}_s(\mathbf{r}_1) \ \mathbf{E}_s(\mathbf{r}_2) \cdots \mathbf{E}_s(\mathbf{r}_N)]. \quad (2)$$

The $M \times 1$ vector $\mathbf{E}_s(\mathbf{r}_i)$ denotes the scattered field due to a single object located at \mathbf{r}_i , $i = 1, 2, \dots, N$, and it is calculated using the Mie series [20]. Therefore, each column of \mathbf{A}_{ss} contains the scattered field that is a function of position of only one particular object. In fact, in this definition of matrix \mathbf{A}_{ss} , each column is equivalent to those of \mathbf{A}_{ms} in (1) when all other scatterers are infinitely far away. The vector \mathbf{x} is the same as before, $\mathbf{x} = [1 \ 1 \cdots 1]^T$.

2.1 Review of Plane wave MUSIC

For plane wave MUSIC, the data model is in the following form:

$$\mathbf{y} = \mathbf{A}_{pw}\mathbf{x} + \mathbf{n}, \quad (3)$$

where $\mathbf{A}_{pw} = [\mathbf{a}(\theta_1) \ \mathbf{a}(\theta_2) \ \cdots \ \mathbf{a}(\theta_N)]$, and $\mathbf{a}(\theta_i) = [1 \ e^{jkd \cos \theta_i} \ e^{j2kd \cos \theta_i} \ \cdots \ e^{j(M-1)kd \cos \theta_i}]^T$.

Here $\mathbf{a}(\theta_i)$ is known as the direction vector, θ_i is the direction of arrival (DOA) of the i th plane

wave, k is the wave number in the medium of propagation, and d is the distance between two sensors. The experiment as represented by the data model in (3) is repeated many times in order to determine the statistics of \mathbf{y} . In particular, if L experiments are performed, then the maximum likelihood estimate of the spatial autocovariance matrix $\mathbf{R} = E\{\mathbf{y}\mathbf{y}^H\}$ is given by [4]:

$$\hat{\mathbf{R}} = \frac{1}{L} \sum_{l=1}^L \mathbf{y}_l \mathbf{y}_l^H \quad (4)$$

where \mathbf{y}_l is the data measured at the l th experiment, and superscript H denotes the complex conjugate transpose. The sample covariance matrix $\hat{\mathbf{R}}$ is used in MUSIC algorithm to separate signal and noise subspaces through the eigenspace decomposition [4]:

$$\hat{\mathbf{R}} = \hat{\mathbf{U}}_s \hat{\Lambda}_s \hat{\mathbf{U}}_s^H + \hat{\mathbf{U}}_n [\hat{\sigma}^2 \mathbf{I}] \hat{\mathbf{U}}_n^H \quad (5)$$

where $\hat{\mathbf{U}}_s$ is the estimated signal subspace matrix and contains the N signal eigenvectors, and $\hat{\mathbf{U}}_n$ is the estimated noise subspace matrix and contains $M - N$ noise eigenvectors of multiple noise eigenvalue $\hat{\sigma}^2$. The projection operator onto the noise subspace is defined as [4]:

$$\hat{\Pi}_n = \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H. \quad (6)$$

The basic idea behind the planewave MUSIC algorithm is that the reciprocal of the “distance” between the estimated noise subspace and the true noise subspace has sharp peaks around the DOAs. Thus, if one plots this quantity versus all possible angles, estimates of DOAs can be determined by the maxima of the angular spectrum. The spatial spectrum of the MUSIC algorithm is given by [4]:

$$P_{MUSIC}(\theta) = \frac{\mathbf{a}(\theta)^H \mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \hat{\Pi}_n \mathbf{a}(\theta)} \quad (7)$$

where $\mathbf{a}(\theta) = [1 \quad e^{jkd \cos \theta} \quad e^{j2kd \cos \theta} \dots e^{j(M-1)kd \cos \theta}]^T$ is the direction vector that accounts for a plane wave impinging on the array from the direction θ .

It is important to realize that the formulation of the array processing problem presented in this section implicitly assumes that the radiator is infinitely distant so that the scattered field has planar wavefronts at the sensor array. Thus, the elements of the direction vector $\mathbf{a}(\theta)$ are complex

exponentials indicative of plane wave signals. However in many applications, including near field object detection, the receivers are in the near-field region of the radiating sources, resulting in non-planar wavefronts. The target localization problem, therefore, not only requires the DOA relative to the array but also the range of the target from a point on the array.

3 Matched Field Processing for Object Localization

As mentioned in the introduction, matched field processing [6] uses the spatial complexities of the fields to localize sources in underwater acoustics. In a similar manner, plane wave MUSIC outlined in Section 2 can be modified so that the direction vector is filled with the type of the wavefront impinging on the array [2, 3, 8–10]. We divided this section into three parts. In the first part we will describe the single object localization. Even though single object localization is a subset of multiple object localization, we present it separately in order to explain the multiple object case clearly. The second part will deal with multiple objects, but the electromagnetic interaction between objects will be ignored. In the last part, we will treat multiple object detection problem when the multiple scattering effects are taken into account.

3.1 Single Object Localization

By using the spatial distribution of the scattered field, we can fill the direction vector in plane wave MUSIC with non-planar scattered fields to locate the near field objects. By modifying the spatial MUSIC spectrum in (7), we form the following spectrum:

$$P_{MUSIC}(r, \theta) = \frac{\mathbf{E}_s(r, \theta)^H \mathbf{E}_s(r, \theta)}{\mathbf{E}_s(r, \theta)^H \hat{\Pi}_n \mathbf{E}_s(r, \theta)} \quad (8)$$

where the new “direction” vector (actually, the *locus vector* in r and θ space) $\mathbf{E}_s(r, \theta)$ is now filled with the scattered field observed for an object located at $\mathbf{r} = (r, \theta)$. Then, the location $(\hat{r}, \hat{\theta})$ maximizing the MUSIC spectrum in (8) is selected as the estimated object center.

Because a two dimensional search requires that the exact scattered field be calculated at each point of the parameter mesh, this technique is in general computationally intensive. When the objects to be detected are modeled as simple shapes, computing the exact scattered field can be relatively simple. Here we consider the case of infinitely long cylinders so that the scattered field due to a plane wave can be calculated using the Mie series:

$$E_s(\rho, \phi) = \sum_{n=-\infty}^{\infty} c_n H_n^{(2)}(k\rho) e^{jn(\phi + \pi - \phi_{inc})} \quad (9)$$

where $H_n^{(2)}(.)$ is the Hankel function of the second kind of order n representing cylindrical outgoing waves, k is the wavenumber in the homogeneous, possibly lossy, background, and ϕ_{inc} is the incidence angle of the plane wave. The coefficients c_n are determined from the boundary conditions when ρ is equal to the radius of the cylinder. For computer implementations, the infinite sum in (9) is truncated at a finite value beyond which the coefficients c_n are below machine precision. Here ρ and ϕ denote the coordinates of the receivers since (9) assumes that the center of the cylinder is located at the origin. In implementing (8), translations from object-to-receiver coordinate system (ρ, ϕ) to array-to-object-position coordinate system (r, θ) are required. These translations do not significantly alter the computational load or functional form implied by (9).

Fig. 2(b) shows the MUSIC spectrum for a localization scenario when a small metallic object in 7.5 cm diameter is placed 15 cm below the array as depicted in Fig. 2(a). The 33-element linear, uniform receiver array spans an aperture of 1.5 meters. All sensors are assumed to be ideal, isotropic receivers. The operating frequency is 1.0 GHz and the plane wave is incident with 90 degrees. The lossy, homogeneous background has the same electrical characteristics of 5% moist San Antonio clay loam or 10% moist Puerto Rico clay loam ($\epsilon_b = 6\epsilon_0$, $\sigma_b = 5 \times 10^{-2} S/m$) at around 1.0 GHz [21]. The signal to noise ratio (SNR) is fixed at 0 decibels. As Fig. 2(b) depicts, the location of the object ($r = 15cm$, $\theta = 90^\circ$) is indicated with a very sharp peak. We note that

the prominent peak structure of the spectrum in this example is representative of a wide range of other cases for which the background electrical properties, target electrical characteristics and the signal to noise ratios are varied [2].

3.2 Multiple Object Localization: No Interaction

In this section, we describe multiple object localization assuming that the objects are located sufficiently far apart, so that the interactions among them can be ignored. For this case, multiple object localization is equivalent to having a superposition of cylindrical waves of the form in (9) incident on the array. Since the interactions are ignored, the scattered field due to one object is independent of the positions of the other objects, and thus we can use the single object localization approach described in previous section to search for multiple peaks in MUSIC spectrum given in (8) to determine positions of multiple objects. As in Section 3.1, the scattered field due to a single scatterer is calculated using the Mie series in (9), and a two-dimensional search is carried out to find N peaks corresponding to estimated target locations $(\hat{r}_1, \hat{\theta}_1), (\hat{r}_2, \hat{\theta}_2), \dots, (\hat{r}_N, \hat{\theta}_N)$.

The use of MUSIC in this manner raises two issues which are considered in the following paragraphs. First, because we are operating in the frequency domain, there are signal coherence difficulties which we address through the use of frequency diversity. The second issue is the need to develop a quantitative means of understanding the circumstances under which the no-interaction approximation is valid. In analyzing both of these issues, we present closed form analytical results valid for near field objects of infinitesimal radius and verify through numerical calculation that the insight provided by these expressions carries over for objects of finite size.

Because of the structure of \mathbf{x} in (1), the rank of $\mathbf{P} = E\{\mathbf{x}\mathbf{x}^H\}$, is one and so is the rank of \mathbf{R} . Thus, the signals impinging on the array are coherent. To increase the rank of \mathbf{R} (i.e. to decorrelate the signals) we repeat the scattering experiment as represented by (1) at N different frequencies

where N is the number of objects whose locations are to be determined. Thus (4) becomes:

$$\hat{\mathbf{R}}_i = \frac{1}{L} \sum_{l=1}^L \mathbf{y}_{li} \mathbf{y}_{li}^H \quad (10)$$

where \mathbf{y}_{li} is the l th data vector observed at the i th operating frequency, f_i , and $\hat{\mathbf{R}}_i$ is the spatial autocovariance matrix at f_i , $i = 1, 2, \dots, N$. Then, the rank enhanced autocovariance matrix $\hat{\mathbf{R}}_d$ is obtained by

$$\hat{\mathbf{R}}_d = \frac{1}{N} \sum_{i=1}^N \hat{\mathbf{R}}_i. \quad (11)$$

In Appendix A, we prove for $N = 2$ that in the limiting case of infinitesimal object radius, frequency diversity does, in fact, increase the rank of the autocovariance matrix $\hat{\mathbf{R}}_d$ to two. In addition, for objects with finite radii, we can computationally demonstrate that using frequency diversity we obtain a full rank autocovariance matrix. Having decorrelated the signals, the projector onto the noise subspace, $\hat{\Pi}_n$ in (8), is found from eigendecomposition of $\hat{\mathbf{R}}_d$.

While neglecting the interactions reduces the computational demand of localization considerably, it may introduce large estimation errors if the interactions are in fact sizeable. To analyze this issue we consider a scenario in which two objects are located a fixed distance beneath a receiver array and separated by a distance d .² We define the following normalized interaction term to measure the relative import of the multiple scattering between the two objects

$$\delta(d) = \frac{\|\mathbf{E}_{s1}^{(2)} - \mathbf{E}_{s1}\|_2^2}{\max_{d \in [d_{min}, d_{max}]} \|\mathbf{E}_{s1}^{(2)} - \mathbf{E}_{s1}\|_2^2} \quad (12)$$

where $\mathbf{E}_{s1}^{(2)}$ is the scattered field vector observed at the array due to the first object in the presence of the second object, \mathbf{E}_{s1} is the scattered field from the first object when there are no other scatterers, and d_{min} and d_{max} are limits over which we want to perform the analysis as dictated by the nature of the underlying application. In Appendix B we provide a closed form expression for δ in the case of two infinitesimally thin scatterers and show in Section 5.1, that (a) the δ for filamentary

²While not the most general setup, this configuration is representative of the types of application of interest in this paper (eg. mine detection) in which at most a couple of objects are in the array's field of view at about the same depth. Moreover, this simplified problem provides insight which may well be of use in more complex situations.

objects bounds those for the extended objects for a wide range of distances and (b) the error in the localization procedure tracks closely the value of δ . Thus, the expression in Appendix B can be used to determine the degree of interaction between two scatterers in terms of spacing and electrical properties of medium of propagation. By selecting a desired interaction level, we may calculate the minimum distance between scatterers that satisfy this a priori level. If for a given scenario we suspect that the actual distance is smaller than the minimum distance, then we should not ignore the interactions, and use the technique in Section 3.3 to localize the objects.

3.3 Multiple Object Localization: With Interaction

In this section, we describe multiple object localization using MUSIC when the electromagnetic interactions between objects are completely taken into consideration. In a sense, this is the multi-dimensional equivalent of single object detection where we aggregate all scatterers into one, large scatterer whose electrical characteristics are defined by the $2N$ co-ordinates of the individual objects. For object localization, then, we form the following MUSIC spectrum:

$$P_{MUSIC}(r_1, \theta_1, r_2, \theta_2, \dots, r_N, \theta_N) = \frac{\mathbf{E}_s(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)^H \mathbf{E}_s(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)}{\mathbf{E}_s(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)^H \hat{\Pi}_n \mathbf{E}_s(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)} \quad (13)$$

where $\mathbf{E}_s(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ denotes the total scattered field due to objects located at $\mathbf{r}_i, i = 1, 2, \dots, N$ and is filled using the recursive T-matrix algorithm [16–18]. In order to find the positions of the objects, then, we perform a $2N$ -dimensional search in location space of all objects, $(r_1, \theta_1), (r_2, \theta_2), \dots, (r_N, \theta_N)$. The coordinates $(\hat{r}_1, \hat{\theta}_1), (\hat{r}_2, \hat{\theta}_2), \dots, (\hat{r}_N, \hat{\theta}_N)$ at which the spectrum (13) reaches maximum give us the estimated target locations. Finally we note that since there is only a single signal vector defined by the position of all the objects, we have no signal coherence issue here.

The computational complexity of this approach is considerably larger than the no interaction case for two inter-related reasons: the repeated use of the forward scattering model many times for the $2N$ -dimensional search, and the extra cost of taking electromagnetic interactions into account

in the exact forward model. To keep the computational load at reasonable levels, we used the recursive T-matrix algorithm given in [18] to calculate the scattered fields due to multiple objects. This variant of T matrix algorithm is specifically designed for the efficient solution of near field scattering problems involving heterogeneous collections of metallic and dielectric objects. Finally, to further reduce the computational load, we employ a variable step-size steepest ascent approach to maximize (13) which, for the problems of interest here typically converges to the neighborhood of the maximum in 7-10 iterations after which point we use a less costly simplex search [22] to locate the exact maximum. Explicit, closed form expressions for the gradient of the MUSIC spectrum with respect to the target coordinates are provided through the use of the T matrix forward model. The associated long and tedious formulae are provided in [23].

4 Cramér-Rao Performance Bounds on Object Localization

The Cramer-Rao Bound (CRB) provides very valuable information about the lower limit for the variance of any unbiased estimator. In order to find CRB, however, one should have a closed form expression of the log-likelihood function. In this section, we will extend the results in [15] to find the Cramér-Rao bounds for the near field, multiple object detection geometries. Since the additive noise in (1) is white and Gaussian, the log-likelihood function can be written as [15]:

$$\ln \mathcal{L} = \text{constant} - 2ML \ln \sigma - \frac{1}{\sigma^2} \sum_{l=1}^L [\mathbf{y} - \mathbf{A}_{ms} \mathbf{x}]^H [\mathbf{y} - \mathbf{A}_{ms} \mathbf{x}] \quad (14)$$

where σ^2 is the noise variance, M is the number of receivers and L is the number of data vectors used for estimating the covariance matrix $\hat{\mathbf{R}}$ in (4).

Given the log-likelihood function, the Fisher Information Matrix (FIM) can be written as:

$$\mathbf{J} = E\{\psi\psi^T\} \quad (15)$$

where $\psi = [\frac{\partial \ln \mathcal{L}}{\partial r_1} \frac{\partial \ln \mathcal{L}}{\partial \theta_1} \dots \frac{\partial \ln \mathcal{L}}{\partial r_N} \frac{\partial \ln \mathcal{L}}{\partial \theta_N}]^T$. The FIM is then expressed in matrix form as:

$$J = \begin{bmatrix} \Gamma_{r_1 r_1} & \Gamma_{r_1 \theta_1} & \cdots & \Gamma_{r_1 r_N} & \Gamma_{r_1 \theta_N} \\ \Gamma_{\theta_1 r_1} & \Gamma_{\theta_1 \theta_1} & \cdots & \Gamma_{\theta_1 r_N} & \Gamma_{\theta_1 \theta_N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Gamma_{r_N r_1} & \Gamma_{r_N \theta_1} & \cdots & \Gamma_{r_N r_N} & \Gamma_{r_N \theta_N} \\ \Gamma_{\theta_N r_1} & \Gamma_{\theta_N \theta_1} & \cdots & \Gamma_{\theta_N r_N} & \Gamma_{\theta_N \theta_N} \end{bmatrix} \quad (16)$$

where $\Gamma_{pq} = E[\frac{\partial^2 \ln \mathcal{L}}{\partial p \partial q}]$, $\{p, q\} = \{r_1, \theta_1, r_2, \theta_2 \dots r_N, \theta_N\}$. The entries of FIM are [15]:

$$\Gamma_{pq} = \frac{2}{\sigma^2} \sum_{l=1}^L \text{Re}\{\mathbf{x}^H \mathbf{D}_p^H \mathbf{D}_q \mathbf{x}\}, \quad (17)$$

where $\mathbf{D}_p = \frac{\partial \mathbf{A}_{ms}}{\partial p}$ and $\mathbf{D}_q = \frac{\partial \mathbf{A}_{ms}}{\partial q}$. The columns of matrix \mathbf{A}_{ms} contain the scattered fields as defined in (1), and since the scattered fields are calculated using the recursive T-matrix algorithm, derivatives of \mathbf{A} with respect to object positions, \mathbf{D}_p and \mathbf{D}_q , can be easily obtained [23]. Since \mathbf{x} is constant over L experiments, (17) can be further reduced to:

$$\Gamma_{pq} = \frac{2L}{\sigma^2} \text{Re}\{\mathbf{x}^H \mathbf{D}_p^H \mathbf{D}_q \mathbf{x}\}. \quad (18)$$

The Cramér-Rao bound by definition is, then, the inverse of the FIM:

$$\text{CRB}(r_1, \theta_1, \dots, r_N, \theta_N) = J^{-1}. \quad (19)$$

The i th diagonal entry in the Cramér-Rao bound expression in (19) gives the Cramér-Rao lower bound for the i th variable in the parameter set $\{r_1, \theta_1, r_2, \theta_2 \dots r_N, \theta_N\}$. In Section 5, we will verify the analytical expressions given by (19) with the Monte-Carlo simulated error variances.

5 Examples

In this section, we present numerical examples on localization of multiple objects and verification of Cramér-Rao lower bounds with Monte-Carlo simulations. In order to simplify the scattering phenomenon associated with the detection problem, the targets are modeled as simple, circular objects with diameters of 7.5 cm. The system parameters are kept constant throughout the examples to allow meaningful comparisons. The scattered field due to an incident plane wave is observed

along a 33-element, uniform, linear receiver array which spans an aperture of 1.5 m. The sensors are assumed to be ideal, isotropic receivers. Unless otherwise noted, the operating frequency is set to 1.0 GHz. The objects are placed in a lossy, homogeneous background which has the same electrical characteristics of 5% moist San Antonio clay loam ($\epsilon_b = 6\epsilon_0$, $\sigma_b = 5 \times 10^{-2} \text{S/m}$) at around 1.0 GHz [21]. In all examples the scattering simulations are repeated 250 times to estimate the autocovariance matrices, i.e. $L = 250$ in (4) and (10).

In order to show the performance of the algorithms, we consider two objects geometries. In the first case (Fig. 3(a)), the objects are located quite far from each other. For this object geometry, due to the lossy background, the interactions between the targets are very weak. In the second case (Fig. 3(b)), the objects are located closely, and the electromagnetic interactions between the objects are considerably stronger.

For the simulations, the definition of signal to noise ratio (SNR) is not obvious. In practical problems, SNR is imposed by the nature of the system noise. However, in computer simulations we want to reference the noise power to a fixed quantity that does not change as the positions of the objects change. For this purpose SNR is referenced to the scattered field strength of a single, cylindrical, metallic object placed at the same depth as the objects, in the same lossy medium. The radius of the reference scatterer is the same as the radii of the targets. With this definition, the noise power is always proportional to the power of reference scattered field.

5.1 Multiple Object Localization: No Interaction

As described in Section 3.2, all scatterers are assumed to be scattering the incident plane wave independently. Therefore, we used the Mie series to generate the MUSIC spectrum in (8). The issue of signal coherence is solved by using two operating frequencies, 1.0 and 1.2 GHz. The autocovariance matrix at each frequency is estimated using (10), and the full rank autocovariance

matrix is calculated using (11). For a wide range of target locations, we have verified that the autocovariance matrix $\hat{\mathbf{R}}_d$ has two distinct signal eigenvalues that are quite different from the noise eigenvalues. The order of magnitude difference, of course, depends on the signal-to-noise ratio. At the 20 dB SNR level used in this example, the smallest signal eigenvalue is approximately two orders of magnitude larger than the largest noise eigenvalue.

In Fig. 4, we plot the MUSIC spectrum in (8) for the case of well separated objects. As the figure depicts, the target locations are indicated by two peaks that are easy to distinguish from the background. The spectrum for the closely packed object case is shown in Fig. 5. It is clear from the figure that the no-interaction approach fails for this case. This is expected, since the no interaction model ignores the multiscattering effects that are very strong for closely spaced scatterers.

Since ignoring electromagnetic interactions may result in estimation errors, or even prevent resolution of targets as in Fig. 5, we investigated the interaction between two extended objects and compared the results with filamentary objects. Consider a problem where two objects are located 15 cm below the array. The position of one is kept fixed at (125, -15) cm (i.e. about 1/4 of the way from the right edge of the array) while the second is moved from left to right such that the inter-object distance, d , varies from eight centimeters to one meter. The interaction term given by (12) for both extended objects (computed using the T matrix method) and the filamentary objects (obtained from (28)) are shown as a function of d in Fig. 6(a). It is clear from this figure that for both extended and infinitesimally thin objects, the interaction terms decay very similarly.

The link between estimation error and object separation is shown in Fig. 6(b). A comparison of Figs. 6(a) and (b) reveals the expected result that both estimation errors and interaction terms decay as the objects are located farther apart. Figs. 6(c) and (d) show the same comparisons in (a) and (b) when the objects are located 25 cm below the array, further indicating that the

simple interaction term of Appendix B clearly provides a good measure of the level of interaction between extended objects, and one may use this simple interaction criterion to decide whether the no-interaction case would result in acceptable estimation errors. In fact, we can use Figs. 6(a) and (b) to estimate how the no-interaction case would perform for geometries given in Fig. 3. When the objects are 1 meter apart, Fig. 6 indicates that the interaction term is insignificant, so is the estimation error. As seen from Fig. 4, for this case the object centers are clearly defined by two peaks around the true center coordinates. When the objects are separated by 10 cm, however, there is a considerable electromagnetic coupling between objects, and the estimation errors are so large that the objects centers cannot be resolved as verified in Fig. 5.

5.2 Multiple Object Localization: With Interaction

We have applied the algorithm given in Section 3.3 to localize the targets for the two geometries: well separated and closely spaced objects. The signal to noise ratio is fixed at 20 dB and the autocovariance matrix, and the projection operator onto the noise subspace is calculated.

To determine the positions of the two targets, we searched for the maximum of the spectrum $P_{MUSIC}(r_1, \theta_1, r_2, \theta_2)$ given by (13) in $(r_1, \theta_1, r_2, \theta_2)$ space. To find this maximum we use the steepest descend algorithm described in Section 3.3. Fig. 7 shows the intermediate object positions at each iteration during the multi-dimensional search. The objects are located 1 meter apart, and the initial guesses are denoted with *'s³. Large circles indicate the support of the objects. Although not clear from this figure, as the number of iterations increase, the location estimates get closer to the true estimates at (25, -15) and (125, -15) cm. Fig. 8 shows the object positions at each iteration during the multi-dimensional search when the objects are closely spaced. The

³Here we have initialized the search method by hand to illustrate the manner in which the search process functions. As described below, to verify the Cramer-Rao bound analysis, we employ an automatic method for initializing the nonlinear optimization.

distance between the targets for this case is 10 cm. Again the initial guesses are indicated with *'s, and as the number of iterations increase, the location estimates get closer to the true estimates at $(70, -15)$ and $(80, -15)$ cm. For both cases, we observed that this approach quickly converges to the general maximum area in 7-10 iterations using the steepest decent algorithm. After the initial quick convergence, we switch to a simplex search [22] to locate the precise maximum.

5.3 Verification of Cramér-Rao bounds

In this section, we will compare the analytical Cramér-Rao bound results obtained in Section 4 with the Monte-Carlo simulated error variances. For this purpose, the algorithms described in Sections 3.1 and 3.3 are repeated 500 times at different signal to noise ratios. For each Monte-Carlo simulation, the multi-dimensional search routine requires initial values for the positions of the targets. Thus, to supply the algorithm with an initial value, we used the subarray processing method described in [1]. This approach can efficiently provide rough estimates of the object locations. Once we initialize the object positions with subarray processing, we perform the multi-dimensional search described in the previous section. The position vector maximizing the spectrum is declared as the position estimates and the error variances are calculated from the estimates.

Fig. 9 shows the comparison of analytical CRB calculated for single object geometry of Fig. 2 using (19) with 500 Monte-Carlo simulations of the algorithm described in Section 3.1. The solid line is the CRB for the radial position of the object and measured in squared meters. The dashed line denotes the CRB for the angular position and measured in squared radians. The radial and angular positions are referenced to the center point of the array. Monte-Carlo simulated error variances for range and bearing variables are shown on the same plot with \circ and \diamond symbols, respectively.

Fig. 10 compares the Cramér-Rao bounds with simulated error variances for multiple object geometries of Fig. 3 using the algorithm of Section 3.3. Fig. 10(a) shows the comparison for two

objects located one meter from each other, and Fig. 10(b) show the comparison when the objects are 10 centimeters apart. Again, the solid and dashed lines are the CRBs for radial and angular positions of the objects, respectively. The range variables have the unit of squared meters, and the bearing variables have the unit of squared radians. As before, both coordinate variables are defined with respect to the center point of the array. The symbols \circ and \diamond show the Monte-Carlo simulated error variances for the range and bearing variables, respectively. For this case, since there are two targets present in the region, there are two symbols at each SNR value. The problem geometry is symmetric, thus, one would expect that the error variances should be the same. Indeed, for analytical Cramér-Rao bounds calculated from (19), the bounds for coordinates of both objects are the same. For the simulated error variances, however, we can notice the different values for the range variables in Fig. 10(a) and for the bearing variables in Fig. 10(b). The difference in simulated variances pronounced for low signal to noise ratios, which, we believe, implies that more Monte-Carlo simulations are needed at those SNR values.

In both Fig. 9 and Fig. 10, we observe that the simulated error variances approach the lower limit provided by the Cramér-Rao bounds as the signal to noise ratio increases. This is expected by the results reported in [15] which states that MUSIC is an efficient estimator for large SNR values. The results presented here, therefore, reveal that at high signal to noise ratios MUSIC is an efficient estimator for near field object localization problems, too.

6 Conclusions

In this paper we presented a matched field based high resolution array processing technique for localization of near field targets. The algorithm is presented in three parts: single object localization, multiple object localization ignoring the interactions between objects, and multiple object localization taking multiple scattering into account. When the interactions are ignored,

the computational load is alleviated since the electromagnetic model is a simple Mie series, and a two-dimensional search in parameter space suffices to localize all objects. The drawbacks, on the other hand, are the necessity to take care of the coherent signal issue, and the failure to resolve and localize closely spaced targets. When the electromagnetic interactions are accounted for, these drawbacks are remedied, but the computational load due to multi-dimensional search, and complex, multiple scattering forward model increases.

We have also calculated the analytical Cramér-Rao bound expressions for coordinates of multiple objects when the interactions are taken into consideration. These lower bounds are then verified with Monte-Carlo simulated error variances. We have shown that as the signal to noise ratio increases, simulated error variances approach the lower limit set by the Cramér-Rao bounds.

In terms of future work, our primary interest is the extension of this method to three dimensional buried objects for problems in which the halfspace between the air and the earth is explicitly modeled. This will require the generalization of the single object, half-space T matrix method of Kristensson and Strom [19] to the multi-object case as well as the development of MFP-type methods which are built around this new forward model.

Appendices

A Frequency Diversity

In this appendix, we will show that under simplifying assumptions the frequency diversity described by (10) and (11) increases the rank of the autocovariance matrix. Consider the data model $\mathbf{y}_i = \mathbf{s}_i + \mathbf{n}_i$ where \mathbf{s}_i is the noise-free scattered field due to a scatterer at frequency f_i , \mathbf{n}_i is the white, Gaussian noise as before, and \mathbf{y}_i is the observed data vector at frequency f_i , $i = 1, 2, \dots, N$. To simplify the problem, we will assume that the number of scatterers is limited

to two ($N = 2$), and the scatterers are infinitesimally thin.

The exact expression for the rank enhanced autocovariance matrix, \mathbf{R}_d is:

$$\mathbf{R}_d = \frac{1}{2} \mathbb{E} \{ \mathbf{y}_1 \mathbf{y}_1^H + \mathbf{y}_2 \mathbf{y}_2^H \} = \frac{1}{2} [\mathbf{s}_1 \mathbf{s}_1^H + \mathbf{s}_2 \mathbf{s}_2^H + 2\sigma^2 \mathbf{I}]. \quad (20)$$

For two scatterers, the frequency diversity should ensure that the signal subspace of \mathbf{R}_d is two dimensional. To see the conditions for which the signal subspace of \mathbf{R}_d is two dimensional, we rewrite (20) as:

$$\mathbf{R}_d = \underbrace{\frac{1}{2} [\mathbf{s}_1 \mathbf{s}_2] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{s}_1^H \\ \mathbf{s}_2^H \end{bmatrix}}_{\mathbf{S}} + \sigma^2 \mathbf{I}. \quad (21)$$

For the signal subspace of \mathbf{R}_d to be two dimensional, the matrix \mathbf{S} should have rank two which is possible if vectors \mathbf{s}_1 and \mathbf{s}_2 are linearly independent. Since the maximum likelihood estimate of the rank enhanced autocovariance matrix $\hat{\mathbf{R}}_d$ in (11) is asymptotically equal to \mathbf{R}_d , the proof for \mathbf{R}_d is valid for $\hat{\mathbf{R}}_d$ for large snapshots, i.e. $L \rightarrow \infty$. In the rest of the appendix, we show that \mathbf{R}_d has a two dimensional signal subspace, by proving that \mathbf{s}_1 and \mathbf{s}_2 are independent.

For this purpose, we show that the scattered fields due to the filament scatterers observed at two points, A and B , in space at two distinct frequencies, f_1 and f_2 , are independent. Thus, the data vectors measured over an array at more than two points will also be independent. The general Mie scattering series given in (9), is reduced to the following expressions for filament scatterers:

$$\begin{aligned} E_{s1}^A &= c_0 H_0^{(2)}(k_1 r_A) & E_{s2}^A &= d_0 H_0^{(2)}(k_1 r_A) \\ E_{s1}^B &= c_0 H_0^{(2)}(k_1 r_B) & E_{s2}^B &= d_0 H_0^{(2)}(k_2 r_B) \end{aligned} \quad (22)$$

where r_A and r_B denote the observation points, coefficients c_0 and d_0 are dependent on the frequency and object properties, $H_0^{(2)}(.)$ is the zeroth order outgoing Hankel function, and k_i is the wavenumber at f_i , $i = 1, 2$. We will show that the vectors

$$\mathbf{s}_1 = \begin{bmatrix} E_{s1}^A \\ E_{s1}^B \end{bmatrix} \quad \text{and} \quad \mathbf{s}_2 = \begin{bmatrix} E_{s2}^A \\ E_{s2}^B \end{bmatrix} \quad (23)$$

are independent, i.e. equations

$$E_{s1}^A = \alpha E_{s2}^A \quad \text{and} \quad E_{s1}^B = \alpha E_{s2}^B \quad (24)$$

cannot be satisfied as long as $f_1 \neq f_2$.

Equations (22) and (24) imply that:

$$c_0 H_0^{(2)}(k_1 r_A) = \alpha d_0 H_0^{(2)}(k_2 r_A) \quad \text{and} \quad c_0 H_0^{(2)}(k_1 r_B) = \alpha d_0 H_0^{(2)}(k_2 r_B) \quad (25)$$

which can be written as

$$\gamma = \frac{\alpha d_0}{c_0} = \frac{H_0^{(2)}(k_1 r_A)}{H_0^{(2)}(k_2 r_A)} = \frac{H_0^{(2)}(k_1 r_B)}{H_0^{(2)}(k_2 r_B)}. \quad (26)$$

Properties of Hankel functions require that $|H_0^{(2)}(k_1 r)|$ and $|H_0^{(2)}(k_2 r)|$ are monotonically decreasing, and do not intersect at any r . These two properties and the fact that $H_0^{(2)}(k_1 r)/H_0^{(2)}(k_2 r)$ is not a constant imply that if we choose a γ such that $H_0^{(2)}(k_1 r_A) = \gamma H_0^{(2)}(k_2 r_A)$, then there would be no r_B which satisfies $H_0^{(2)}(k_1 r_B) = \gamma H_0^{(2)}(k_2 r_B)$ unless $k_1 = k_2$ and $\gamma = 1$. Therefore, the vectors in (23) are mutually dependent only when $k_1 = k_2$, i.e. $f_1 = f_2$.

B Bound on Electromagnetic Interactions

In this appendix, we derive the interaction terms between two infinitesimally thin scatterers. Using the definition of the interaction term in (12), we calculate the electromagnetic interactions between two infinitesimal scatterers here, and then compare these simple interaction terms with those of extended objects in Section 5.1.

First we derive the scattered field from two filamentary objects separated by a distance d , when they are excited with a plane wave, $e^{j\mathbf{k} \cdot \mathbf{r}}$. With a simple approach, schematically shown in Fig. 11, we account for the multiple scattering between these objects iteratively. The field scattered from the first object due to the plane wave is obtained by setting $n = 0$ in (9) and is $\mathbf{E}_{s1} = c_0 H_0^{(2)}(k|\mathbf{r}|)$ where c_0 depends on the object radius, and $|\mathbf{r}|$ is the distance between the scatterer and the receivers

in the array. Thus, the value of $\mathbf{E}_{\mathbf{s1}}$ at the location of the second object is $c_0 H_0^{(2)}(kd)$, where d is the distance between the two objects.

Still assuming no interaction, we use $c_0 H_0^{(2)}(kd)$ as incident field on the second object, and find the scattered field due to the wave scattered from the first object as: $\mathbf{E}_{\mathbf{s12}} = d_0 H_0^{(2)}(k|\mathbf{r}|)$ where $d_0 = c_0 [c_0 H_0^{(2)}(kd)]$. When this field arrives at the location of the first object, it becomes $d_0 H_0^{(2)}(kd)$. Using this as the incident field on the first object, we calculate the third term in our scattering series as $\mathbf{E}_{\mathbf{s121}} = e_0 H_0^{(2)}(k|\mathbf{r}|)$ with $e_0 = c_0 [d_0 H_0^{(2)}(kd)]$. By carrying no-interaction scattered fields between scatterers in this manner, the scattered field due to the first object, in the presence of the second is given in the form of an infinite sum:

$$\begin{aligned} \mathbf{E}_{\mathbf{s1}}^{(2)} &= \mathbf{E}_{\mathbf{s1}} + \mathbf{E}_{\mathbf{s12}} + \mathbf{E}_{\mathbf{s121}} + \mathbf{E}_{\mathbf{s1212}} + \dots \\ &= c_0 \left\{ 1 + c_0 H_0^{(2)}(kd) + [c_0 H_0^{(2)}(kd)]^2 + [c_0 H_0^{(2)}(kd)]^3 + \dots \right\} H_0^{(2)}(k|\mathbf{r}|). \end{aligned} \quad (27)$$

The difference between with-interaction and no-interaction fields can now be expressed as:

$$\mathbf{E}_{\mathbf{s1}}^{(2)} - \mathbf{E}_{\mathbf{s1}} = c_0 \{ c_0 H_0^{(2)}(kd) + [c_0 H_0^{(2)}(kd)]^2 + [c_0 H_0^{(2)}(kd)]^3 + \dots \} H_0^{(2)}(k|\mathbf{r}|),$$

which can be written concisely as:

$$\mathbf{E}_{\mathbf{s1}}^{(2)} - \mathbf{E}_{\mathbf{s1}} = \frac{c_0^2 H_0^{(2)}(kd)}{1 - c_0 H_0^{(2)}(kd)} H_0^{(2)}(k|\mathbf{r}|). \quad (28)$$

Expression in (28) is warranted since $|c_0 H_0^{(2)}(kd)| < 1$, where $c_0 = \frac{J_0(ka)}{H_0^{(2)}(ka)}$ and a is the radius of the object, $a < d$. Evaluation of (28) for \mathbf{r} equal to the positions of the receivers then provides the means of determining $\delta(d)$ in (12).

References

- [1] A. Şahin and E.L. Miller, "Object Detection Using High-Resolution Near-Field Array Processing", *IEEE Trans. on Geoscience and Remote Sensing*, Submitted for Review.
- [2] A. Şahin and E.L. Miller, "Object-Based Localization of Buried Objects Using High Resolution Array Processing Techniques", in *Proc. of SPIE- AeroSense: Detection and Remediation Technologies for Mines and Minelike Targets*, May 1996, vol. 2765, pp. 409–419.
- [3] A. Şahin and E.L. Miller, "Performance Bounds for Matched Field Processing in Subsurface

- Object Detection Applications”, in *Proc. of SPIE– AeroSense: Detection and Remediation Technologies for Mines and Minelike Targets III*, April 1998, vol. 3392.
- [4] H.Krim and M.Viberg, “Sensor Array Signal Processing: Two Decades Later”, Tech. Rep. CICS-P-448, Center for Intelligent Control Systems, January 1995.
 - [5] M. Bouvet and G. Bienvenu, Eds., *High-Resolution Methods in Underwater Acoustics*, Springer-Verlag, Heidelberg, Germany, 1991.
 - [6] A.B. Baggeroerr, W.A. Kupperman, and P.N. Mikhalevsky, “An Overview of Matched Field Methods in Ocean Acoustics”, *IEEE J. of Oceanic Engineering*, vol. 18, no. 4, pp. 401–424, October 1993.
 - [7] D.F. Gingras, P. Gerstoft, and N.L. Gerr, “Electromagnetic Matched-Field Processing: Basic Concepts and Tropospheric Simulations”, *IEEE Trans. on Antennas and Propagation*, vol. 45, no. 10, pp. 1536–1544, October 1997.
 - [8] J.W. Choi and Y.H. Kim, “Spherical Beam-Forming and MUSIC Methods for the Estimation of Location and Strength of Spherical Sound Sources”, *Mechanical Systems and Signal Processing*, vol. 9, no. 5, pp. 569–588, 1995.
 - [9] J.W. Choi and Y.H. Kim, “Estimation of Locations and Strengths of Broadband Planar and Spherical Noise Sources Using Coherent Signal Subspace”, *Journal of Acoustical Society of America*, vol. 98, no. 4, pp. 2082–2093, October 1995.
 - [10] Y.D. Huang and M. Barkat, “Near-Field Multiple Source Localization by Passive Sensor Array”, *IEEE Transactions on Antennas and Propagation*, vol. 39, no. 7, pp. 968–974, July 1991.
 - [11] N. Yuen and B. Friedlander, “Performance Analysis of Higher Order ESPRIT for Localization of Near-Field Sources”, *IEEE Transactions on Signal Processing*, vol. 46, no. 3, pp. 709–719, March 1998.
 - [12] D. Starer and A. Nehorai, “Passive Localization of Near-Field Sources by Path Following”, *IEEE Trans. Signal Processing*, vol. 42, no. 3, pp. 677–680, March 1994.
 - [13] N. V. Budko and P. M. van den Berg, “Two-Dimensional Object Characterization with an Effective Model”, *Journal of Electromagnetic Waves and Applications*, vol. 12, pp. 177–190, 1998.
 - [14] S. Valaee, B. Champagne, and P. Kabal, “Parametric Localization of Distributed Sources”, *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2144–2153, September 1995.
 - [15] P. Stoica and A. Nehorai, “MUSIC, Maximum Likelihood, and Cramer-Rao Bound”, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.
 - [16] W.C. Chew, *Waves and Fields in Inhomogeneous Media*, Van Nostrand Reinhold, 1990.
 - [17] A. Şahin and E.L. Miller, “Recursive T-matrix Algorithm for Multiple Metallic Cylinders”, *Microwave Opt. Tech. Lett.*, vol. 15, no. 6, pp. 360–363, August 1997.

- [18] A. Şahin and E.L. Miller, “Recursive T-matrix Methods for Scattering from Multiple Dielectric and Metallic Objects”, *IEEE Trans. on Antennas and Propagation*, vol. 46, no. 5, pp. 672–678, May 1998.
- [19] G. Kristensson and S. Strom, “Scattering from buried inhomogeneities—a general three dimensional formalism”, *J. Acoust. Soc. Am*, vol. 64, pp. 917–936, 1978.
- [20] J.A. Kong, *Electromagnetic Wave Theory*, John Wiley & Sons, 1986.
- [21] J.E. Hipp, “Soil Electromagnetic Parameters as Functions of Frequency, Soil Density, and Soil Moisture”, *Proceedings of the IEEE*, vol. 62, no. 1, pp. 98–103, January 1974.
- [22] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes*, Cambridge University Press, Cambridge, UK, 1986.
- [23] Adnan Şahin, *Near Field Forward Scattering, and Object-Based Localization Algorithms for Subsurface Objects*, PhD thesis, Northeastern University, Boston, MA, 1998.

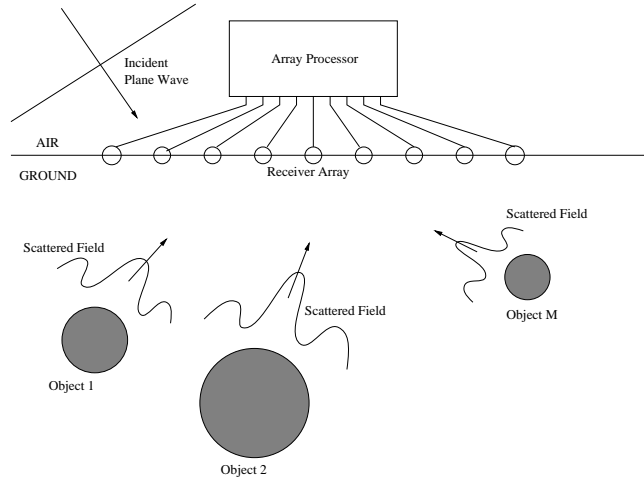


Figure 1: Problem geometry

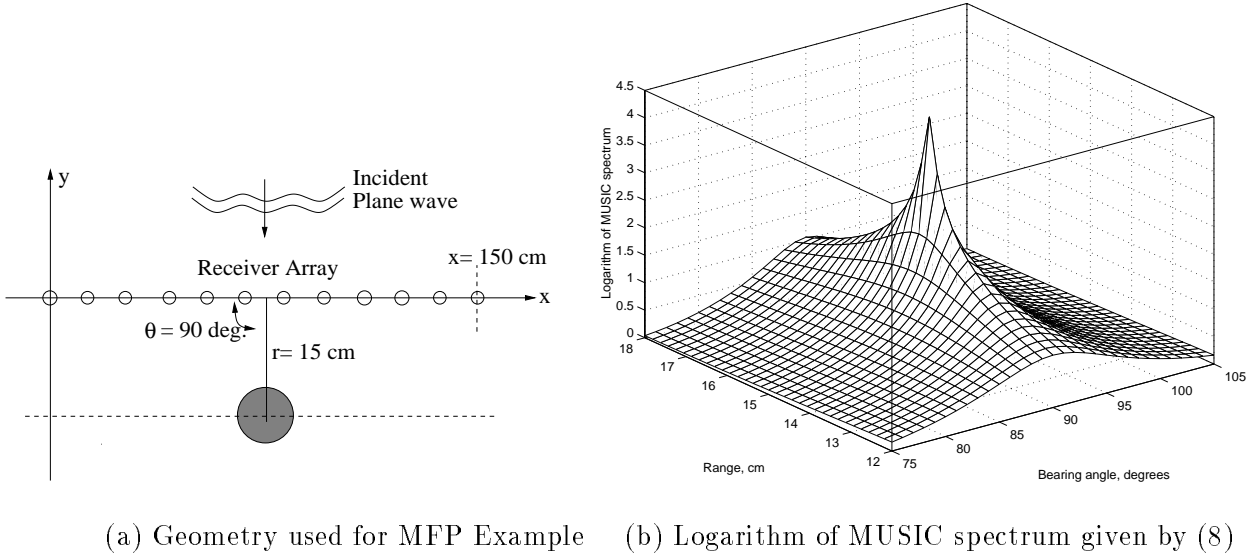


Figure 2: MFP localization example: single metallic object in a lossy, homogeneous background

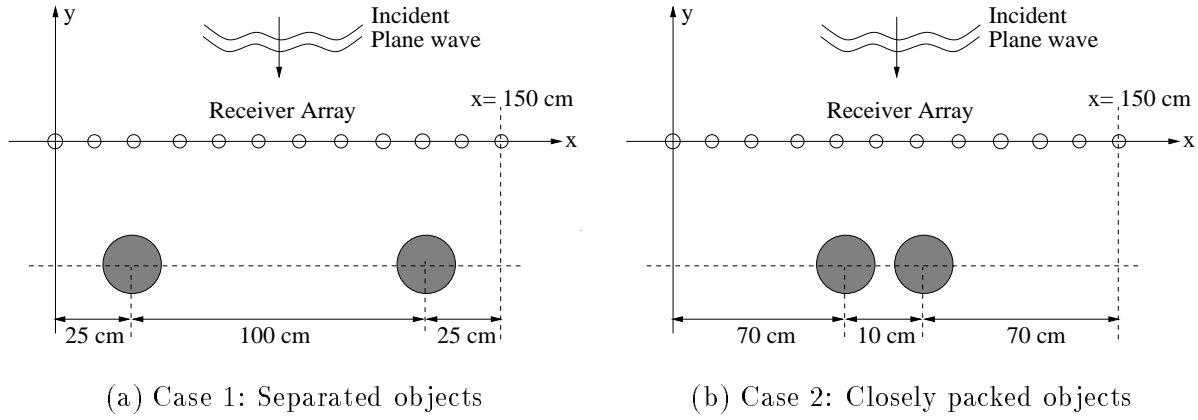


Figure 3: Example geometries used in multiple object localization

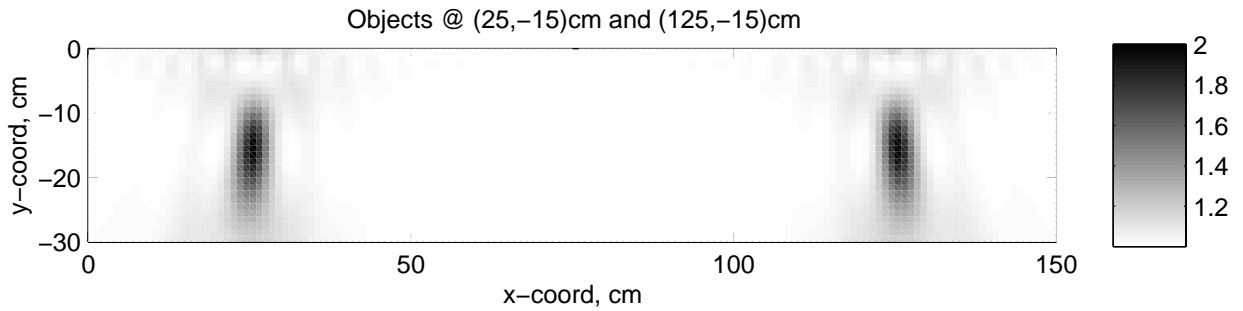


Figure 4: Localization of two targets located 1 meter apart, electromagnetic interactions are ignored

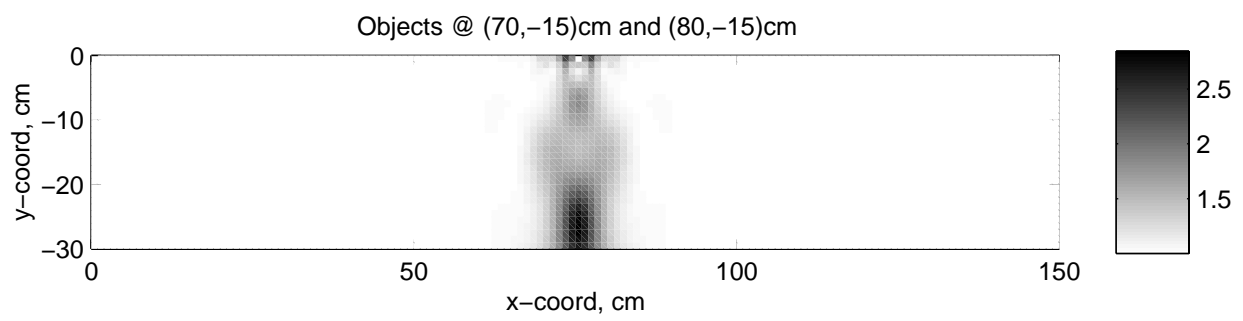
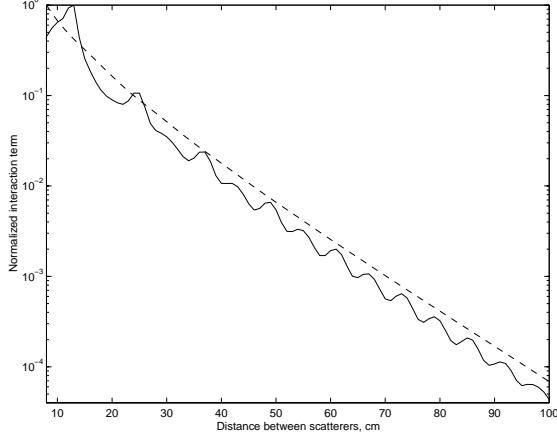
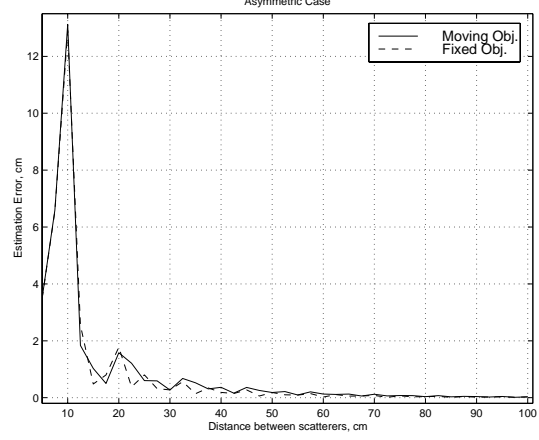


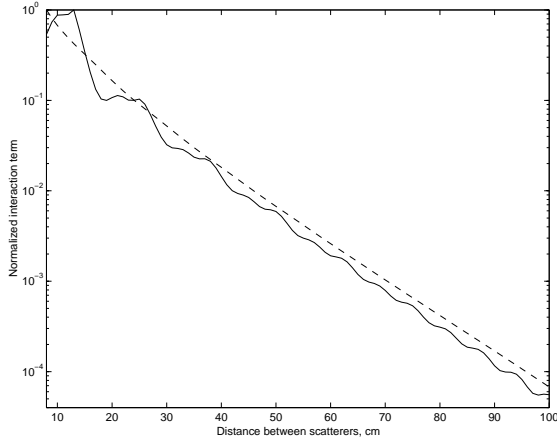
Figure 5: Localization fails when the target centers are 10cm apart in the no interaction case



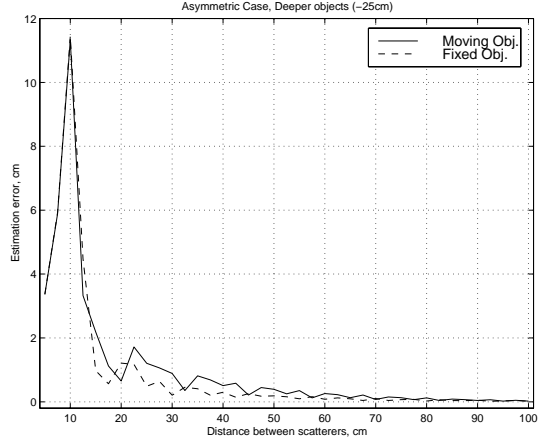
(a) Comparison of normalized interaction terms; solid line extended objects, dashed line filament objects, objects 15 cm away from the array



(b) Estimation error vs. separation for no interaction case, objects 15 cm away from the array

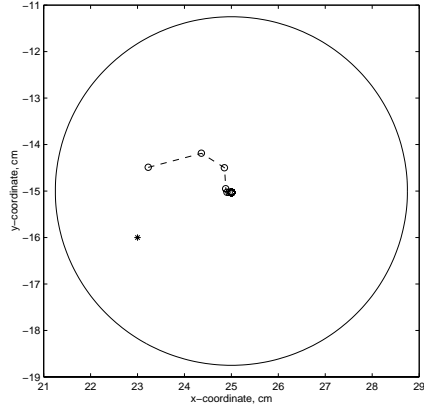


(c) Same as (a), objects 25 cm away from the array

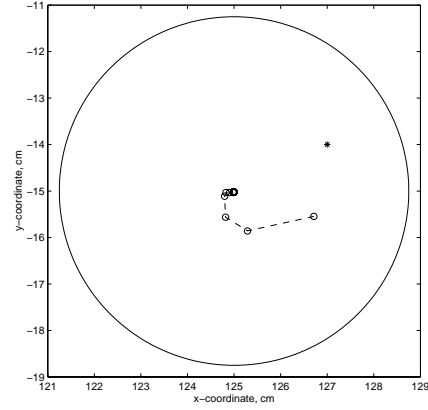


(d) Same as (b), objects 25 cm away from the array

Figure 6: Electromagnetic interaction between scatterers, and its effect on estimated object coordinates when the objects are 15 cm away from the receiver array (top plots) and 25 cm away (bottom plots)



(a) Object @ (25,-15)cm



(b) Object @ (125,-15)cm

Figure 7: Convergence of four-dimensional search into the true object coordinates when the objects are 1 meter apart, *'s indicate the initial guess, and the large circle indicates the support of the objects

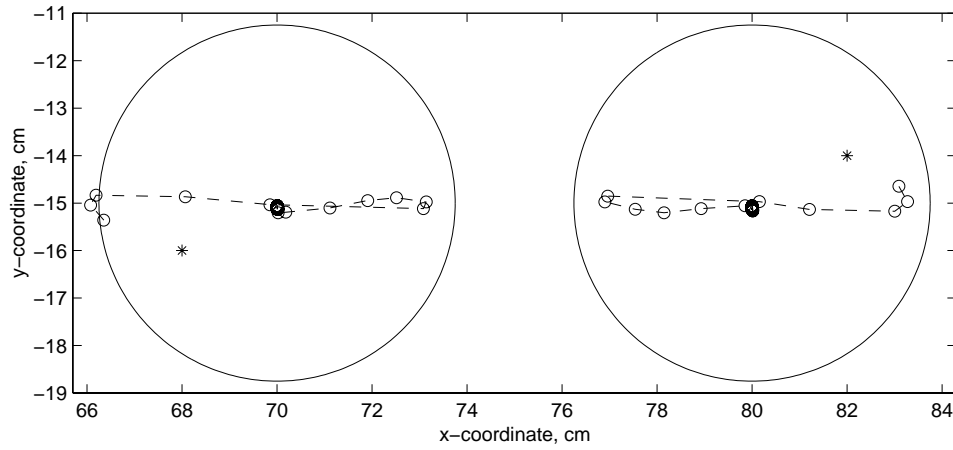


Figure 8: Convergence of four-dimensional search into the true object coordinates when the objects are 10 cm apart. *'s indicate the initial guess, and large circles indicate the support of the objects

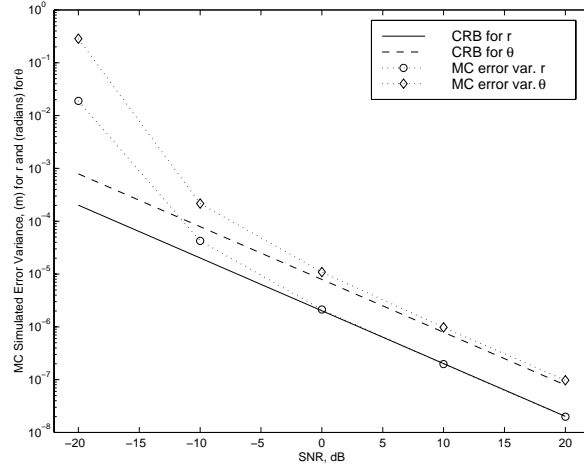
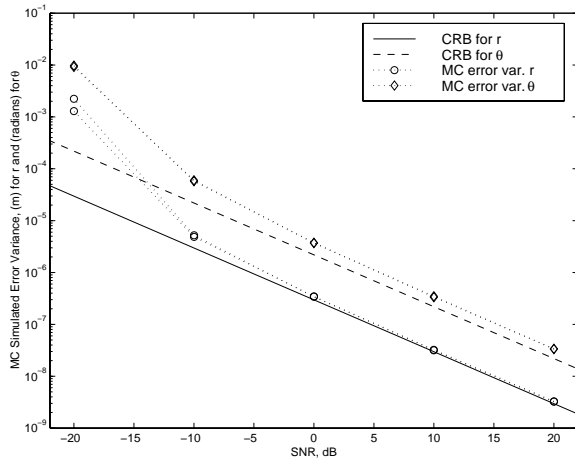
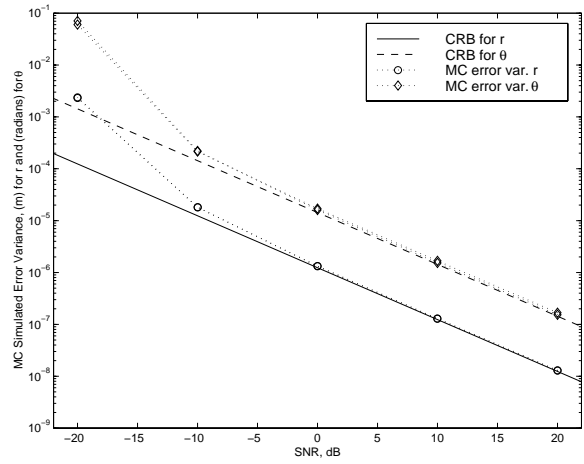


Figure 9: Comparison of analytical CRB with Monte-Carlo simulations for single object geometry given in Fig 2(a)



(a) Well separated objects, Fig 3(a)



(b) Closely spaced objects, Fig 3(b)

Figure 10: Comparison of analytical CRB with Monte-Carlo simulations for multiple object geometries given in Fig. 3

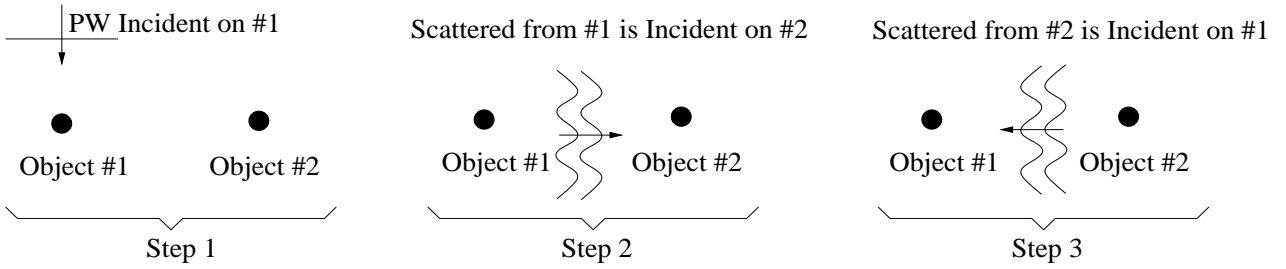


Figure 11: Finding the scattered field due to a filamentary object by incorporating the electromagnetic interactions between two scatterers step by step. Steps 2 and 3 are repeated infinitely many times successively to account for all interactions

Statistically-based Sequential Detection of Buried Mines from Array Ground Penetrating Radar Data

Xiaoyin Xu, Eric L. Miller, and Carey M. Rappaport
Center for Electromagnetics Research
Dept. of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115
Telephone: (617) 373-8386
Telefax: (617) 373-8627

ABSTRACT

We consider the problem of detecting and localizing buried landmines from a ground penetrating radar (GPR) array. A simplified, ray-optics-based physical model for time domain GPR returns is presented. Under this model in the absence of an object from the field of view of the array, there exist well defined symmetries in the structure of the radar returns. In particular, for a bistatic system composed of one length M transmit array and a second length M array of receivers, we identify M subsets of signals from the M^2 total transmitter/receiver pairs such that the mean value of the signals within each subset should be the same when no object is present. This relationship then forms the basis for a modified Hotelling's T^2 -test to detect the presence of objects when there is noise in the signal. Simulation results demonstrate the validity of these methods.

1. INTRODUCTION

Mines kill or maim hundreds of people every week, mostly innocent and defenseless civilians. Among the various demining methods developed in recent years, ground penetrating radar (GPR) holds substantial promise because of its sensitivity to non-metallic, plastic objects which traditional metal detectors are largely incapable of finding. Nevertheless, using GPR to detect and localize mines is both difficult and complex. One key problem is the rejection of interference caused by the signal arising from scatter off the air-earth interface. Both the magnitude of the ground bounce and its timing are such that they can easily swamp the relatively small signal arising from the interaction of the transmitted GPR waveform with the buried mine. To help overcome this difficulty, we consider the use of a GPR array to provide a richer and more diverse set of data thereby making accurate detection and localization possible in circumstances where a single GPR is unable to perform well.

In this paper we discuss a method to detect mines which exploits both the physics of the problem as well as the geometry of the array system. We assume that the GPR array is deployed as shown in Fig. 3 with one linear array of transmitters and a second array of receivers both traveling down a track. As described in Section 3, the geometric symmetry inherent in this configuration introduces a “statistical symmetry” in the received signals. In particular, this symmetry is preserved precisely when there is no mine. On the other hand, the presence of a mine will break this symmetry and therefore provide information as to the existence of an object. Based on this relationship, we develop a statistical test of homogeneity to ascertain the presence of an object in the field of view of the array.

The organization of this paper is as follows. Section 2 introduces the physical model of GPR signals and the configuration of the GPR array. In Section 3 we present the detection algorithm, using a modified

Other author information: XX: xxu@cdsp.neu.edu, ELM: elmiller@ece.neu.edu, CR: rappaport@neu.edu. This work was supported by the Army Research Office Demining MURI under Grant DAAG55-97-1-0013

form of Hotelling's T^2 -test. Examples are given in Section 4 to demonstrate validity of the algorithm. Section 5 summarizes present work and points out future research direction.

2. PHYSICAL MODEL AND PROBLEM FORMULATION

2.1. A Single GPR System

To detect and localize mines, a ground penetrating radar array is implemented. Fig. 1 shows a typical single GPR system with the signals it generates.¹ In this paper we assume a simplified model where the signal seen by the receiver is composed of at most two components. The first signal is the reflected signal from the ground and is always present in the data. The second component (if it exists) is the reflected contribution from an object in the field of view of the array.

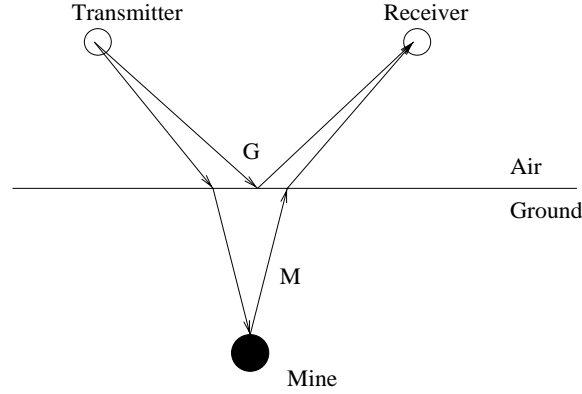


Figure 1. Schematic drawing of a single GPR, transmitter and receiver.

The received signal, $\phi(t)$, is taken to be the sum of delayed and attenuated versions of two “template” signals indicating the nominal behavior of the ground bounce signal and the nominal behavior of a signal arising from scattering from a mine. Mathematically we have

$$\phi(t) = a\psi_g(t - \tau_g) + b\psi_m(t - \tau_m) \quad (1)$$

where ψ_g and ψ_m are the nominal ground bounce and mine reflected signal, a and b are attenuation factors, τ_g is the delay of the ground reflection, and τ_m is the delay of the mine signal. Note that if no mine is present, ϕ is just equal to the first term of (1).

To find the delays and the attenuation factors we assume that the propagation of the signal from the transmitter to the receiver can be described using a ray-optics-type model shown in Fig. 1. That is, the ground bounce is composed of signal reflected from the interface at the specular point midway between the transmitter and receiver while the four-part path of the mine component of the signal can be determined via the judicious use of Snell's law.

To begin, the τ_g and τ_m are determined by the travel time of two-way paths and can be calculated as

$$\text{Delay} = \frac{\text{2-way path length}}{\text{velocity of the wave}}. \quad (2)$$

To find τ_g and τ_m , we need to locate reflecting point and refracting point shown in Fig. 2. Let media 1 be air and media 2 be soil, with electric permittivity ϵ_0 and ϵ_1 , respectively, Fig. 2.(a), the reflecting point on the boundary between two points (x_1, y_1) and (x_2, y_1) in media 1 is simply the mid-point $(x_4, 0)$, where $x_4 = \frac{x_1 + x_2}{2}$.

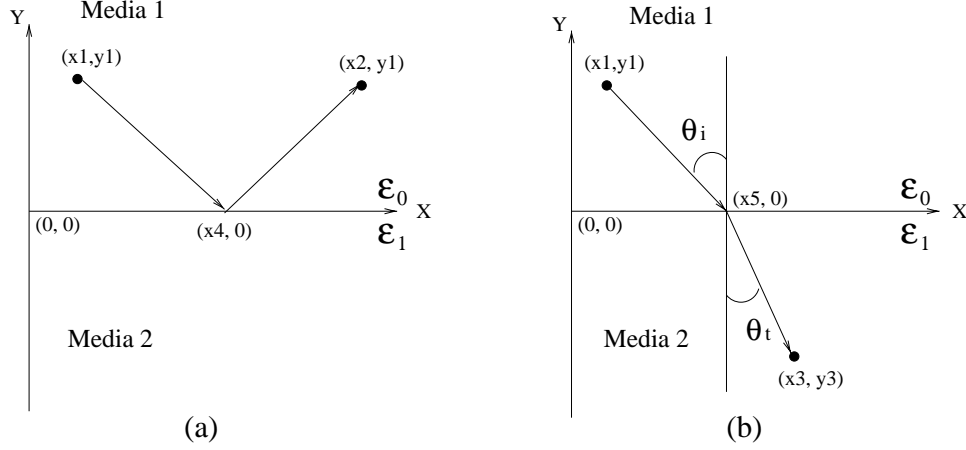


Figure 2. Geometries for determining (a) the reflecting point and (b) the refracting point, $\epsilon_1 > \epsilon_0$.

For the refracting point, according to Snell's law, for a source located at (x_1, y_1) in media 1 and target at (x_3, y_3) in media 2, the refracted ray from source to target must intersect the boundary at a point $(x_5, 0)$, Fig. 2.(b), such that²

$$\frac{Re\{\epsilon_1\}}{\epsilon_0} = \frac{\frac{(x_1 - x_5)^2}{(x_1 - x_5)^2 + y_1^2}}{\frac{(x_3 - x_5)^2}{(x_3 - x_5)^2 + y_3^2}}. \quad (3)$$

Solution of this quartic equation has four roots. By Fermat's principle, which states that of all possible paths joining two given points on a wave path, the wave path has actual least travel time, we can discard three physically impossible roots and retain the true refracting point. Once the reflecting point and the refracting point are established, the delay τ_g and τ_m can be found as,

$$\tau_g = 2 \frac{\sqrt{(x_1 - x_4)^2 + y_1^2}}{c} \quad (4)$$

$$\tau_m = 2 \frac{\sqrt{(x_1 - x_5)^2 + y_1^2}}{c} + 2 \frac{\sqrt{(x_3 - x_5)^2 + y_3^2}}{c/Re\sqrt{\epsilon_1}} \quad (5)$$

where c is the speed of light in air and $\frac{c}{Re\{\sqrt{\epsilon_1}\}}$ is the speed of the wave in soil.

In addition to the time delays, the received signal $\phi(t)$ has an amplitude reduction caused by propagation through the soil as well as geometric spreading as it traverses both the air and the earth. In soil, the wave attenuates exponentially with the distance it travels, $e^{-\alpha_s d}$. The quantity α_s is the attenuation constant of the soil which is related to the conductivity and permittivity of the medium³ while d is the distance the wave travels in the earth. We assume geometric spreading results in an inverse path length amplitude reduction. Referring to the setup of Fig. 2, then we have the overall amplitude reduction factors given by

$$a = \frac{1}{2\sqrt{(x_1 - x_4)^2 + y_1^2}} \quad \text{and} \quad b = \left(\frac{e^{-\alpha_s \sqrt{(x_3 - x_5)^2 + y_3^2}}}{\sqrt{(x_1 - x_5)^2 + y_1^2} + \sqrt{(x_3 - x_5)^2 + y_3^2}} \right)^2.$$

2.2. GPR Array

In this work, the GPR array is assumed to consist of M pairs of transmitters and receivers. Data are collected by the GPR array as it travels step by step down track, Fig. 3. At each stop of the array, M^2 signals (time-traces) are collected; one for each transmitter/receiver pair.

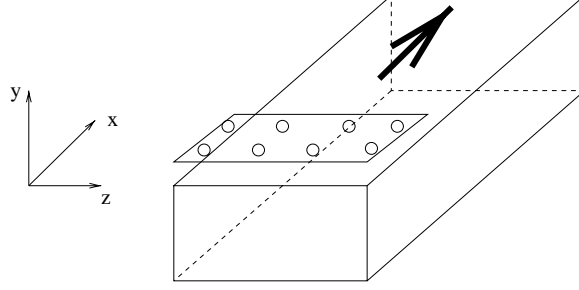


Figure 3. Array of GPR, moves in x-direction(down-track).

Using the model developed in the previous section and assuming that the GPR array is at its k -th stop, the signal seen at receiver j due to input from transmitter i is written as

$$\phi_{ij}^k(t) = a_{ij}\psi_g(t - \tau_{g,ij}) + b_{ij}^k\psi_m(t - \tau_{m,ij}^k), \quad i, j = 1, \dots, M, \quad k = 1, \dots, K. \quad (6)$$

Note that a and τ_g do not depend on k , the down-track GPR array position. This is easy to understand because the ground-reflected signal only depends on the relative position of transmitter and receiver*. To simplify matters, in the future we use the following shorthand

$$g_{ij}(t) = a_{ij}\psi_g(t - \tau_{g,ij}) \quad s_{ij}^k(t) = b_{ij}^k\psi_m(t - \tau_{m,ij}^k).$$

Each $\phi_{ij}(t)$ is densely sampled P times over a time interval. The interval is chosen to be long enough to embrace both ground bounce and mine signal. When no ambiguity will arise, we refer to the vector of samples, $\underline{\phi}_{ij}^k$, rather than the temporal signal, $\phi_{ij}^k(t)$ with a similar interpretation holding for \underline{g}_{ij} and \underline{s}_{ij}^k . Note $\underline{\phi}_{ij}^k$, \underline{g}_{ij} , and \underline{s}_{ij}^k are column vectors of size P .

For a given location of the GPR array, to detect mines, we carry out a binary hypothesis test. Under the null hypothesis, H_0 , the received signal $\underline{\phi}_{ij}^k$ is comprised of ground bounce \underline{g}_{ij} plus measurement noise, which is assumed to be a white Gaussian vector, $\underline{w} \sim N(0, \sigma^2 \mathbf{I})$. Under the alternate hypothesis, H_1 , $\underline{\phi}_{ij}^k$ consists of ground bounce, noise, and mine signal, \underline{s}_{ij}^k . Mathematically we have

$$\begin{aligned} H_0^k : \quad & \underline{\phi}_{ij}^k = \underline{g}_{ij} + \underline{w} \\ H_1^k : \quad & \underline{\phi}_{ij}^k = \underline{g}_{ij} + \underline{s}_{ij}^k + \underline{w}. \end{aligned} \quad (7)$$

Our processing method is based on the observation that under H_0 , the M^2 received signal should display certain symmetries, as illustrated by Fig. 4. For example if no mine is present then $\underline{\phi}_{12}$ should be “statistically equal” to $\underline{\phi}_{21}$, $\underline{\phi}_{23}$, $\underline{\phi}_{32}$, $\underline{\phi}_{34}$, and $\underline{\phi}_{43}$ because the ground bounce in each case depends only on the relative spacing of the sensors which is identical for these six pairs. Similarly, $\underline{\phi}_{13}$ should be statistically equal to $\underline{\phi}_{31}$, and so on. By “statistically equal” we mean that any variations in these signal are caused by random sensor noise. In other words, signals from these sets will, on average, possess the same means with some variability (variance) caused by the noise. Thus statistical tests designed to determine homogeneity of a population (i.e. equality of mean vectors) can be used to test whether an object is present (lack of homogeneity) or absent (all the data vectors are about the same). Finally, for the M^2 received signals it is not hard to show that there are only M sets of statistically different signals because of this symmetry. Note the M sets of signals are not of equal size, some sets consist of more signals than the others.

*Here we are assuming that the ground is locally flat over the extent of the sensing system. Extension of the results in this work to smoothly changing ground is an area of current work.

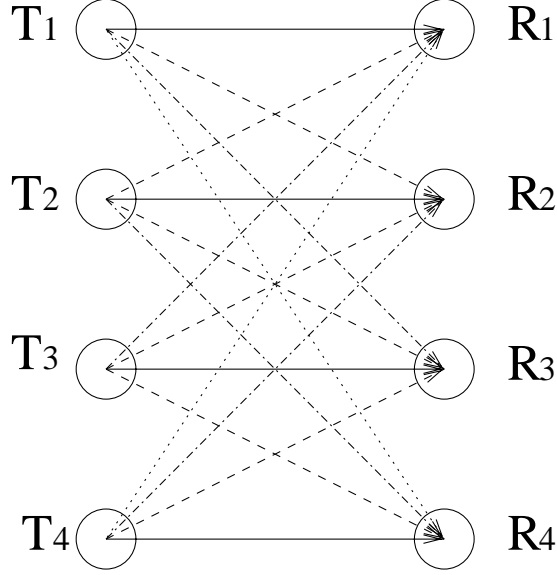


Figure 4. Symmetry of GPR signals, 4 different path lengths are shown in solid line, dashed line, dashed-dot line, and dot line.

3. INFORMATION EXTRACTION ALGORITHM

From the above discussion, we know that if a mine is present, the symmetry of received signals will be disrupted. Therefore, to detect mines, we can look for asymmetry of received signals. Basically, our approach is to sequentially detect any changes in the means of the received signals. For the purpose of illustration, we use $M = 4$ pairs of transmitter and receiver. At each stop k of the GPR array, we then have 4 sets of statistically different signals, $S_1 = \{\phi_{11}^k, \phi_{22}^k, \phi_{33}^k, \phi_{44}^k\}$, $S_2 = \{\phi_{12}^k, \phi_{21}^k, \phi_{23}^k, \phi_{32}^k, \phi_{34}^k, \phi_{43}^k\}$, $S_3 = \{\phi_{13}^k, \phi_{31}^k, \phi_{24}^k, \phi_{42}^k\}$, and $S_4 = \{\phi_{14}^k, \phi_{41}^k\}$.

3.1. Homogeneity Test of Single Set

Here we discuss the homogeneity test of a single set, say $S_1 = \{\phi_{11}^k, \phi_{22}^k, \phi_{33}^k, \phi_{44}^k\}$. We consider the acceptance test of the hypothesis

$$H_0: \underline{\phi}_{ii}^k \text{ in the set are from same multivariate Gaussian distribution, with equal and known covariance matrices, and identical, unknown means, } i = 1, \dots, 4.$$

Because the test is identically the same at each stop k for the remainder of this subsection we drop the k notation.

As the problem currently stands, we do not assume that the ground bounce signals are known. In fact, a simple transformation of the data allows us to perform the test without ever having to know the ground bounces or estimate them. To see this, we begin by forming the collection of all pairwise differences of the signals within the set under consideration. For S_1 we get the set D_1 defined as

$$D_1 = \left\{ \phi_{11} - \phi_{22}, \phi_{11} - \phi_{33}, \phi_{11} - \phi_{44}, \phi_{22} - \phi_{33}, \phi_{22} - \phi_{44}, \phi_{33} - \phi_{44} \right\} \equiv \{\underline{\delta}_1, \underline{\delta}_2, \underline{\delta}_3, \underline{\delta}_4, \underline{\delta}_5, \underline{\delta}_6\}.$$

Testing H_0 using the original set of $\underline{\phi}_{ii}$ signals then is equivalent to testing the following hypothesis using the $\underline{\delta}_n$ vectors:

$$H'_0: \underline{\delta}_n \text{ in the set are from same multivariate Gaussian distribution, with equal and known covariance matrices and means equal to zero}$$

To test this equality of means, we carry out a Hotelling's T^2 -test.⁴ As described more fully in [5], this test amounts to the following comparison

$$\xi_1 = \sum_{n=1}^6 \frac{1}{2} \underline{\delta}_n^T \mathbf{R}^{-1} \underline{\delta}_n \leqslant threshold \quad (8)$$

with $\mathbf{R}(= \sigma^2 \mathbf{I})$ is the covariance matrix of the measurement noise and *threshold* is chosen to ensure an *a priori* specified probability of false acceptance. The hypothesis is accepted if the left hand side is less than or equal to *threshold* and rejected otherwise. Noting that \mathbf{R} is diagonal, we write ξ_1 as

$$\xi_1 = \frac{1}{2\sigma^2} \Phi_1^T \mathbf{A}_1 \Phi_1 = \frac{1}{2\sigma^2} \begin{bmatrix} \underline{\phi}_1^T & \underline{\phi}_2^T & \underline{\phi}_3^T & \underline{\phi}_4^T \end{bmatrix} \begin{bmatrix} 3\mathbf{I} & -\mathbf{I} & -\mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & 3\mathbf{I} & -\mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & -\mathbf{I} & 3\mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & -\mathbf{I} & -\mathbf{I} & 3\mathbf{I} \end{bmatrix} \begin{bmatrix} \underline{\phi}_1 \\ \underline{\phi}_2 \\ \underline{\phi}_3 \\ \underline{\phi}_4 \end{bmatrix}. \quad (9)$$

For a set S_i of size N_i , it is easy to show that \mathbf{A}_i is an $N_i \times N_i$ block matrix. Its block diagonal elements are $(N_i - 1)\mathbf{I}_{P \times P}$ and the off diagonal elements are $-\mathbf{I}_{P \times P}$.

3.2. Homogeneity Test of Multiple Sets

For the application of interest in this paper, we have multiple sets (four for our sample system) for which we wish to test homogeneity. Here we adopt a simple, generalized form of the T^2 test in which the four ξ_i 's are weighted and added to yield one final test statistic,

$$\xi = \xi_1 \eta_1 + \xi_2 \eta_2 + \xi_3 \eta_3 + \xi_4 \eta_4. \quad (10)$$

The η_i are weight factors, defined by

$$\eta_i = \frac{1}{l_i} \left(\sum_{j=1}^4 l_j^{-1} \right)^{-1} \quad (11)$$

where l_j are the path lengths of ground bounces for set S_j . More will be said about this weight factor in Section 3.3. Stacking all the 16 signals ϕ_{ij} to form a long column vector, we obtain a new vector

$$\Phi^T = [\Phi_1^T \ \Phi_2^T \ \Phi_3^T \ \Phi_4^T]$$

and

$$\xi = \frac{1}{2\sigma^2} \Phi^T \mathbf{A} \Phi = \frac{1}{2\sigma^2} \begin{bmatrix} \Phi_1^T & \Phi_2^T & \Phi_3^T & \Phi_4^T \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 \eta_1 & 0 & 0 & 0 \\ 0 & \mathbf{A}_2 \eta_2 & 0 & 0 \\ 0 & 0 & \mathbf{A}_3 \eta_3 & 0 \\ 0 & 0 & 0 & \mathbf{A}_4 \eta_4 \end{bmatrix} \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \\ \Phi_4 \end{bmatrix}. \quad (12)$$

Matrix \mathbf{A} is a block diagonal matrix. Obviously, \mathbf{A} is symmetric. Note \mathbf{A} is positive semidefinite by its buildup. According to Mathai,⁶ the first two moments of ξ are given by,

$$\begin{aligned} \mu_\xi &= (6\eta_1 + 15\eta_2 + 6\eta_3 + \eta_4)P \\ \sigma_\xi^2 &= (24\eta_1^2 + 90\eta_2^2 + 24\eta_3^2 + 2\eta_4^2)P.^\dagger \end{aligned} \quad (13)$$

It has been shown that ξ asymptotically has a Gaussian distribution⁶ for large NP , i.e., $\xi \sim N(\mu_\xi, \sigma_\xi^2)$. The generalized T^2 -test amounts to $\xi \leq \alpha$ where α is the threshold chosen to ensure an *a priori* specified probability P_{fa} of false acceptance. The α is determined by

$$P_{fa} = \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\xi^2}} e^{-\frac{(\xi-\mu_\xi)^2}{2\sigma_\xi^2}} d\xi. \quad (14)$$

3.3. Sequential Detection

Now we consider the actual GPR array operation as the array moves down-track. At each stop k , we calculate a corresponding ξ^k . When there is no mine in the field of view of the array, we record ξ^k only. That is, $\nu^k = \xi^k$ under hypothesis H_0 . When there is a mine, the mine signal \underline{s}_{ij}^k adds to the ground bounce. Passing \underline{s}_{ij}^k to the generalized Hotelling's T^2 -test produces an output signal, denoted by ζ^k . So the problem is to detect signal ζ^k , given the observed signal sequence ν^k and known "noise" ξ^k , using the additive noise model $\nu^k = \xi^k + \zeta^k$ under the alternate hypothesis H_1 . Though ζ^k is unknown and changes with k , it is always *positive* because of the very nature of quadratic form of the generalized T^2 -test. Fig. 5 shows separate ξ_i^k and $\xi_i^k + \zeta_i^k$. Signals on the left side of Fig. 5 are typical "noise" sequences, taken from four different sets of transmitter-receiver combinations. Signals on the right side are "noise" and mine signals, corresponding to the respective sets of transmitter-receiver combinations. It is seen that ζ_i^k are much weaker in set S_3 and S_4 . This observation can be explained by the increased attenuation associated with the longer distances the mine signals in sets S_3 and S_4 travel. For this reason, we introduced the weight factors in Eq. 11.

When the mine is buried deep, the mine signal attenuates exponentially as explained in Section 2.1. Detecting this unknown low power signal can be aided quite a bit using sequential detection methods which retain information from previous scans to improve the SNR. We therefore employ a sequential detector as the GPR array moves down track. At each stop k , the detector makes one of two decisions⁷: (1) Hypothesis H_0 is true, no mine signal is present, (2) Reject H_0 . Because ζ^k causes a positive displacement of the mean of ν^k , we choose a running average of ν^k as a statistical test^{8,9}

$$g^k = \frac{1}{N} \sum_{j=k-N+1}^k \nu^j, \quad k = N, N+1, \dots, K \quad (15)$$

and make a decision by checking $g^k \leq \beta$ where β is a threshold. At each stop k , g^k is compared to the threshold to make a declaration of mine presence. Fig. 6 illustrates relationship between threshold settings and declaration of mine presence as the GPR arrays move down-track. Each filled dot indicates a location of the array where we say that a mine is present. Such a declaration is made when g^k is above the threshold for consecutive 5 stops of k . A lower threshold allows us to make early declaration of mine. The disadvantage is a high false-acceptance rate. The false-acceptance probability P_{fa} equals the probability under hypothesis H_0 that the g^k crosses the threshold. The detection probability P_d equals to $1 - P_{fa}$. By virtue of the generalized T^2 -test, ν^k are statistically independent. Hence, g^k has the Gaussian distribution, $N(\mu_\xi, \frac{1}{N}\sigma_\xi^2)$. Thus, in principle, for a given P_{fa} , β can be determined similarly as in Eq. 14.

[†]See appendix

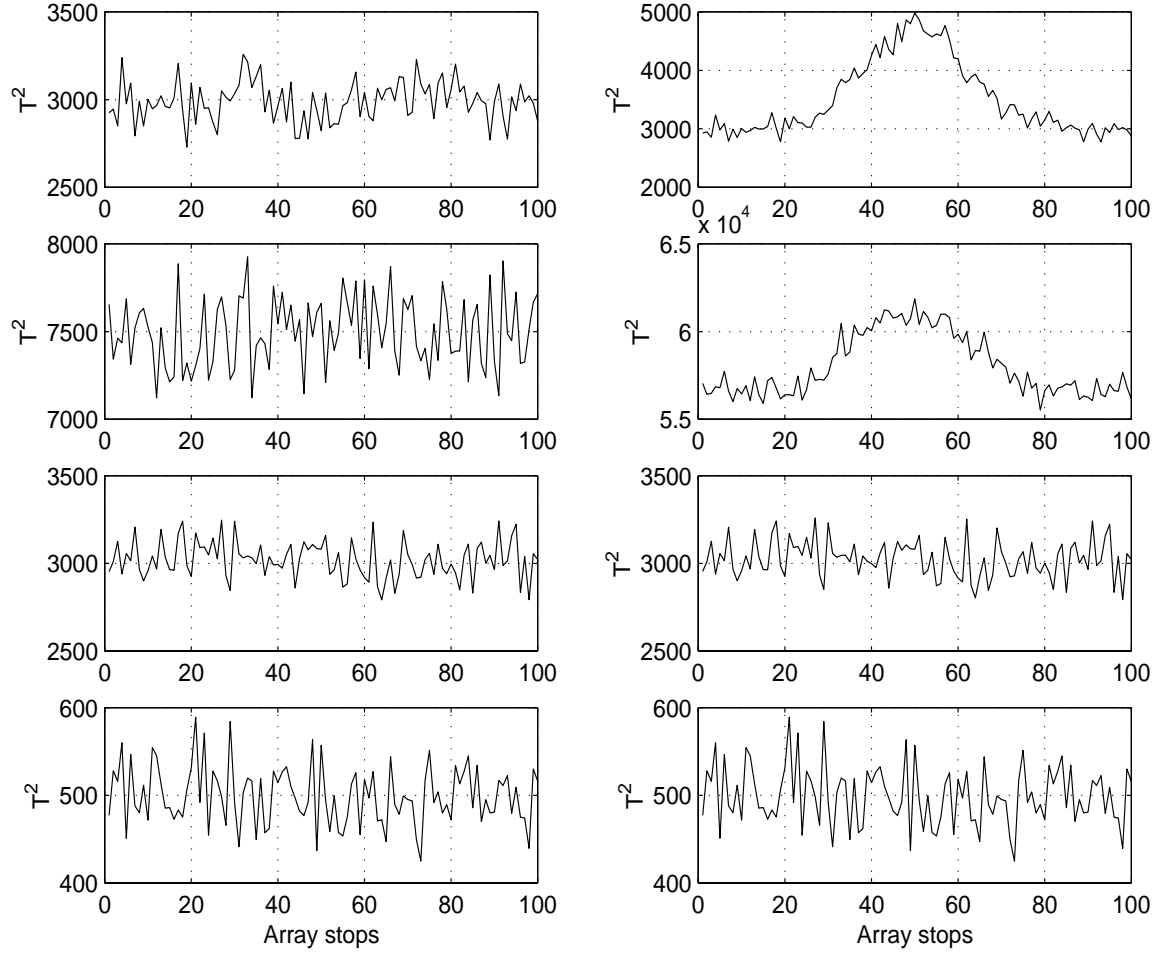


Figure 5. First row of figures corresponds to the T^2 statistic for set S_1 , the left figure is “noise” only, the right figure has both “noise” and mine reflected signal, second row of figures corresponds to set S_2 , third row of figures corresponds to set S_3 , and, fourth row of figures corresponds to set S_4 .

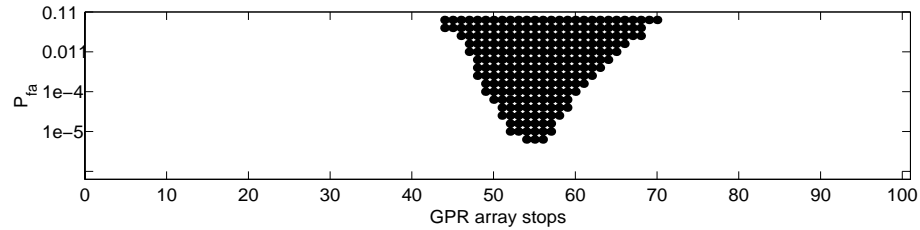


Figure 6. Detection results versus prescribed false alarm probability, P_{fa} . Note that lower P_{fa} results in slower detection while a higher allowable false alarm rate allows for detection prior to the array passing over the mine.

4. EXAMPLES

In this section we consider some examples of the above procedure. Monte-Carlo runs were used to determine the probability of detection and probability of false-alarm for various noise levels and depth of buried mines. In this work, the signal to interference plus noise ratio is defined as

$$\text{SINR} = 10 \log_{10} \frac{\underline{s}^T \underline{s}}{\underline{g}^T \underline{g} + P\sigma^2}. \quad (16)$$

In all cases, we generate synthetic data with an object located around the 50th stops of the GPR array and buried 10 cm underground. The GPR array is composed of four pairs of transmitters and receivers evenly spaced along a baseline width of 80 cm. A transmitter and its corresponding receiver (e.g. transmitter 1 and receiver 1) are 20 cm apart. The GPR array is 40 cm above the ground. For simplicity we assume that the nominal ground bounce and mine-bounce signal take the form of a second derivative of a Gaussian shown in Fig. 7(a). Current work in our group is aimed at developing more sophisticated models for these signals. Fig. 7(b) and (c) show the received signals of two pairs of T/R combinations. Because of the domination of the ground bounce, it is difficult to see any mine signature. Even after pairwise subtraction, the mine signal can hardly be observed in Fig. 7(d). But the generalized T^2 -test can pick up this difference and declare a mine.

Fig. 8(a) shows the receiver operating characteristics of detecting a mine buried 10 cm underground. Fig. 8(b) shows two ROC's for mines buried at different depth under the same noise power. Because of the fast attenuation in soil, the deeper buried mine has a significantly smaller SINR thereby leading to the degradation in performance.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we discuss a statistical approach to detect mines using a GPR array. Exploiting the geometric symmetry of GPR array setting, we have looked into tests for statistical homogeneity of GPR returns as a tool for performing detection. Basically, by treating the received signal as a multivariate Gaussian distribution we test its statistical homogeneity using a generalized, sequential Hotelling's T^2 -test.

Future work will take into consideration of rough ground-air interface and time-delay estimation to actually localize detected mines. An assumption of this work is that the soil conditions are known. Sensitivity analysis involving mismatch in soil parameters will be valuable to apply this method to realistic scenario. Moreover, the additive white Gaussian noise mode will be lifted as we explore issues associated with the modeling of clutter, the incorporation of these models into our processing, and the development of test which are robust (or invariant) to uncertainty in these models.

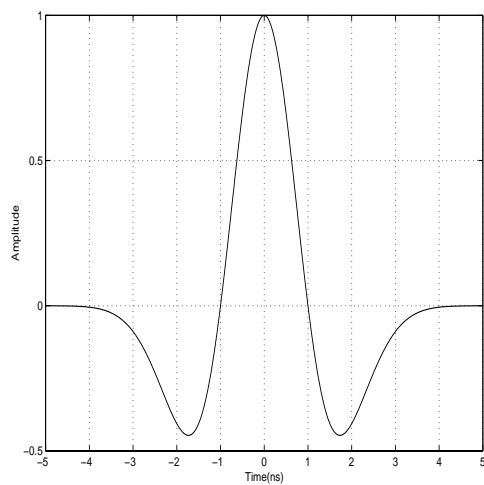
6. APPENDIX

From Mathai's⁶ work, for $\mathbf{X} \sim \mathbf{N}(\mu, \Sigma)$, the first two moments of quadratic $\mathbf{X}^T \mathbf{A} \mathbf{X}$ can be found by

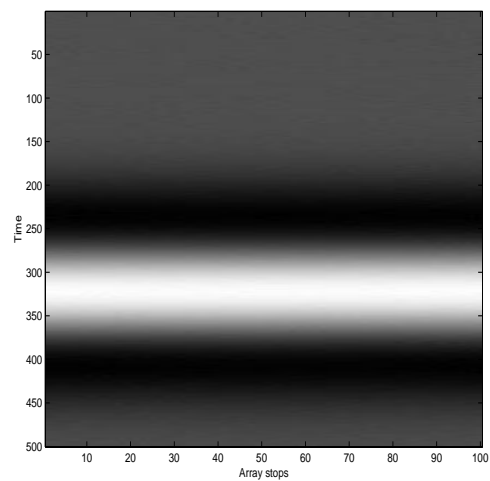
$$\begin{aligned} E[\xi] &= \text{tr}(\mathbf{A}\Sigma) + \underline{\mu}^T \mathbf{A} \underline{\mu} \\ \text{Var}[\xi] &= 2\text{tr}(\mathbf{A}\Sigma)^2 + 4\underline{\mu}^T \mathbf{A} \Sigma \mathbf{A} \underline{\mu}. \end{aligned} \quad (17)$$

For the generalized T^2 -test, we have

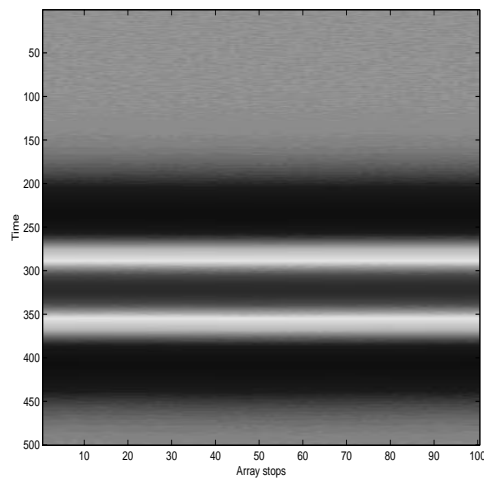
$$\underline{\mu}^T = \left[\underline{\mu}_{11}^T, \dots, \underline{\mu}_{44}^T, \underline{\mu}_{12}^T, \dots, \underline{\mu}_{34}^T, \underline{\mu}_{13}^T, \dots, \underline{\mu}_{24}^T, \underline{\mu}_{14}^T, \underline{\mu}_{41}^T \right] \quad (18)$$



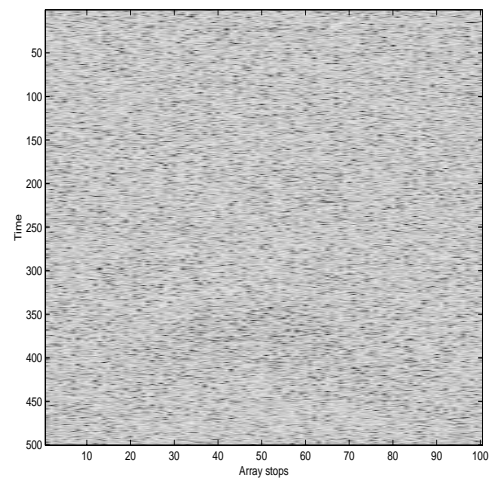
(a) Received signal, the second order derivative of a Gaussian.



(b) Time trace of ϕ_{11}



(c) Time trace of ϕ_{22}



(d) Difference of time traces, $\phi_{11} - \phi_{22}$,
SINR = -11 dB.

Figure 7. Received signal and time traces.

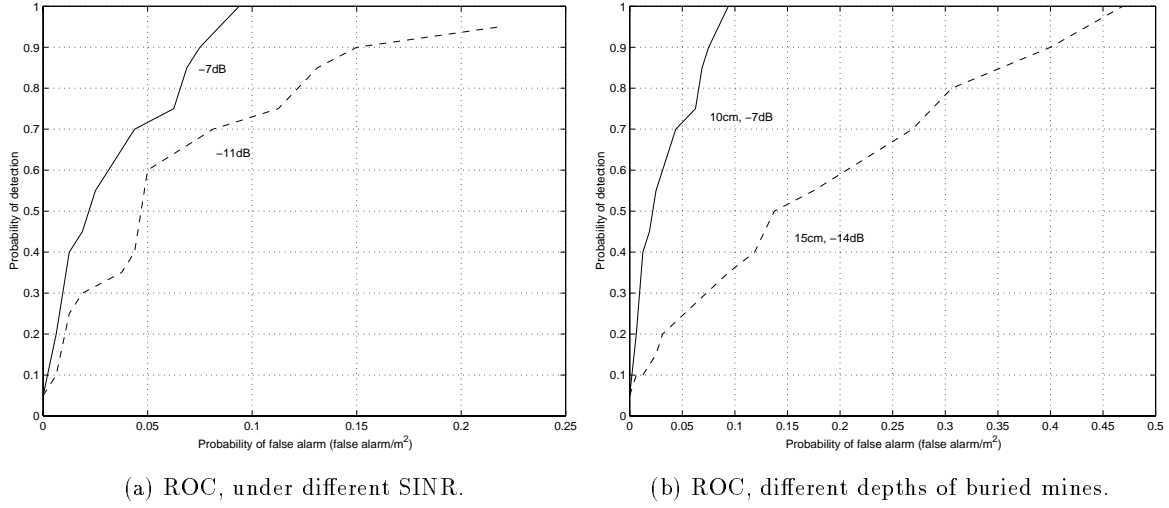


Figure 8. Receiver operating characteristics.

where

$$\begin{aligned}
 \underline{\mu}_{11} = \dots = \underline{\mu}_{44} &\equiv \underline{\mu}^{(1)} & \underline{\mu}_{12} = \dots = \underline{\mu}_{34} &\equiv \underline{\mu}^{(2)} \\
 \underline{\mu}_{13} = \dots = \underline{\mu}_{24} &\equiv \underline{\mu}^{(3)} & \underline{\mu}_{14} = \underline{\mu}_{41} &\equiv \underline{\mu}^{(3)}.
 \end{aligned} \tag{19}$$

Next, we note that $4\underline{\mu}^T \mathbf{A} \Sigma \mathbf{A} \underline{\mu} = 0$ under the condition of Eq. 19,

$$4\underline{\mu}^T \mathbf{A} \Sigma \mathbf{A} \underline{\mu} = 4 \sum_{i=1}^4 \underbrace{\begin{bmatrix} \underline{\mu}^{(i)T} & \dots & \underline{\mu}^{(i)T} \end{bmatrix}}_{N_i} \mathbf{A}_i \Sigma \mathbf{A}_i \begin{bmatrix} \underline{\mu}^{(i)} \\ \vdots \\ \underline{\mu}^{(i)} \end{bmatrix} \tag{20}$$

where

$$\begin{bmatrix} \underline{\mu}^{(i)T} & \dots & \underline{\mu}^{(i)T} \end{bmatrix} \mathbf{A}_i = 0 \tag{21}$$

because of the semidefiniteness of \mathbf{A}_i . Similarly, $\underline{\mu}^T \mathbf{A} \underline{\mu} = 0$. The mean of ξ is then given by $\frac{1}{2\sigma^2} [tr(\mathbf{A} \Sigma)]$. For $\mathbf{A} = \text{block diag} \{ \mathbf{A}_1 \eta_1, \mathbf{A}_2 \eta_2, \mathbf{A}_3 \eta_3, \mathbf{A}_4 \eta_4 \}$ we have

$$E[\xi] = \frac{1}{2\sigma^2} tr(\mathbf{A} \Sigma) = \frac{1}{2} \sum_{i=1}^4 N_i (N_i - 1) P \eta_i. \tag{22}$$

Additionally $\mathbf{A} \mathbf{A} = \text{block diag} \{ \mathbf{A}_1 \mathbf{A}_1 \eta_1^2, \mathbf{A}_2 \mathbf{A}_2 \eta_2^2, \mathbf{A}_3 \mathbf{A}_3 \eta_3^2, \mathbf{A}_4 \mathbf{A}_4 \eta_4^2 \}$ where

$$\mathbf{A}_i \mathbf{A}_i = \begin{bmatrix} (N_i^2 - N_i) \mathbf{I} & -N_i \mathbf{I} & \dots & -N_i \mathbf{I} \\ -N_i \mathbf{I} & (N_i^2 - N_i) \mathbf{I} & \dots & -N_i \mathbf{I} \\ \dots & \dots & \ddots & \dots \\ -N_i \mathbf{I} & -N_i \mathbf{I} & \dots & (N_i^2 - N_i) \mathbf{I} \end{bmatrix}. \tag{23}$$

Here \mathbf{I} is of size $P \times P$ and \mathbf{A}_i is of size $N_i P \times N_i P$. We then have

$$tr(\mathbf{A}_i)^2 = (N_i^2 - N_i)N_i P \quad tr(\mathbf{A})^2 = \sum_{i=1}^4 tr(\mathbf{A}_i)\eta_i^2 = \sum_{i=1}^4 (N_i^2 - N_i)N_i P \eta_i^2.$$

so that $Var[\xi] = \frac{1}{4\sigma^4} 2tr(\mathbf{A}\Sigma)^2 = \frac{1}{2} \sum_{i=1}^4 (N_i^2 - N_i)N_i P \eta_i^2$.

REFERENCES

1. J. Bourgeois and G. Smith, "A Fully Three-Dimensional Simulation of a Ground-Penetrating Radar: FDTD Theory Compared with Experiment," *IEEE Trans. on Geoscience and Remote Sensing* **34**, pp. 36–44, Jan. 1996.
2. C. Rappaport and D. Reidy, "Focused Array Radar for Real Time Imaging and Detection," *Proc. SPIE, Radar Sensor Technology* **2747**, pp. 202–213, 1996.
3. C. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons, Inc., New York, New York, 1989.
4. S. Kullback, *Information Theory and Statistics*, Dover, Mineola, New York, 1997.
5. E. Lehman, *Testing Statistical Hypotheses*, Springer-Verlag, New York, New York, 1986.
6. A. Mathai and S. Provost, *Quadratic Forms in Random Variables: Theory and Applications*, Marcel Dekker, Inc., New York, New York, 1992.
7. C. Helstrom, *Elements of Signal Detection and Estimation*, Prentice-Hall, Englewood Cliffs, New Jersey, 1995.
8. S. Mason, G. Birch, and M. Ito, "Improved Single-Trial Signal Extraction of Low SNR Events," *IEEE Trans. on Signal Processing* **42**, pp. 423–426, Feb. 1994.
9. B. Ghosh, *Sequential Tests of Statistical Hypotheses*, Addison-Wesley, Reading, Mass., 1970.

Nonlinear inverse scattering methods for thermal wave slice tomography

by Eric Miller¹, Lena Nicolaides² and Andreas Mandelis²

¹Dept. Electrical and Computer Engineering, Northeastern University, Boston, MA 02215, USA; ²Photothermal and Optoelectronic Diagnostics Laboratories Dept. of Mechanical Engineering, Univ. of Toronto, Toronto M5S 3G8, CANADA

Abstract

A wavelet domain, non-linear inverse scattering approach is presented for imaging sub-surface defects in a material sample based on observations of scattered thermal waves. We use newly developed wavelet-based regularization methods to resolve better the edge structures of defects relative to reconstructions obtained with smoothness-type regularizers. A non-linear approximation to the exact forward scattering model is introduced to simplify the inversion with little loss in accuracy. We demonstrate this approach on cross-section imaging problems using synthetically generated scattering data from transmission and backprojection geometries.

Keywords: Thermal wave slice tomography, inverse scattering, wavelet transform, edge preserving regularization

1. Introduction

Thermal wave slice tomography (TWST) has evolved in recent years as a useful tool for non-invasively imaging and detecting defects in the bulk properties of a material sample [1, 2]. This non-destructive evaluation (NDE) technique makes use of a modulated laser source illuminating an external surface of the material under test to induce internal thermal waves. Interaction of the thermal wave field with material inhomogeneities gives rise to scattered fields which propagate and are ultimately measured at the surface of the material. The problem of interest in this paper is, given knowledge of the applied thermal wave field and the observed scattered fields, to produce a reconstruction of the internal structure which reproduces as faithfully as possible features of interest such as defects.

The techniques we use to solve this inverse problem are based on the results of Mandelis who has shown that the thermal wavefield obeys a scalar Helmholtz equation with a complex valued, space varying propagation constant [3]. The spatial structure of this constant is related to the thermal diffusivity of the material. Because defects are reflected in changes in the thermal diffusivity, a reconstruction of the propagation constant, or a normalized form known as the object function [1], yields quantitative information about the material's bulk structure.

In this work, we extend the use of linearized wavefield inversion methods initially developed in [1, 2] for the TWST problem in a number of ways. First, a full, non-linear inverse scattering approach is used to generate the reconstruction. The resulting inversion method is iterative in nature and may allow for the more accurate reconstruction of defects whose structure falls outside of the bounds where the Born approximation is valid. We also build on our previous efforts in wavelet-based non-linear inverse scattering techniques [4, 5]. First, the TWST inverse problem is highly ill-posed in the sense discussed in [6]. That is, small perturbations in the data, as would come from noise, can result in reconstructions with high amplitude, oscillatory structure. We introduce and demonstrate the utility of a class of wavelet regularization schemes which are appropriate for the reconstruction of objects (defects) with sharply defined boundaries. Tikhonov type regularizers typically are designed to produce smooth, low-pass reconstruc-

tions which blur important features such as edges in the image [6]. Using certain mathematical results stating that wavelets are bases for a wide range of function spaces including spaces containing “edgy” objects [7], we develop a new scale-space regularization method which produces significantly sharper reconstructions.

Additionally, we employ wavelet methods to lower the computational burden of the inversion procedure. The wavelet transform is known to make sparse the matrix representations of many integral operators including those arising in the TWST scattering problem [8]. Thus, a transform domain formulation can build on this sparsity to reduce the computational burden of generating a reconstruction. We also develop a wavelet-based, reduced complexity approximation to the forward TWST problem. At each stage of the iterative solution method we must compute explicitly the inverse of a large, dense matrix, a computationally intensive task. This matrix is related to the solution of the thermal-wave forward scattering problem for an object function equal to the reconstruction obtained at the previous iteration plus a small correction. By linearizing the expression for this matrix about this correction we obtain a recursive formula in which the inverse matrix at the current iteration is equal to the one from the last iteration plus an easily-constructed increment due to the change in the material properties.

The remainder of this paper is organized as follows. In Section 2. the mathematical model for the TWST problem is defined. Section 3. is devoted to the application of the wavelet transform to the TWST problem. The non-linear inverse scattering algorithm is developed in Section 4. with examples of its application to synthetic problems provided in Section 5.. In Section 6., conclusions are discussed.

2. Physical Model for TWST

As illustrated in *Fig. 1*, we consider TWST problems in which a modulated laser illuminates a point at the top of the material sample inducing thermal waves in the bulk. Upon scattering from inhomogeneities, whose spatial structure is defined by the real-valued object function the field is measured along an array of points either at the top (backpropagation geometry) or the bottom (transmission mode). The inversion routines are based on K such scattering experiments. The data for the i th experiment form the vector of in-phase and quadrature components of the scattered thermal-wave field measurements obtained along one of the arrays due to illumination by the laser at a given point on the top surface.

As discussed in [1], the object function is related to the measured data non-linearly through a pair of coupled integral equations. Our algorithms are based on discrete representations of these equations obtained using the method of moments [9] with a pulse basis and Dirac testing functions. Upon discretization the model takes the form

$$\mathbf{y}_i = \mathbf{L}_i \mathcal{D}(\mathbf{T}_i) \mathbf{g} + \mathbf{n}_i \quad (1)$$

$$\mathbf{T}_i = \bar{\mathbf{T}}_i + \mathbf{G} \mathcal{D}(\mathbf{g}) \mathbf{T}_i \quad (2)$$

where \mathbf{y}_i is the vector of observations along the array for the i th experiment, \mathbf{g} and \mathbf{T}_i are vectors containing the lexicographically ordered pulse basis expansion coefficients for the object function and the thermal fields internal to the medium respectively, $\bar{\mathbf{T}}_i$ is the discrete representation of the thermal field in the absence of defects, $\mathcal{D}(\mathbf{x})$ is a diagonal matrix whose entries are the elements of the vector \mathbf{x} , and \mathbf{L}_i and \mathbf{G} are the matrices obtained by discretizing the integral kernels in the underlying continuous model. \mathbf{n}_i are taken to be mutually uncorrelated, zero mean, white Gaussian noise vectors. By solving for \mathbf{T}_i in (2), and substituting the result into (1), the data are related to the object function via the non-linear model

$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{g}) + \mathbf{n}_i \quad \text{with} \quad \mathbf{h}_i(\mathbf{g}) = \mathbf{L}_i \mathcal{D} [(\mathbf{I} - \mathbf{G} \mathcal{D}(\mathbf{g}))^{-1} \bar{\mathbf{T}}_i] \mathbf{g}. \quad (3)$$

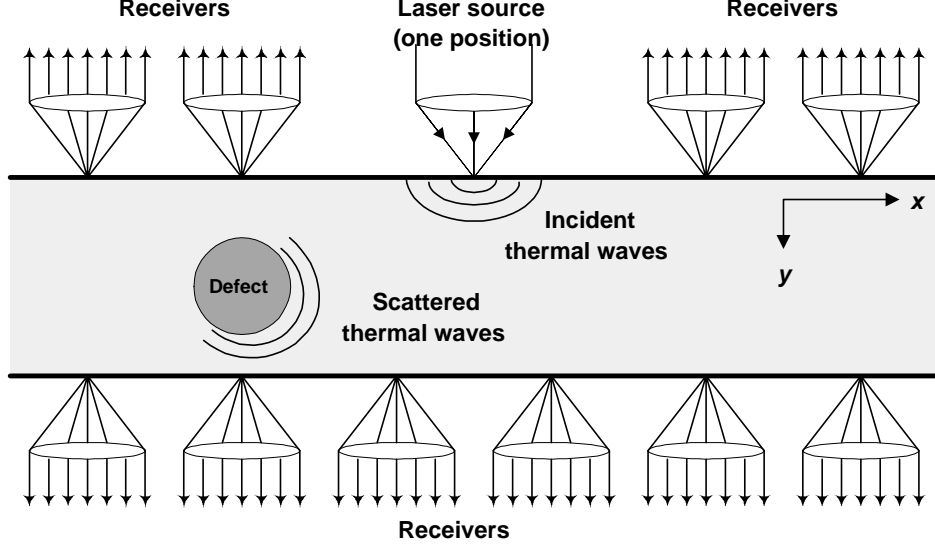


Figure 1: *Experimental setup for thermal wave slice tomography.*

Given the model in (3), the TWST problem is to determine g based on the data, y_i from K scattering experiments.

3. Wavelet Domain Model

Here we employ discrete, orthonormal wavelet transforms to take the inverse problem from physical to scale space. As described more fully in [10], given a signal or image \mathbf{a} , we obtain its wavelet transform, denoted α via multiplication by a unitary wavelet transform matrix, \mathbf{W}_a . That is, $\alpha = \mathbf{W}_a^T \mathbf{a}$ and $\mathbf{a} = \mathbf{W}_a^T \alpha$. The transform of \mathbf{a} is composed of wavelet coefficients over the scales and dyadic shifts of interest in the particular problem under investigation. We subscript the matrix \mathbf{W}_a to make explicit that this is the transform for \mathbf{a} . We will use different wavelet transforms for the different variables.

We use the discrete wavelet transform to take (1) and (2) from physical space to scale space. Defining \mathbf{W}_g as the 2D transform for g and \mathbf{W}_i for $i = 1, 2, \dots, K$ as the 1D transforms for y_i yields:

$$\begin{aligned} \mathbf{W}_i \mathbf{y}_i &= [\mathbf{W}_i \mathbf{L}_i \mathbf{W}_g^T] [\mathbf{W}_g \mathcal{D}(\mathbf{T}_i) \mathbf{W}_g^T] (\mathbf{W}_g \mathbf{g}) + \mathbf{W}_i \mathbf{n}_i \\ \mathbf{W}_i \mathbf{T}_i &= \mathbf{W}_i \bar{\mathbf{T}}_i + [\mathbf{W}_i \mathbf{G} \mathbf{W}_g^T] [\mathbf{W}_g \mathcal{D}(\mathbf{g}) \mathbf{W}_g^T] (\mathbf{W}_i \mathbf{T}_i) \end{aligned}$$

which, upon making the obvious definitions become

$$\eta_i = \Lambda_i \Delta(\theta_i) \gamma + \nu_i \quad (4)$$

$$\theta_i = \bar{\theta}_i + \Gamma \Delta(\gamma) \theta_i \quad (5)$$

where, $\Delta(\gamma)$ is a function of γ since $\Delta(\gamma) = \mathbf{W}_g \mathcal{D}(\mathbf{W}_g^T \gamma) \mathbf{W}_g^T$. From (4) and (5), the complete model relating the transform of the object to those of the data is

$$\eta_i = \Upsilon_i(\gamma) + \nu_i \quad \text{with} \quad \Upsilon_i(\gamma) = \Lambda_i \Delta \left[(\mathbf{I} - \Gamma \Delta(\gamma))^{-1} \bar{\theta}_i \right] \gamma. \quad (6)$$

Finally, we aggregate the η_i into a single data vector and define the stacked system $\eta = \Upsilon(\gamma) + \nu$ with $\eta = [\eta_1^T \dots \eta_K^T]$ and $\Upsilon(\gamma)$ and ν defined accordingly. Thus, the TWST inverse problem is the recovery of γ , the DWT of g , from measurements η , knowledge of Υ , and the statistics of ν .

4. Inversion Algorithm

In this paper, the reconstruction of γ , denoted by $\hat{\gamma}$, given the data, η , is defined to be the solution to the following non-linear least squares type of optimization problem

$$\hat{\gamma} = \arg \min_{\gamma} \mathcal{C}(\gamma) \quad (7)$$

$$\mathcal{C}(\gamma) = \frac{1}{2} \|\eta - \Upsilon(\gamma)\|_{\mathcal{R}^{-1}}^2 + \lambda^2 \rho^T(\gamma) \rho(\gamma) \quad (8)$$

where $\|\mathbf{x}\|_{\mathbf{A}} \equiv \mathbf{x}^T \mathbf{A} \mathbf{x}$, \mathcal{R} is a diagonal weighting matrix whose entries reflect the noise levels in the data, $\rho^T(\gamma) \rho(\gamma)$ is used to regularize the problem, and λ^2 is the regularization parameter. We employ a form of the Levenberg-Marquardt algorithm (LMA) for finding $\hat{\gamma}$. This iterative technique defines a sequence of reconstructions, $\hat{\gamma}_n$, whose costs as measured by (8) are steadily decreasing. Starting from an initial guess, $\hat{\gamma}_0$, the form of the LMA used here is

$$\begin{aligned} \hat{\gamma}_{n+1} &= \hat{\gamma}_n + \mathbf{s}_n \\ \mathbf{s}_n &= \arg \min_{\sigma^2} \mathcal{C}(\hat{\gamma}_n + \mathbf{s}) \end{aligned} \quad (9)$$

$$\mathbf{s} = [\mathcal{J}^T(\hat{\gamma}_n) \mathcal{R}^{-1} \mathcal{J}^T(\hat{\gamma}_n) + \sigma^2 \mathcal{L}^T(\hat{\gamma}_n) \mathcal{L}(\hat{\gamma}_n)]^{-1} \times \mathcal{J}^T(\hat{\gamma}_n) \mathcal{R}^{-1} [\eta - \Upsilon(\hat{\gamma}_n)] - \mathcal{L}^T(\hat{\gamma}_n) \mathcal{L}(\hat{\gamma}_n). \quad (10)$$

In (10), $\mathcal{J}(\hat{\gamma}_n)$ (resp. $\mathcal{L}(\hat{\gamma}_n)$) is the Jacobian matrix of Υ (resp. ρ) evaluated at the vector $\hat{\gamma}_n$ and σ is a regularization parameter whose value is determined adaptively at each iteration of the algorithm.

4.1. Edge Preserving Wavelet Regularization

In our previous work [4, 5], we have concentrated on the use of wavelet domain regularizers with $\rho(\gamma) = \mathbf{D}\gamma$ where the matrix \mathbf{D} was diagonal with

$$[\mathbf{D}]_{ii} \equiv d_i = 2^{-(\alpha_x j_{x,i} + \alpha_y j_{y,i})}. \quad (11)$$

In (11) $j_{x,i}$ and $j_{y,i}$ are the horizontal and vertical scale indices for the i th wavelet coefficient and α_x and α_y are constants. For this regularization approach, $\rho^T(\hat{\gamma}_n) \rho(\hat{\gamma}_n)$ is a weighted two-norm of $\hat{\gamma}_n$ and functions in much the same way as traditional smoothness type regularizers [4, 5]. One consequence of this choice is that edges tend to be blurred in the reconstruction.

Recently, there has been considerable work performed in the area of “edge-preserving” regularizers [11]. The idea is to construct a physical-space regularization scheme which results in reconstructions whose discontinuities are better preserved as compared to a Tikhonov approach [12]. One way of implementing this regularization technique is to choose an expression for $\rho^T \rho$ which is a norm in a function space containing “edgy” objects. Adding this as the second term in (8) produces an object which lies in such a space and therefore retains the desired edge-like structure.

It is the case that, in addition to spanning the space of square integrable functions, orthonormal wavelets are also bases for these more exotic function spaces [7]. Here we make use of the fact that the norm in such a space may be computed in terms of the wavelet coefficients via $\sum_i d_i |\gamma_i|^p$ with $1 \leq p \leq 2$ and d_i as in (12). With this motivation, we make the following choice for $\rho(\gamma)$

$$\rho^T(\gamma) = \left[d_1^{1/2} |\gamma_1|^{p/2} \quad d_2^{1/2} |\gamma_2|^{p/2} \quad \dots \quad d_N^{1/2} |\gamma_N|^{p/2} \right]. \quad (12)$$

4.2. Approximate Physical Model

A large computational burden associated with the LMA is the need to invert the matrix $\mathbf{I} - \Gamma\Delta(\hat{\gamma}_n)$ for numerous $\hat{\gamma}_n$. This operation amounts to explicitly inverting the generally large and dense system matrix for the forward scattering problem. To bypass this operation we note that at the beginning of stage $n + 1$ of the LMA, we require

$$\Xi_{n+1} \equiv [\mathbf{I} - \Gamma\Delta(\hat{\gamma}_{n+1})]^{-1} = [\mathbf{I} - \Gamma\Delta(\hat{\gamma}_n + \mathbf{s}_n)]^{-1}.$$

Making use of the fact that Δ is a linear operator [5],

$$\Xi_{n+1} = [\mathbf{I} - \Gamma\Delta(\hat{\gamma}_n + \mathbf{s}_n)]^{-1} = [\mathbf{I} - \Gamma\Delta(\hat{\gamma}_n) - \Gamma\Delta(\mathbf{s}_n)]^{-1} = \Xi_n [\mathbf{I} - \Gamma\Delta(\mathbf{s}_n)\Xi_n]^{-1} \quad (13)$$

$$\approx \Xi_n + \Xi_n \Gamma\Delta(\mathbf{s}_n)\Xi_n \quad (14)$$

where (14) follows from (13) under the assumption that $\Gamma\Delta(\mathbf{s}_n)\Xi_n$ is “small” relative to \mathbf{I} . We note that, unlike the Born approximation, this is not a linearization of the physics about the current estimate of the object function. Additionally, our approximation provides a recursive method for updating Ξ_n from one iteration of the LMA to the next.

5. Example of Thermal-Wave Tomographic Imaging

To illustrate the inversion algorithm developed in the previous sections, we consider the imaging of a square hole centered near the top of a 3 mm by 3 mm block of aluminum (thermal diffusivity of $0.82 \text{ cm}^2\text{s}^{-1}$). The object function for this case is shown in *Fig. 2(a)* is of amplitude $(\alpha_{\text{Al}}/\alpha_{\text{air}}) - 1 \approx 3$ [3] and the SNR is 50dB. The reconstruction obtained under the Born approximation is displayed in *Fig. 2(b)*. For this case, the Born approximation yields coarse-scale localization of the defect. That is, the reconstruction is non-zero over a region of space which includes the area of the true defect. However, the amplitude of the reconstruction is at best a third of the true amplitude and the shape of the reconstructed defect is in fact larger than that of the true structure.

The final results of the LMA using $p = 2$ (which corresponds to a traditional Tikhonov smoothness regularization scheme) and $p = 1.2$ (an edge preserving case) are shown in *Figs. 2, (c), and (d)* respectively. In *Fig. 2(e)*, the value of $\|\hat{\gamma}_n - \gamma\|_2$ is plotted for both regularization schemes as a function of iteration. It is evident that both regularization schemes produce reconstructions which are better localized with more accurate amplitude information than the Born inversion. The $p = 1.2$ case is a slightly better than the $p = 2$ reconstruction both from a quantitative and qualitative perspective. Quantitatively, the error in the $p = 1.2$ estimate is somewhat lower than that of the $p = 2$ case after 10 iterations. This improvement is due to the fact that the $p = 2$ case overestimates the amplitudes in a couple of pixels, while the amplitude of the $p = 1.2$ version is quite accurate. Visually, as we expect, the edges on the $p = 1.2$ reconstruction are sharper and the overall reconstruction looks much more like a box than a smoothed blob as in *Fig. 2(c)*.

6. Conclusions

We have presented a new approach to image formation from scattered thermal waves based on the use of non-linear inverse scattering methods and wavelet-domain techniques. We build our inversion routine on the full wavefield physics developed by Mandelis in [3] resulting in a highly non-linear relationship between the data on which a reconstruction is to be based and the desired image of the object function. The reconstruction problem was formulated as a solution to a non-linear least-squares optimization problem in the wavelet transform domain. We chose to work in a multiscale setting for a number of reasons. First, the matrices comprising the physical model are sparse in this domain thereby lowering the

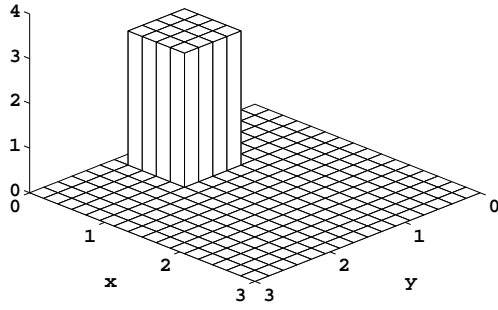
computational cost of generating a reconstruction. Second, we were able to make use of a new class of edge-preserving regularization methods which are easily specified and implemented in the wavelet transform domain. Finally, the computational burden was further reduced by employing the methods of [5] for rapidly solving non-linear inverse scattering problems in a multiscale domain.

7. Acknowledgments

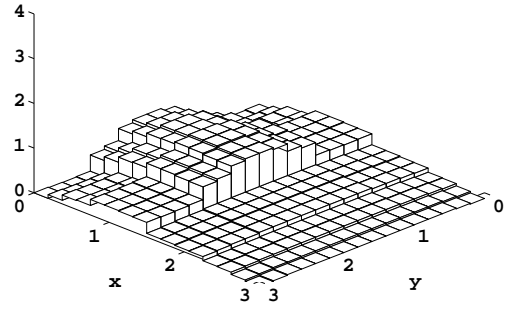
The work of the first author was supported in part DOE contract DE-FC07-95ID13395, NSF Grant MIP-9623721, and by subcontract GC123920NDG from Boston University under the AFOSR MURI Program on Reduced Signature Target Recognition. The work of the other authors was supported by the Natural Sciences and Engineering Research Council of Canada, through a Research Grant.

References

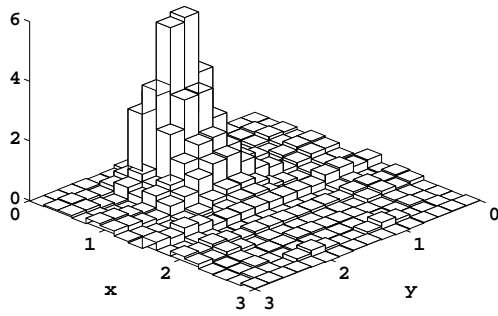
- 1] NICOLAIDES (L.) and MANDELIS (A.).- Image-enhanced thermal-wave slice diffraction tomography with numerically simulated reconstructions: Theoretical and computational. *Inverse Problems*, in press.
- 2] PADE (O.) and MANDELIS (A.).- Thermal-wave slice tomography using wave-field reconstruction. *Inverse Problems*, 10:185–197, 1994.
- 3] MANDELIS (A.).- Theory of photothermal-wave diffraction and interference in condensed media. *J. Opt. Soc. Am.*, A6(298), 1989.
- 4] MILLER (E.) and WILLISKY (A.).- Multiscale, statistically-based inversion scheme for the linearized inverse scattering problem. *IEEE Trans. Geosc. Remote Sens.*, 34(2):346–357, March 1996.
- 5] MILLER (E.) and WILLISKY (A.).- Wavelet-based methods for the nonlinear inverse scattering problem using the Extended Born Approximation. *Radio Sci.*, 31(1):51–67, Jan–Feb 1996.
- 6] BERTERO (M.).- Linear inverse and ill-posed problems. In Hawkes Peter W, editor, *Advances in Electronics and Electron Physics*, volume 75, pages 1–120. Academic Press, Boston, 1989.
- 7] MEYER (Y.).- *Wavelets and Operators*. Cambridge, Cambridge, England, 1995.
- 8] BEYLKIN (G.), COIFMAN (R.), and ROKHLIN (V.).- Fast wavelet transforms and numerical algorithms, I. *Commun. Pure Appl. Math.*, 44:141–183, 1991.
- 9] HARRINGTON (R.).- *Field Computations by Moment Methods*. Macmillan, New York, 1968.
- 10] STRANG (G.) and NGUYEN (T.).- *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, 1996.
- 11] CHARBONNIER (P.), BLANC-FERAUD (L.), AUBERT (G.), and BARLUND (M.).- Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Processing*, 6(2):298–311, February 1997.
- 12] NICOLAIDES (L.), MUNIDASA (M.), and MANDELIS (A.).- Thermal-wave diffraction tomographic microscopy. In *III International Workshop Advances in Signal Processing for Nondestructive Evaluation of Materials*, Universite Laval, Quebec City, Canada, August 5–8 1997.



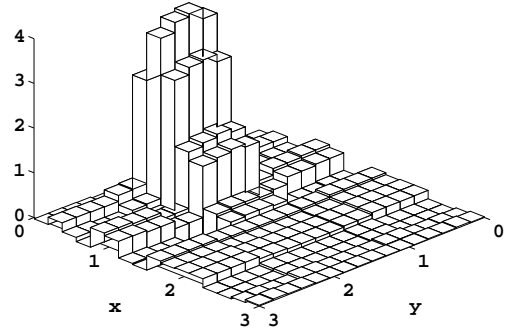
(a) Object Function



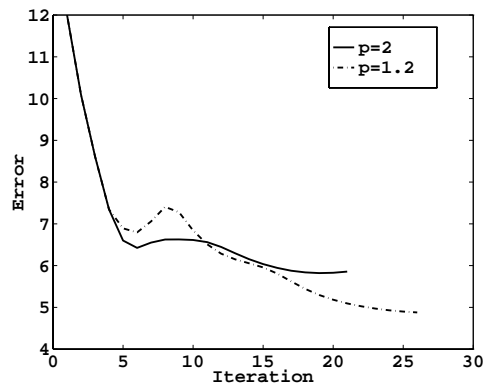
(b) Born reconstruction



(c) 2 norm reconstruction



(d) p norm reconstruction



(e) $\|\hat{\gamma}_n - \gamma\|_2$

Figure 2: True object function, inversion results and error vs. iteration for single defect located near the top of the material. The laser source position is assumed to be along the line $x = 0$.

On the detection of buried mines from array inductive measurements

Eric L. Miller^a, W. Clem Karl^b, and Stephen J. Norton^c

^a Center for Electromagnetics Research, Northeastern University, Boston, MA.

^b Multidimensional Signal Processing Laboratory, Boston University, Boston, MA 02215

^c Oak Ridge National Laboratory, Oak Ridge, TN 37831-6200

ABSTRACT

The problem of mine detection and localization from array-based electromagnetic induction data is addressed. An efficient forward scattering model based on the Born approximation is employed. Using insight obtained from this model, a clutter model in the form of a state space system is developed which describes the correlation of the noise both across the sensing array and from one position of the array to the next as the measurement device proceeds down track. A multiple-model detection scheme based on the whitening properties of the Kalman filter is employed to perform the actual mine detection. This approach allows for the detection and localization of buried objects well before the array physically moves over mines. Examples are provided for mines buried directly in front and off to one side of the array.

Keywords: Mine detection, EMI arrays, Kalman filter

1. INTRODUCTION

The problem of detecting and localizing buried landmines from indirect observations of scattered electromagnetic radiation has received considerable attention in recent years. Of interest in this paper is the development of statistical signal processing methods for solving such problems based on data collected using an array of very low frequency electromagnetic induction sensors. From a signal processing perspective, the physics of this problem represents an unusual challenge. The majority of array processing algorithms are based on the assumption that the observed waves are propagating in free space and the unknown sources are in the far field. The resulting planewave model for the radiation incident on the sensors serves as the basis for all beamforming and high-resolution localization methods. For the applications of interest here, the objects are in the nearfield of the array and while the sensors are in freespace, the targets are located in a lossy halfspace. The typical array approaches are therefore not applicable.

In addition to the basic physics describing these problems, the data are typically obtained by sweeping the array in a fixed direction over a region of interest, as illustrated in Figure 1. The result is an “image” acquired over time whose number of rows is equal to the number of sensors in the array and whose number of columns is determined by the number of stops the array makes as it sweeps over the area. This implied temporal structure is the second unique image processing element present in this problem. In particular, it is imperative to detect the mine *before* the sensor array actually passes over its location. This requirement implies that any algorithm must process the image data in a real-time, column-by-column manner, continuously updating its decision as to mine presence and its estimate concerning the object’s location as new data are obtained by the scanning procedure.

Other author information: ELM: elmiller@ece.neu.edu WCK: wckarl@bu.edu. SJN nortonsj@ornl.gov. This work was supported by the Army Research Office Demining MURI under Grant DAAG55-97-1-0013 and the Air Force Office of Scientific Research under Grant F49620-96-1-0028

The final challenging characteristic of this problem is the need to process data contaminated by “clutter.” Physically, such clutter can arise from a variety of inhomogeneities present in the soil including a random distribution of rocks and other discrete structures as well as the continuous, natural changes in the electromagnetic characteristics of the earth. From a signal processing perspective, we regard clutter as a correlated noise source present in the data. The correlation is evident first along the array from one sensor to the next and second along the collection track from one array position to the next. To adequately detect mines whose signals are obscured by such noise requires both careful modeling of the clutter and the development of processing methods explicitly based on these models.

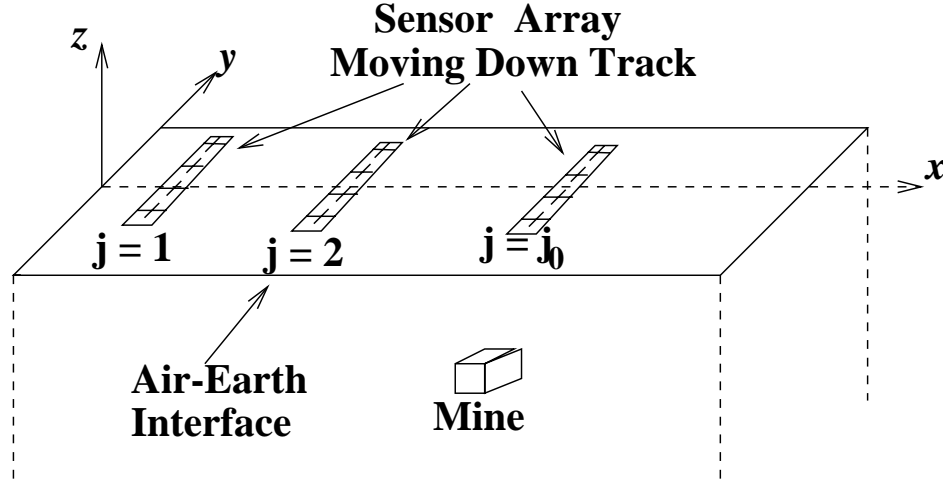


Figure 1. Data Acquisition Method for Subsurface Detection Problem

In this paper, we consider statistical signal processing methods for solving the mine detection problem outlined in the previous paragraphs. Our focus is on shallowly buried metallic anti-personal mines arising in the case of humanitarian demining. In Section 2, we describe a physical model relating the characteristics of the buried mine to the observations obtained over an array. We next use this physical model as the basis for a simple, analytical, stochastic model describing the distribution of clutter. In particular, the clutter model takes the form of a so-called “state-space” system. This system accounts both for the inter-element and inter-array correlation of the noise. In Section 4, we take advantage of the fact that the Kalman filter can be used to decorrelate (or whiten) such noise processes to obtain a method for detecting mines. At each stop of the array, we perform a series of hypothesis tests. Under the null hypothesis for each test, no mine is present. Under the alternative, there is a mine located at one of a number of positions over a coarse scale grid. The whitening properties significantly simplify the formulation of the tests while retaining the optimality property of the procedure. We stop the array and declare the presence of a mine when the sequence of tests for a mine at a hypothesized location are “sufficiently convincing.” We show in the examples in Section 5 that this process is capable of detecting mines reliably (a) up to 20cm before the array actually passes over the mine and (b) under low clutter conditions, when the mine is located up to 30 cm from one side or another of the array. Finally, conclusions and future work are described in Section 6.

2. FORWARD MODEL

To fix notation, we assume that the array is situated in the $x - y$ plane and proceeds from $x = 0$ to $x = x_{max}$ stopping at N_s equally spaced locations to obtain data. The array is composed of a single

transmitter at $y = 0$ and N_r receivers equally spaced on either side of the array. All transducers are taken to be small current loops of radii r_p located z_p above the air-earth interface carrying I A of current with an $e^{-i\omega t}$ time dependence. We define $d_{m,n}$ to be the field measured by the m th element of the array at the n th stop between $x = 0$ and $x = x_{max}$. In this paper, we assume that the mine represents a small perturbation of constant conductivity σ S/m in the electromagnetic properties of the earth so that the first Born approximation¹ can be used to obtain a tractable physical model. For the case of a single mine, it can be shown that the relationship between the physical properties of the mine and y_{mn} is given by¹

$$d_{nm} = -\frac{\sigma(\omega\mu I r_p)^2}{4} \int_V A(\mathbf{r} - \mathbf{r}_0 + \mathbf{r}_n) A(\mathbf{r} - \mathbf{r}_0 + \mathbf{r}_m) d\mathbf{r} \quad (1)$$

where the quantity A represents the Green's function for the problem of a loop over a half-space, $\mathbf{r} = [x \ y \ z]^T$ is a point in three space represented in either Cartesian or polar coordinates, \mathbf{r}_0 is the position of the center of the mine, \mathbf{r}_n (\mathbf{r}_m) is the position of the transmitter (receiver), μ is the magnetic permeability and the integral is taken over the volume, V , of the prototypical mine centered at the origin. In [1], it is shown that the Green's function A can be well approximated as

$$A(\mathbf{r}) = \frac{2}{\pi k(\mathbf{r}) \sqrt{r r_p}} \left[\left(1 - \frac{1}{2} k(\mathbf{r})^2 \right) K(k(\mathbf{r})) - E(k(\mathbf{r})) \right] \quad (2)$$

$$k(\mathbf{r}) = \frac{4 r r_p}{(z + z_p)^2 + (r + r_p)^2} \quad (3)$$

where $r = \|\mathbf{r}\|$ is the distance to the physical point, $K(x)$ is the complete elliptic integral of the first kind, $E(x)$ is the complete elliptic integral of the second kind and the parameter $k(\mathbf{r})$ is in the range 0 to 1.

For the problem of interest in this paper, we have found by direct evaluation that $k(\mathbf{r}) \in [0, 0.6]$. Under this restriction, we have developed the following highly accurate, yet simple approximation to (2):

$$A(\mathbf{r}) = \frac{2}{\pi k(\mathbf{r}) \sqrt{r r_p}} \left(\beta_4 k^4(\mathbf{r}) + \beta_6 k^6(\mathbf{r}) \right) \quad (4)$$

with $\beta_4 = 0.0882$, $\beta_6 = 0.1321$ and $k(\mathbf{r})$ as in (3). The particular values of the β parameters are obtained by a least squares fit and the use of the fourth and sixth powers of k is motivated by the Taylor series expansion for the right hand side of (2). A semilogarithmic comparison of the right hand sides of (2) and (4) is shown in Figure 2. The solid line represents the exact combination of elliptic integrals while the dashed line is the polynomial approximation. It is clear that the approximation is, for all intents and purposes, the same as the exact expression. From a computational perspective however, (4) offers definite advantages over (2) in terms of speed of evaluation and ease of manipulation.

Given this model, the goals of the processing in this paper are the detection of the presence of a mine from noisy observations of d_{ij} and determination of \mathbf{r}_0 once a detection has been declared.

3. CLUTTER MODEL

Typically, white Gaussian noise is used as the clutter model for mine detection problems due to its tractability. Unfortunately, this simplified assumption leads to a degradation in detection performance. In this work we instead explicitly model the clutter or background process, leading to more robust performance. In particular, we develop a tractable yet realistic stochastic model describing the distribution of the data obtained over our receiver in the absence of a target. This statistical model then forms the basis for a decision-theoretic procedure for the localization of buried mines discussed in the remainder of this paper.

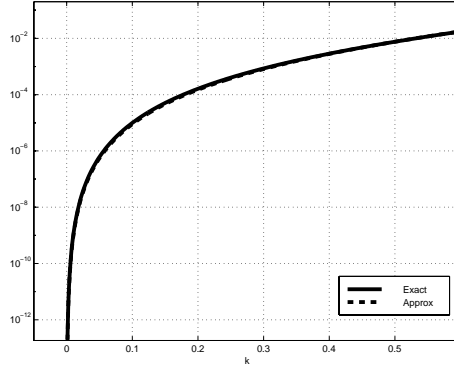


Figure 2. Comparison of exact and polynomial approximation to forward scattering model

The motivation for the mathematical form of our clutter model is actually provided by the nature of the signal observed over the array due only to a mine buried some distance away. The wavefield over the aperture of the array is rather smooth when a mine does in fact exist near the array. In Fig. 3, noiseless simulated data is shown for the case of a $.05 \times .05 \times .05$ m³ mine located at $\mathbf{r}_0 = \{x_0, y_0, z_0\} = \{2, .1, -.05\}$ m where an 8 element array of length .6m is positioned at $x \in \{0, .5, 1, 1.5\}$ m. These plots indicate the more general feature that the data over the array can be well modeled by a low order polynomial in y_i , the y position of the i th sensor in the array. In other words we can write

$$d_{m,n} = \sum_{j=0}^J \alpha_{j,n} y_m^j. \quad (5)$$

By defining the vectors $\mathbf{d}^T(n) = [d_{1,n} \ d_{2,n} \ \dots \ d_{N_r,n}]$ and $\boldsymbol{\alpha}^T(n) = [\alpha_{0,n} \ \alpha_{1,n} \ \dots \ \alpha_{J,n}]$, (5) is rewritten as

$$\mathbf{d}(n) = \mathbf{C}\boldsymbol{\alpha}(n). \quad (6)$$

where the (m, j) th element of \mathbf{C} is y_m^j .

The clutter model we employ in this work is based on the assumption that the random signal over the array in the absence of a mine is the sum of two components. The first portion is an additive zero mean, white Gaussian component of variance γ^2 representing sensor noise. The second component is intended to capture the effects due to the statistical distribution of the environmental factors not of primary interest in the detection problem. Here we assume that this signal will display the same basic smoothness as that seen in Fig. 3. That is it will be described by a low order polynomial. In the case of a mine, the coefficients of the polynomial are deterministically related to the position of the object relative to the array. For clutter however, these coefficients are taken to be random quantities. That is, $\boldsymbol{\alpha}(n)$ is a random vector. Such a model arises if we assume that the clutter represents the average effect on the sensor array of a random distribution of small scatterers. If each such scatterer results in a smooth signal over the array then so too will their sum. We note here that this choice of model certainly satisfies the criteria set forth in the Introduction that clutter essentially represents noise which is correlated from one array element to the next.

The second criteria is that the clutter should be correlated from one position of the array to the next. For example, a clutter vector at position n should be correlated to that at position $n + 1$. This implies

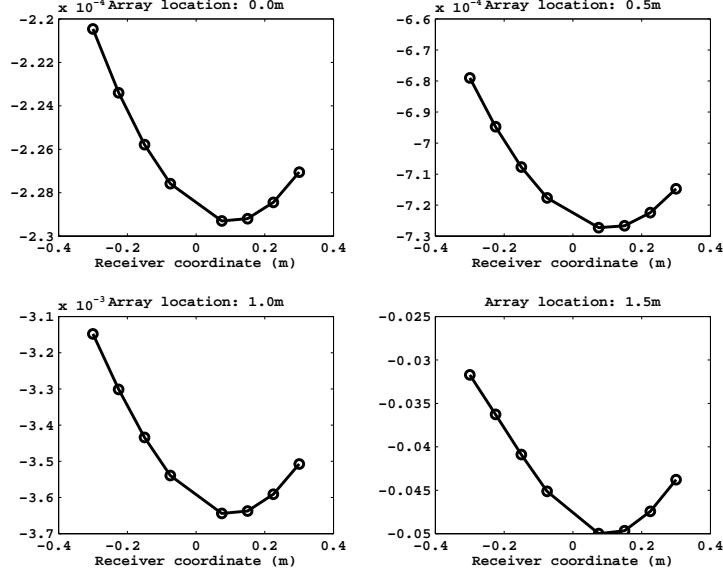


Figure 3. Data vectors, \mathbf{d}_j , obtained as array proceeds toward buried mine. Note the scale change in the plots.

the need to specify a dynamic model describing the evolution of $\boldsymbol{\alpha}(n)$. Toward this end, we employ the following model of the spatial evolution of $\boldsymbol{\alpha}(n)$:

$$\boldsymbol{\alpha}(n+1) = \boldsymbol{\alpha}(n) + \mathbf{B}(n)\mathbf{w}(n) \quad (7)$$

where $\mathbf{B} = \nu \text{diag}[\delta^{J/2}, \delta^{(J-1)/2}, \dots, 1]$, $\nu > 0$, $\delta \in (0, 1)$, and $\mathbf{w}(n)$ is a zero mean, unit variance, uncorrelated vector of Gaussian random variables. Finally, $\boldsymbol{\alpha}(0)$ is taken to be a zero mean Gaussian random vector with covariance matrix given by $\mathbf{B}^T \mathbf{B}$. Eq. (7) represents a random walk model for the vector $\boldsymbol{\alpha}(n)$. The matrix \mathbf{B} is such that variance of the additive noise is greatest for the higher order coefficients in (5) and decreases for the lower order α 's representing lower frequency components. The constant ν controls the overall power in the process and δ determines the relative variances among the coefficients.

To summarize, in the absence of a mine, the overall signal model we use for the remainder of the work is described by the following state-space model with $\boldsymbol{\alpha}(n)$ representing the state of the system:

$$\boldsymbol{\alpha}(n+1) = \boldsymbol{\alpha}(n) + \mathbf{B}\mathbf{w}(n) \quad (8)$$

$$\mathbf{d}(n) = \mathbf{C}\boldsymbol{\alpha}(n) + \mathbf{v}(n) \quad (9)$$

where $\mathbf{v}(n) \sim N(0, \gamma^2 \mathbf{I})$ is the additive white, sensor noise.

Sample paths of these clutter models are shown in Figs. 5(a) and 7(a) for an array consisting of 20 receiver elements for parameter choices described in greater depth in Section 5. In both cases, the variance of the $\mathbf{w}(n)$ is greater than that of $\mathbf{v}(n)$ so that the background “clutter to noise” ratio is assumed to be relatively large. These images indicate that our model of clutter possesses the desired features. As functions of y , they are quite smooth. As x (and hence n) increases, they display gradual changes except at a few points where the random walk model results in a sharp alteration of the “shape” of the profile.

In closing this section, we wish to emphasize that the clutter model described above is neither the only nor even necessarily the best such model for the processing of real data. Rather the precise model based on the underlying polynomial regression is intended more as a vehicle for demonstrating the efficacy of the detection algorithm described and tested in subsequent sections. Indeed, our approach towards detection is based only on the reasonable and generally unrestrictive assumption that one can describe the clutter using *some* state space model to capture the temporal correlation of the process. Any valid state space model will work and the problem of fitting field data to such a modeling structure represents an interesting and very worthwhile piece of future research.

4. DETECTION ALGORITHM

In this section, we develop our approach to the detection of buried mines from progressively obtained array EMI data. We begin by solving a simpler problem in which we assume that the mine can exist only at a known point in space and want only to determine whether the mine is present. The solution to this problem then allows for the solution to the problem where we lift the assumption of known location.

4.1. Case I: Known mine location

Assuming that the mine may only be located at position \mathbf{r}_0 , the first problem of interest is to determine whether the data we observe as the array proceeds down the track is just clutter or is comprised of clutter plus mine. From a statistical signal processing perspective, this amounts to a binary hypothesis test. Under the null hypothesis, H_0 , the data $\mathbf{d}(n)$ is comprised only of clutter and is therefore modeled using (8) and (9). Under the alternate hypothesis, the data are described by combining (8) and a modified observation equation:

$$\mathbf{d}(n) = \mathbf{s}_0(n) + \mathbf{C}\boldsymbol{\alpha}(n) + \mathbf{v}(n) \quad (10)$$

where $\mathbf{s}_0(n)$ is the signal over the array at time n arising from the mine at position \mathbf{r}_0 which can be computed using the model of Section 2.

A key difficulty in solving this binary decision problem is that the clutter is correlated. To overcome this problem, we employ a Kalman filter to “whiten” the data. In most cases, the Kalman filter is used to track the state of a stochastically evolving system. It is well known however that a bi-product of the minimum mean square error property of the filter is that it may be interpreted as a means of de-correlating a non-white noise sequence. Specifically, the Kalman filter takes as input the sequence $\mathbf{d}(n)$ and produces as output a new sequence $\mathbf{q}(n)$ termed the innovations. When there is no mine present, $\mathbf{q}(n)$ is a zero mean, unit variance, white Gaussian noise sequence. In the event that a mine is present, $\mathbf{q}(n)$ is again unit variance, white, and Gaussian, but now there is an added mean vector, $\boldsymbol{\rho}(n)$ (also known as a signature), which can be computed directly from $\mathbf{s}(n)$ and the matrices in the state space model for the noise.² Mathematically we have:

$$\text{Under } H_0 \quad \mathbf{q}(n) \sim N(0, \mathbf{I}) \quad (11)$$

$$\text{Under } H_1 \quad \mathbf{q}(n) \sim N(\boldsymbol{\rho}(n), \mathbf{I}) \quad (12)$$

That is, the detection problem now is one of determining whether or not there exists a known mean in a sequence of independent, unit covariance Gaussian random vectors.

In the case where the two hypotheses are to be distinguished based on a batch of data obtained between $n = 0$ and $n = N$, the problem stated in the previous paragraph has a well known solution given in the form of a likelihood ratio test. First, the following likelihood statistic is computed

$$l(N) = \sum_{j=0}^N \boldsymbol{\rho}^T(j)\mathbf{q}(j) - 0.5\boldsymbol{\rho}^T(j)\boldsymbol{\rho}(j). \quad (13)$$

Next, we compare $l(N)$ with a threshold $\tau(N)$. If $l(N) > \tau(N)$ we say that the mine is present otherwise, we continue to collect data. Typically, $\tau(N)$ is chosen such that the probability of false alarm P_{fa} (that is the probability of rejecting H_0 when H_0 is in fact true) is less than or equal to a predefined threshold.³ Under our Gaussian assumptions on the innovations, this threshold is³

$$\tau(N) = d(N) \operatorname{erfc}_*^{-1}(P_{fa}) - d^2(N)/2 \quad (14)$$

with

$$\operatorname{erfc}_*(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-x^2/2} dx \quad \text{and} \quad d^2(N) = \sum_{j=0}^N \boldsymbol{\rho}^T(j) \boldsymbol{\rho}(j)$$

For the demining application, we want the ability to declare the presence of a mine before the array actually passes the object. That is, we want to solve a sequential detection problem wherein at each n we decide whether to declare a mine, declare no mine, or move the array and collect more data. While well known methods exist for optimally solving this problem when $\boldsymbol{\rho}$ is independent of n , the solution to the problem for a time varying mean is less well studied. The algorithm in this paper represents a first attempt to solve this problem for the demining application.

Since we know where the mine is located, we can, in principle, take data up to but not including the time the sensor passes the mine. Thus, at each n we compute $l(n)$ as in (13) and for a fixed P_{fa} , perform the comparison to $\tau(n)$. We do not stop taking data until that point, n_0 , when the x coordinate of the array is such that at step $n_0 + 1$ the array would pass the mine. At that time we look at the sequence of decisions made for all previous n . If H_1 has been accepted more than Q times (in this paper $Q = 3$, that hypothesis is declared to be true.

We observe that we could have chosen the simpler approach of collecting data until time n_0 and then solving the fixed size binary hypothesis test. The methods described in the previous paragraph though are more easily generalized and provide greater insight into the more complex problem where we cannot wait until n_0 but must make the mine/no mine decision as rapidly as possible.

4.2. Case II: Unknown mine location

In the more realistic event that we know neither whether a mine is in a region of interest nor where it may be, we employ a “multiple model”² approach to the detection problem. As illustrated in Fig. 4, we hypothesize the existence of mines over a coarse grid extending in front of the moving array. Associated with the i th such mine location is a separate signature sequence $\boldsymbol{\rho}_i(n)$. At each n , we perform a separate binary likelihood ratio test for each location in the grid. When the array is set to pass those possible mine locations with a given x coordinate, we look to see if any of the hypotheses in that collection have been chosen more than M times. If none satisfies this criteria, we continue moving the array. If one location passes, we declare a mine. If more than one has been passed more than M times, we take as the actual mine detection that hypotheses which has exceeded its threshold by the widest margin at any time during the scan. Essentially, we take the most likely target if there are multiple possibilities.

The implementation of this procedure can be made quite efficient. Rather than keeping track of a large number of mine locations extending over the entire track of the array, we perform this procedure using a sliding window of hypothesized target positions. Let Δx be the distance the array moves from n to $n + 1$. We choose as the coarse grid of possible mine locations the points at the center of the squares on a grid of size $M\Delta x$. That is, it takes the array M steps to proceed through one column of the coarse scale grid. As shown in Fig. 4, assuming the array has been moving for a sufficiently long time so that the Kalman filter is operating in the steady state, there is a certain shift invariance to the detection problem. In particular, the mathematical forms of the hypotheses generated when the array is in column 1 and the mines in column

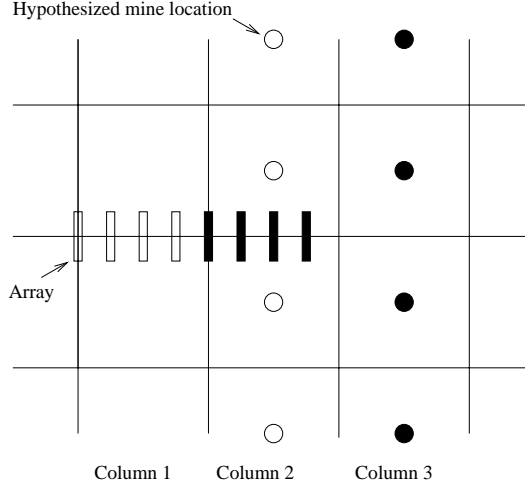


Figure 4. Relationship of moving array to coarse grid of possible mine locations. Rectangles represent positions of the array as it moves down the track. Circles represent hypothesized mine locations. There are $M = 4$ array positions per coarse grid cell.

2 (represented by the open boxes and circles) are identical to those obtained when the array is in column 2 and the mines are postulated to lie in column 3 (the darkened rectangles and circles). Clearly this holds for any array/mine combinations with identical relative spacings in the along-track direction.

We use this observation in the following manner. We begin by postulating mines over a $P \times P$ coarse grid in front of the array where P is chosen such that the response of the array when it is $PM\Delta x$ meters away is negligible. This grid (i.e. these hypotheses) are retained for the first M steps of the array. At the $M + 1$ st step, when the array is entering the second column, the grid looks very similar as in step 1 except the columns need to be renumbered. The first column, of the grid is removed since the array is leaving it. Each of the remaining columns are shifted to the left by one and a new column of hypotheses is added to the rightmost edge of the grid so that the overall array is still $P \times P$ only now looking further down track than was the case before. This procedure is repeated every M moves of the array.

This sliding window testing process is advantageous because it requires the storage of a fixed and relatively small number of signature vectors. Rather than having to maintain a $\rho(n)$ sequence for every conceivable combination of array locations and mine positions, we need only compute and store P^2M such sequences. For the EMI problems of interest here, excellent results are obtained in Section 5 for $P = 10$ and $M = 4$. The computation, storage, and manipulation of only 400 $\rho(n)$ sequences is relatively easy making this processing methodology quite amenable to real time implementation.

5. EXAMPLES

In this section, the performance of the detection approach is demonstrated and analyzed under a variety of noise conditions and mine placements. In all cases, the array is composed of a single transmitter located at $x = 0$ and 20 receivers with half equally spaced between $y = 5$ cm and $y = 50$ cm and the other half spaced between $y = -5$ cm and $y = -50$ cm. Thus, the overall length of the array is 1 meter. The transmitters and receivers are modeled as small circular loops of radius 2 cm located 10 cm above the air-earth interface and operating at 3 kHz. The array starts at $x = 0$ meters and proceeds in 75 equally spaced steps to $x = 3.7$ meters. The rolling grid of coarse scale hypothesized mine locations is composed of a 10 by 10

array of square cells stretching from about -90 cm to 90 cm in the y direction and extending about 2 meters in front of the array in the x direction. Finally, the mine-like target used in these simulations is modeled as a $5 \times 5 \times 5 \text{ cm}^3$ parallelepiped of contrast 1 S/m above the surrounding soil. Further, it is assumed that the burial depth of the mine is known *a priori*. Thus, the signature vectors are generated using this information. In practice, this assumption is not overly restrictive as range of burial depths is generally quite limited for a given demining application. Our scenario is meant to model the case of a shallowly buried anti-personnel mine, of interest for humanitarian demining.

For the low-frequency, Born-based inductive model discussed in Sec. 2, the electrical conductivity of the soil is approximated as zero. Moreover, the frequency and conductivity contrast of the mine enter the model only as multiplicative constants scaling the amplitude of the data. At higher frequencies and when an exact scattering model is employed, these various parameters will indeed play a larger role on the performance of the algorithm. For the work in this paper, however the key determination of performance is the relative strengths of the noise, clutter and target signatures. The measure used here is the signal to clutter plus noise ratio (SCNR) taken to be 20 times the base 10 logarithm of the ratio of the signal power to the noise plus clutter power taken over the entire data set obtained from $x = 0$ to $x = 3.7\text{m}$.

As a first example, we consider the detection of a mine located at $(x, y, z) = (2.2, -0.22, -0.05)$ m under a variety of SCNR. In Fig. 5(a) we display a sample data set. The mine, located in the image at $(x, y) = (2.2, -0.22)$ m, is barely discernible at this SCNR. In Fig. 5(b) the detection performance as a function of SCNR is shown. Each point represents the fraction of times the mine was detected in 100 runs of the algorithm using independent realizations of the clutter and noise processes for each run. The SCNR was changed by setting the variance of $\mathbf{v}(n)$ in (9) to 0.01 and increasing the variance of $\mathbf{w}(n)$ in (7) from 0.01 to 0.25 in steps of 0.02. It is clear, that strong performance is seen down to a SCNR of about -20 dB with a gradual decline in detection as the SCNR further degrades. At an SCNR of -15 dB the mine is detected essentially 100% of the time. Finally, since the mine was not located at one of the coarse scale grid points, a detection was declared if the estimated mine location was one of the four neighbors nearest the true mine location on the hypothesis grid.

In Fig. 6, we display the average value of the likelihood statistic less the threshold (i.e. $l - \tau$ from (13) and (14)) as a function of the position of the sensing array as the array proceeds toward the mine. The average is taken over all noise realizations and all experiments where the mine was in fact detected according to the criterion described previously. The SCNR for this problem was about 2.6 dB and the x coordinate of the mine is again 2.2 m. The plot does not reach 2.2 m because of the coarse scale gridding of the mine hypotheses, which for this problem tests for mines located at $x = 2.1$ m and $x = 2.3$ m. Fig. 6 clearly shows that the relevant decision statistic rises to a significant level above the detection threshold of zero well before the array actually passes over the buried object. In particular, these statistics become substantial when the array is located at $x = 1.95$ m which is over 25 cm from the location of the buried mine.

As a second example, we consider the problem of detecting a mine at $(x, y, z) = (2.2, -0.7, -0.05)$. This placement of the mine is interesting because, unlike the first case, the mine is located outside of the direct field of view of the EMI array. In particular, the array extends to $y = -50$ cm; however the mine is located at $y = -70$ cm. A sample data set at an SCNR of about 1.8 dB is shown in Fig. 7(a) and the detection results are given in Fig. 7(b). As in the first problem, the signature of the mine, which should be evident at around $x = 2.2$ m and $y \in (-.3, -.5)$ m, is obscured by the clutter; however the detection rate is quite high even to an SCNR of about -20dB. In Fig. 8, the plot of average likelihood statistics less threshold are shown for this problem at an SCNR of about 1. These plots show that even when the mine is located significantly off the axis of the array, we still retain a strong degree of forward looking, predictive capability. In particular, non-negligible test statistics are reliably observed for this scenario when the array is located at $x = 1.9$ m which is about 30 cm from the mine.

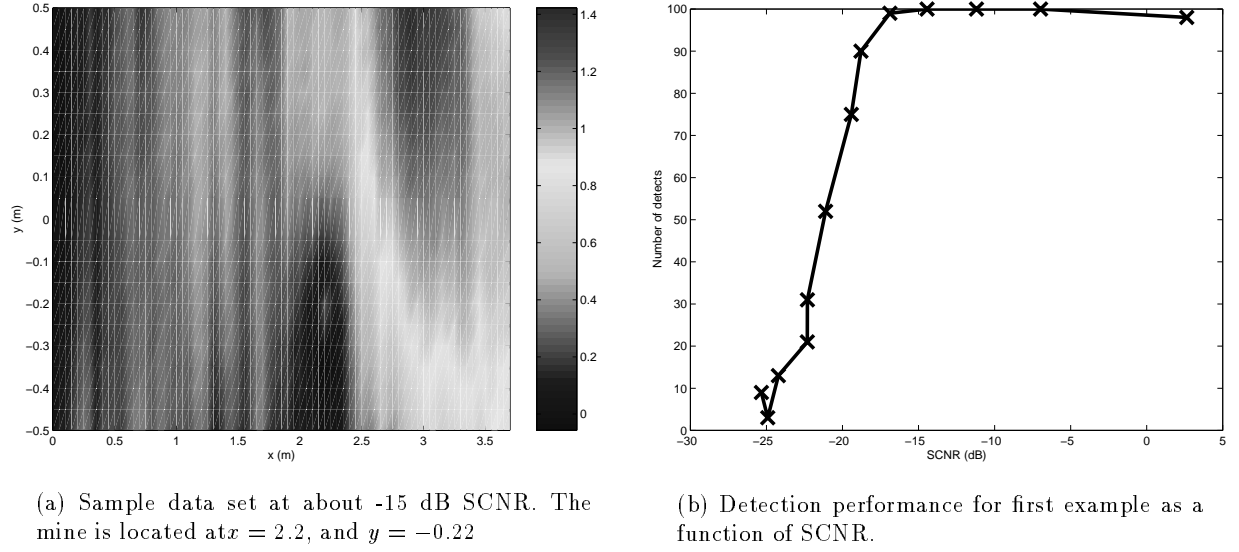


Figure 5. Sample data set and detection results for first example

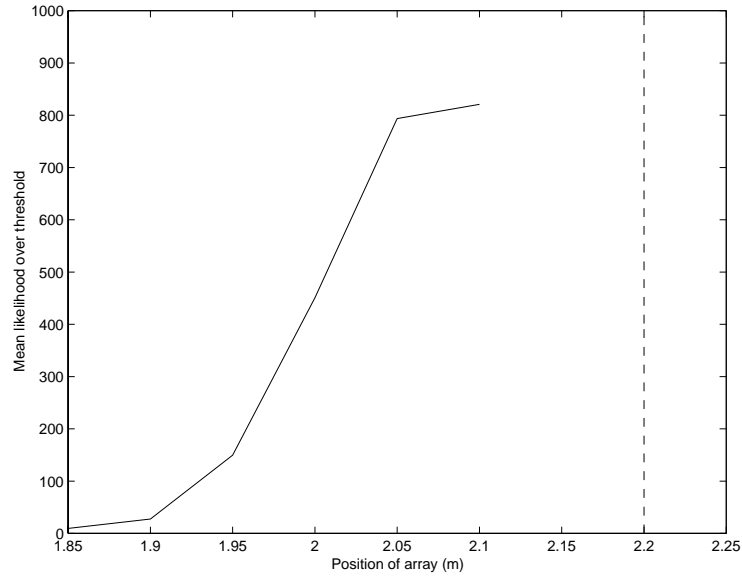


Figure 6. Average value of likelihood statistic in excess of threshold as array approaches mine location for first problem. Detection requires this value to exceed zero. The dashed line indicates the x coordinate of the mine.

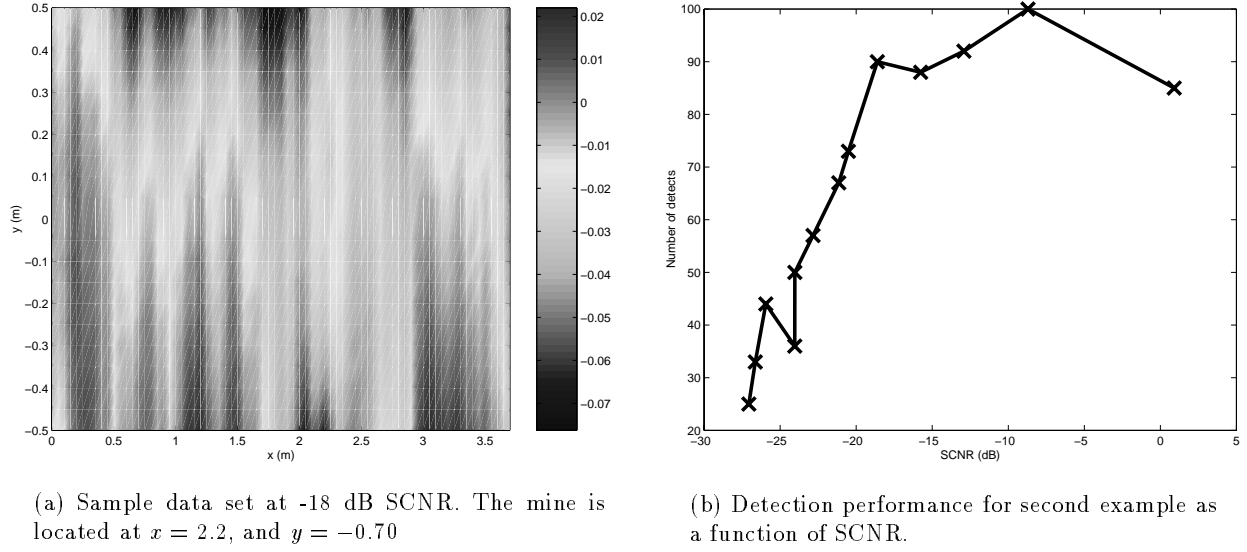


Figure 7. Sample data set and detection results for second example

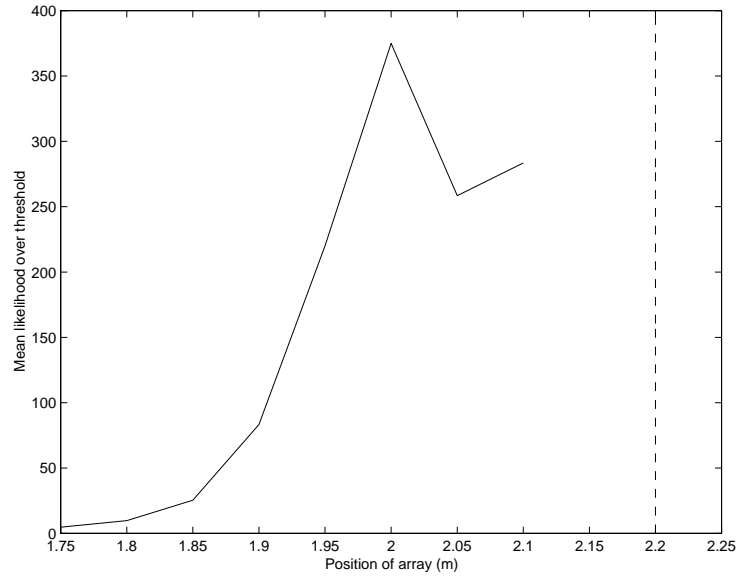


Figure 8. Average value of likelihood statistic in excess of threshold as array approaches mine location for second problem. Detection requires this value to exceed zero. The dashed line indicates the x coordinate of the mine.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented an approach for the detection and localization of buried landmines from EMI sensor data. Our statistical signal processing approach is tailored both to the sensor (an array of EMI transducers) and the data collection method (progressive acquisition as the array proceeds down track). The approach we have taken casts the detection problem in the form of a statistical hypothesis test. Using a novel state-space model for the clutter, we employ a Kalman filter to whiten the data and use a rolling window, multiple-model approach to ascertain the location of a mine. Preliminary simulations indicate that this approach can provide significant forward and side looking capability thereby increasing the area over which an EMI sensor can reliably detect buried mines.

While the work in this paper has been directed towards an EMI sensing device, the general detection methodology can be extended and applied to *any* scheme in which an array of sensors collects data in a progressive manner. All that is required is a forward model to generate the rather modest sized library of target signatures and a state-space or other recursive model for the clutter. With this in mind, future research efforts stemming from this project include the following:

1. Development of an estimation-based methods to better localize the detected object. Here we are interested in developing a non-linear least squares method which takes the coarse grid estimates generated by the algorithm in this paper and provides a more accurate determination of the mine's true position.
2. Development of methods for “discounting” detected mines from the data set. After the mine has been detected, we want to remove its effects from the data thereby allowing the array to continue down track looking for other buried objects. Failure to perform such discounting generally degrades subsequent performance.
3. Development of a full EMI scattering model for use in generation of target signatures.
4. Identification of state space type model for EMI array from real field data.
5. Extension and adaptation of these methods to ground penetrating radar (GPR) sensor.
6. Use of Kalman filtering ideas for fusion of GPR and EMI sensor data.

REFERENCES

1. W. C. Chew and R. J. Kleinberg, “Theory of microinduction measurements,” *IEEE Trans. Geoscience and Remote Sensing* **26**, pp. 707–719, Nov. 1988.
2. M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice Hall, 1993.
3. H. L. Van Trees, *Detection, Estimation and Modulation Theory: Part I*, John Wiley, New York, 1968.

Detection and Localization of Obscured Objects from Near-Field Sensor Array Data: Physical Models and Statistical Processing*

Eric L. Miller
235 Forsyth Building
Northeastern University
Boston, MA 02115
Tel: (617) 373-8386
email: elmiller@ece.neu.edu

W. Clem Karl
Dept. Elect. and Comp. Engineering
Dept. Biomedical Engineering
Boston University
8 St. Mary's St.
Boston, MA 02215
Tel: (617) 353-9788
email: wckarl@bu.edu

Abstract

The problem of near field object detection from array-based electromagnetic induction data is addressed. A forward scattering model based on the Born approximation is employed. Using insight obtained from this model, a clutter model in the form of a state space system is developed which describes the correlation of the noise both across the sensing array and from one position of the array to the next as the measurement device proceeds down track. A multiple-model detection scheme based on the whitening properties of the Kalman filter is employed to perform the actual object detection. This approach is applied to the detection and localization of buried objects. Examples are provided for land mines buried directly in front and off to one side of the array.

1 Introduction

The problem of detecting and localizing highly obscured objects from indirect observations of scattered electromagnetic (EM) radiation has received considerable attention in recent years. In environmental remediation, one is often wants to determine the existence and positions of metallic drums filled with hazardous waste materials. There has been a great deal of attention paid to the problem of finding and removing buried unexploded ordinance using EM sensing technologies. On a smaller scale, the non-destructive testing problem of flaw detection using eddy current sensors possesses essentially the same structure (and physics) as the waste and mine removal applications.

From a signal processing perspective, the physics of these problems represents an unusual challenge to de-

tection and localization tasks. The vast majority of array processing algorithms are based on the assumption that the observed waves are propagating in free space and the unknown sources are in the far field. The resulting planewave model for the radiation incident on the sensors serves as the basis for all beamforming and high-resolution localization methods. For the applications of interest here, the objects are in the nearfield of the array and while the sensors are in freespace, the targets are located in a lossy halfspace. The typical array approaches are therefore not applicable. The problem of information extraction from such data then requires careful modeling of the physics coupled with judicious approximations required for computational efficiency. In this paper we will focus on the problem of detecting buried objects from such array data.

The data for these problems are typically obtained by sweeping an array of transducers in a fixed direction over a region of interest, as illustrated in Figure 1. The result is an “image” whose number of rows is equal to the number of sensors in the array and whose number of columns is determined by the number of stops the array makes. This implied temporal structure is the second unique “image processing” element present in this problem. In applications such as buried landmine detection it is imperative to detect the object (i.e. the mine) *before* the sensor array actually passes over the its location. More generally, there is often a desire in these problem to process the data in a “real time” manner (i.e. as the data is being acquired). This requirement implies that any algorithm must process the image data in a column-by-column manner.

The final challenging characteristic of this problem is the need to process data contaminated by “clutter.” Physically, such clutter can arise from a variety

*This work was supported by the Army Research Office Demining MURI under Grant DAAG55-97-1-0013 and the Air Force Office of Scientific Research under Grant F49620-96-1-0028

of inhomogeneities and man-made non-target discrete structures present in the observed medium as well as the continuous, natural changes in its electromagnetic characteristics. From a signal processing perspective, we regard clutter as a correlated noise source present in the data. The correlation is evident first along the array from one sensor to the next and second along the collection track from one array position to the next.

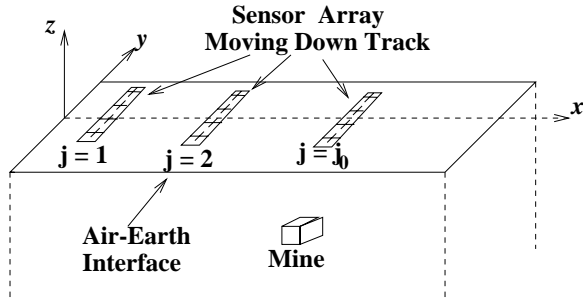


Figure 1: Data Acquisition Method for Subsurface Detection Problem

We consider statistical signal processing methods for solving the object detection problem outlined in the previous paragraphs. In §2, we describe a physical model relating the characteristics of the buried object to the observations obtained over an array. We next use this physical model as the basis for a simple, analytical, stochastic model describing the distribution of clutter. This model takes the form of a state-space system accounting both for the inter-element and inter-array correlation of the noise. In §4, we take advantage of the fact that the Kalman filter can be used to decorrelate (or whiten) such noise processes to obtain a method for detecting objects. We show in the examples in §5 that this process is capable of detecting shallowly buried mine-like objects reliably (a) up to 20cm before the array actually passes over the mine and (b) under low clutter conditions, when the mine is located up to 30 cm from one side or another of the array. Finally, conclusions are described in §6.

2 Forward Model

We assume that the array is situated in the $x - y$ plane and proceeds from $x = 0$ to $x = x_{max}$ stopping at N_s equally spaced locations to obtain data. The array is composed of a single transmitter at $y = 0$ and N_r receivers equally spaced on either side of the array. All transducers are taken to be small current loops of radii r_p located z_p above the air-earth interface carrying I A of current with an $e^{-i\omega t}$ time dependence. We define $d_{m,n}$ to be the field measured by the m th element of the array at the n th stop between $x = 0$ and

$x = x_{max}$. In this paper, we assume that the object represents a small perturbation of constant conductivity σ S/m in the electromagnetic properties of the earth so that the first Born approximation [2] can be used to obtain the following, tractable physical model

$$d_{nm} = -\frac{\sigma(\omega\mu I r_p)^2}{4} \times \int_V A(\mathbf{r} - \mathbf{r}_0 + \mathbf{r}_n) A(\mathbf{r} - \mathbf{r}_0 + \mathbf{r}_m) d\mathbf{r} \quad (1)$$

where the quantity A represents the Green's function for the problem of a loop over a half-space [2], $\mathbf{r} = [x \ y \ z]^T$ is a point in three space represented in either Cartesian or polar coordinates, \mathbf{r}_0 is the position of the center of the object, \mathbf{r}_n (\mathbf{r}_m) is the position of the transmitter (receiver), μ is the magnetic permeability and the integral is taken over the volume, V , of the prototypical object centered at the origin. Given this model, the goals of the processing in this paper are the detection and localization of the presence of an object from noisy observations of d_{ij} .

3 Clutter Model

The motivation for the mathematical form of the clutter model is provided by the nature of the signal observed over the array due only to an object buried some distance away. For the problems of interest here, the wavefield over the aperture of the array is rather smooth when an object does in fact exist near the array and can be well modeled by a low order polynomial in y_m , the y position of the m th sensor in the array [3]. In other words we can write

$$d_{m,n} = \sum_{j=0}^J \alpha_{j,n} y_m^j. \quad (2)$$

By defining the vectors $\mathbf{d}^T(n) = [d_{1,n} \ d_{2,n} \ \dots \ d_{N_r,n}]$ and $\boldsymbol{\alpha}^T(n) = [\alpha_{0,n} \ \alpha_{1,n} \ \dots \ \alpha_{J,n}]$, (2) is rewritten as

$$\mathbf{d}(n) = \mathbf{C}\boldsymbol{\alpha}(n). \quad (3)$$

where the (m, j) th element of \mathbf{C} is y_m^j .

The clutter model we employ in this work is based on the assumption that the random signal over the array in the absence of an object is the sum of two components. The first portion is an additive zero mean, white Gaussian component of variance γ^2 representing sensor noise. The second component is intended to capture the effects due to the statistical distribution of the environmental factors not of interest in the detection problem. Here we assume that this signal can also be described by a model of the form (3). In the case of an object, the coefficients of the polynomial

are deterministically related to the position of the object relative to the array. For clutter however, these coefficients are taken to be random quantities. That is, $\alpha(n)$ is a random vector. Such a model arises if we assume that the clutter represents the average effect on the sensor array of a random distribution of small scatterers.

To account for the fact that the clutter seen at position n of the array should be correlated to that at position $n + 1$, we employ the following model of the spatial evolution of $\alpha(n)$:

$$\alpha(n+1) = \alpha(n) + \mathbf{B}(n)\mathbf{w}(n) \quad (4)$$

where $\mathbf{B} = \nu \text{diag} [\delta^{J/2}, \delta^{(J-1)/2}, \dots, 1]$, $\nu > 0$, $\delta \in (0, 1)$, and $\mathbf{w}(n)$ is a zero mean, unit variance, uncorrelated vector of Gaussian random variables. Finally, $\alpha(0)$ is taken to be a zero mean Gaussian random vector with covariance matrix given by $\mathbf{B}^T \mathbf{B}$. Eq. (4) represents a random walk model for the vector $\alpha(n)$. The matrix \mathbf{B} is such that variance of the additive noise is greatest for the higher order coefficients in (2) and decreases for the lower order α 's representing lower frequency components. The constant ν controls the overall power in the process and δ determines the relative variances among the coefficients.

To summarize, in the absence of an object, the overall signal model we use for the remainder of the work is described by the following state-space model with $\alpha(n)$ representing the state of the system:

$$\alpha(n+1) = \alpha(n) + \mathbf{B}\mathbf{w}(n) \quad (5)$$

$$\mathbf{d}(n) = \mathbf{C}\alpha(n) + \mathbf{v}(n) \quad (6)$$

where $\mathbf{v}(n) \sim N(0, \gamma^2 \mathbf{I})$ is the white sensor noise.

4 Detection algorithm

We begin by assuming that the object may only be located at position \mathbf{r}_0 so that the first problem of interest is to determine whether the data we observe as the array proceeds down the track are just clutter or are comprised of clutter plus object. From a statistical signal processing perspective, this amounts to a binary hypothesis test. Under the null hypothesis, H_0 , the data $\mathbf{d}(n)$ are comprised only of clutter and is therefore modeled using (5) and (6). Under the hypothesis that a mine is present, the data are described by:

$$\alpha(n+1) = \alpha(n) + \mathbf{B}\mathbf{w}(n) \quad (7)$$

$$\mathbf{d}(n) = \mathbf{s}_0(n) + \mathbf{C}\alpha(n) + \mathbf{v}(n) \quad (8)$$

where $\mathbf{s}_0(n)$ is the signal over the array at time n arising from the object at position \mathbf{r}_0 which can be computed using the model of §2.

A key difficulty in solving this binary decision problem is that the clutter is correlated. To overcome this problem, we employ a Kalman filter to “whiten” the data. Specifically, the Kalman filter takes as input the sequence $\mathbf{d}(n)$ and produces as output a new sequence $\mathbf{q}(n)$ termed the innovations [1]. When there is no object present, $\mathbf{q}(n)$ is a zero mean, unit variance, white Gaussian noise sequence. In the event that an object is present, $\mathbf{q}(n)$ is again unit variance, white, and Gaussian, but now there is an added mean vector, $\rho(n)$ (also known as a signature), which can be computed directly from $\mathbf{s}(n)$ and the matrices in the state space model for the noise [1]. Thus, the detection problem now is one of determining whether or not there exists a known mean in a sequence of independent, unit covariance Gaussian random vectors.

In the case where the two hypotheses are to be distinguished based on a batch of data obtained between $n = 0$ and $n = N$, the problem stated in the previous paragraph has a well known solution given in the form of a likelihood ratio test. First, the following likelihood statistic is computed

$$l(N) = \sum_{j=0}^N \rho^T(j) \mathbf{q}(j) - 0.5 \rho^T(j) \rho(j). \quad (9)$$

Next, we compare $l(N)$ with a threshold $\tau(N)$. If $l(N) > \tau(N)$ we say that the object is present otherwise, we continue to collect data. Typically, $\tau(N)$ is chosen such that the probability of false alarm P_{fa} is less than or equal to a predefined threshold [4]. For the application here, we want the ability to declare the presence of an object before the array actually passes the object. That is, we want to solve a sequential detection problem wherein at each n we decide whether to declare an object, declare no object, or move the array and collect more data. While well known methods exist for optimally solving this problem when ρ is independent of n , the solution to the problem for a time varying mean is less well studied. The algorithm here represents a first attempt to solve this problem for our object detection application.

Since we know the object location, we can, in principle, take data up to but not including the time the sensor passes the object. Thus, at each n we compute $l(n)$ as in (9) and for a fixed P_{fa} , perform the comparison to $\tau(n)$. We do not stop taking data until that point, n_0 , when the x coordinate of the array is such that at step $n_0 + 1$ the array would pass the object. At that time we look at the sequence of decisions made for $n \leq n_0$. If H_1 has been accepted more than Q times (here $Q = 3$), that hypothesis is declared to be true.

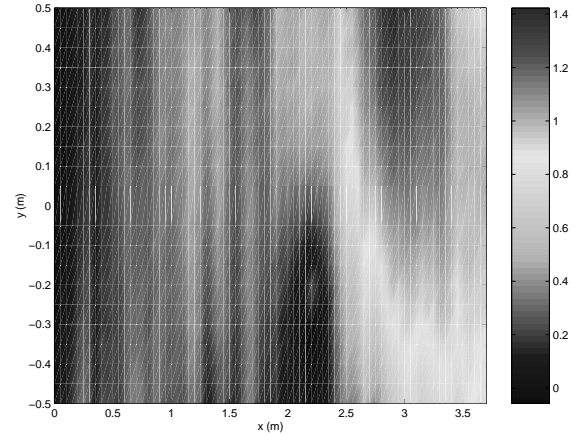
In the more realistic event that we know neither whether an object is in a region of interest nor where it may be, we employ a “multiple model” [1] approach to the detection problem. We hypothesize the existence of objects over a coarse grid extending in front of the moving array. Associated with the i th such object location is a separate signature sequence $\rho_i(n)$. At each n , we perform a separate binary likelihood ratio test for each location in the grid. When the array is set to pass those possible object locations with a given x coordinate, we look to see if any of the hypotheses in that collection have been chosen more than M times. If none satisfies this criteria, we continue moving the array. If one location passes, we declare an object. If more than one has been passed more than M times, we take as the actual object detection that hypotheses which has exceeded its threshold by the widest margin at any time during the scan. Essentially, we take the most likely target if there are multiple possibilities.

As we discuss in the talk accompanying this paper, this sliding window testing process is advantageous because it requires the storage of a fixed and relatively small number of signature vectors. In particular, excellent results are obtained in §5 for using a target signature library of only 400 $\rho(n)$ sequences the computation, storage, and manipulation of which is relatively easy making this processing methodology quite amenable to real time implementation.

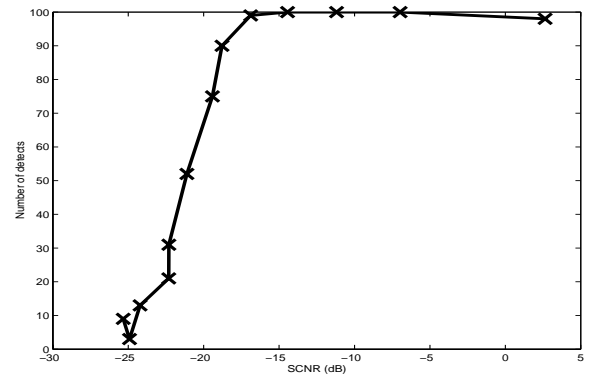
5 Examples

For the following examples, the array is composed of a single transmitter located at $x = 0$ and 20 receivers with half equally spaced between $y = 5$ cm and $y = 50$ cm and the other half spaced between $y = -5$ cm and $y = -50$ cm. The array starts at $x = 0$ m and proceeds in 75 equally spaced steps to $x = 3.7$ m. The target is modeled as a $5 \times 5 \times 5$ cm³ parallelepiped of contrast 1 S/m above the surrounding soil and it is assumed that the burial depth of the object is known *a priori*. Thus, the signature vectors are generated using this information. In practice, this assumption is not overly restrictive as the range of burial depths is generally quite limited for a given demining application. Our scenario is meant to model the case of a shallowly buried anti-personnel mine, of interest for humanitarian demining. The noise measure used here is the signal to clutter plus noise ratio (SCNR) taken to be 20 times the base 10 logarithm of the ratio of the signal power to the noise plus clutter power taken over the entire data set.

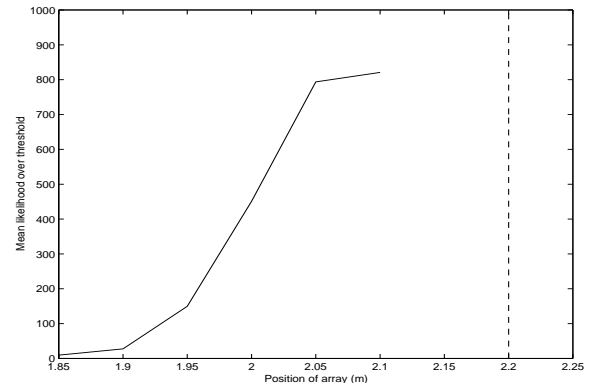
We first consider the detection of an object located at $(x, y, z) = (2.2, -0.22, -0.05)$ m under a variety of SCNR. In Fig. 2(a) we display a sample data set. The



(a) Sample data set at about -15 dB SCNR. The object is located at $x = 2.2$, and $y = -0.22$



(b) Detection performance for first example.



(c) Average value of likelihood statistic in excess of threshold. The dashed line indicates the x coordinate of the object.

Figure 2: Results for first example

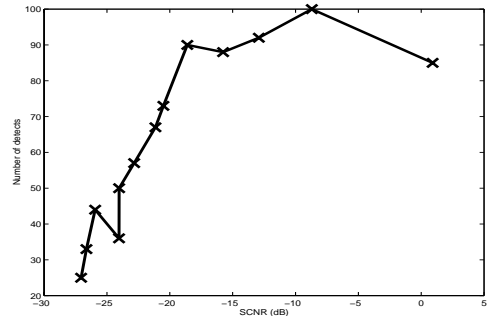
object, located in the image at $(x, y) = (2.2, -0.22)$ m, is barely discernible at this SCNR. In Fig. 2(b) the detection performance as a function of SCNR is shown. Each point represents the fraction of times the object was detected in 100 runs of the algorithm using independent realizations of the clutter and noise processes for each run. Strong performance is seen down to a SCNR of about -20 dB with a gradual decline in detection as the SCNR further degrades. At -15 dB the object is detected essentially 100% of the time.

In Fig. 2(c), we display the average value of the likelihood statistic less the threshold (i.e. $l - \tau$) as a function of the position of the sensing array as the array proceeds toward the object. The average is taken over all noise realizations and all experiments where the object was in fact detected. The SCNR for this problem was about 2.6 dB and the x coordinate of the object is 2.2 m. Fig. 2(c) clearly shows that the relevant decision statistic rises to a significant level above the detection threshold of zero well before the array actually passes over the buried object. In particular, these statistics become substantial when the array is located at $x = 1.95$ m which is over 25 cm from the location of the buried object.

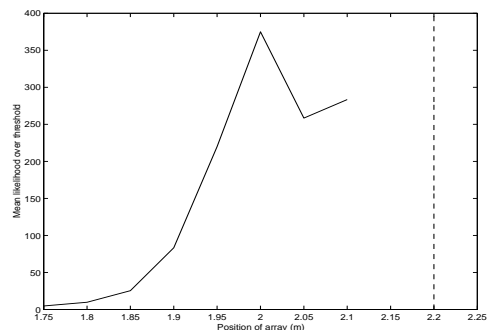
As a second example, we consider detecting an object at $(x, y, z) = (2.2, -0.7, -0.05)$. Unlike the first case, the object is located outside of the direct field of view of the array. The detection results are given in Fig. 3(a) while in Fig. 3(b), the plot of average likelihood statistics less threshold are shown for at SCNR of about 1. These plots show that even when the object is located significantly off the axis of the array, we still retain a strong degree of forward looking, predictive capability. In particular, non-negligible test statistics are reliably observed when the array is located about 30 cm from the object.

6 Conclusions

In this paper, we have presented an approach for the detection and localization of objects from near field sensor data. Our statistical signal processing approach is tailored both to the sensor and the data collection method. The approach we have taken casts the detection problem in the form of a statistical hypothesis test. Using a novel state-space model for the clutter, we employ a Kalman filter to whiten the data and use a rolling window, multiple-model approach to ascertain the location of a object. Preliminary simulations indicate that this approach can provide significant forward and side looking capability thereby increasing the area over which an EMI sensor can reliably detect buried objects. In the future, we look to extend these processing methods to other detection



(a) Detection results for second example



(b) Average value of likelihood statistic in excess of threshold. The dashed line indicates the x coordinate of the object.

Figure 3: Results for second example

problems; specifically those arising in remote surveillance where target signatures have been reduced due to excessive clutter, foliage, or other obscuration.

References

- [1] Michele Basseville and Igor Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, 1993.
- [2] Weng Cho Chew and Robert J. Kleinberg. Theory of microinduction measurements. *IEEE Trans. Geoscience and Remote Sensing*, 26(6):707–719, Nov. 1988.
- [3] Eric L. Miller, W. Clem Karl, and Stephen J. Norton. On the detection of buried mines from array inductive measurements. In *SPIE AeroSense Symposium, Detection Technologies for Mines and Minelike Targets*. SPIE, April 1998.
- [4] Harry L. Van Trees. *Detection, Estimation and Modulation Theory: Part I*. John Wiley, New York, 1968.

On the detection of buried mines from array inductive measurements

Eric L. Miller^a, W. Clem Karl^b, and Stephen J. Norton^c

^a Center for Electromagnetics Research, Northeastern University, Boston, MA.

^b Multidimensional Signal Processing Laboratory, Boston University, Boston, MA 02215

^c Oak Ridge National Laboratory, Oak Ridge, TN 37831-6200

ABSTRACT

The problem of mine detection and localization from array-based electromagnetic induction data is addressed. An efficient forward scattering model based on the Born approximation is employed. Using insight obtained from this model, a clutter model in the form of a state space system is developed which describes the correlation of the noise both across the sensing array and from one position of the array to the next as the measurement device proceeds down track. A multiple-model detection scheme based on the whitening properties of the Kalman filter is employed to perform the actual mine detection. This approach allows for the detection and localization of buried objects well before the array physically moves over mines. Examples are provided for mines buried directly in front and off to one side of the array.

Keywords: Mine detection, EMI arrays, Kalman filter

1. INTRODUCTION

The problem of detecting and localizing buried landmines from indirect observations of scattered electromagnetic radiation has received considerable attention in recent years. Of interest in this paper is the development of statistical signal processing methods for solving such problems based on data collected using an array of very low frequency electromagnetic induction sensors. From a signal processing perspective, the physics of this problem represents an unusual challenge. The majority of array processing algorithms are based on the assumption that the observed waves are propagating in free space and the unknown sources are in the far field. The resulting planewave model for the radiation incident on the sensors serves as the basis for all beamforming and high-resolution localization methods. For the applications of interest here, the objects are in the nearfield of the array and while the sensors are in freespace, the targets are located in a lossy halfspace. The typical array approaches are therefore not applicable.

In addition to the basic physics describing these problems, the data are typically obtained by sweeping the array in a fixed direction over a region of interest, as illustrated in Figure 1. The result is an “image” acquired over time whose number of rows is equal to the number of sensors in the array and whose number of columns is determined by the number of stops the array makes as it sweeps over the area. This implied temporal structure is the second unique image processing element present in this problem. In particular, it is imperative to detect the mine *before* the sensor array actually passes over its location. This requirement implies that any algorithm must process the image data in a real-time, column-by-column manner, continuously updating its decision as to mine presence and its estimate concerning the object’s location as new data are obtained by the scanning procedure.

Other author information: ELM: elmiller@ece.neu.edu WCK: wckarl@bu.edu. SJN nortonsj@ornl.gov. This work was supported by the Army Research Office Demining MURI under Grant DAAG55-97-1-0013 and the Air Force Office of Scientific Research under Grant F49620-96-1-0028

The final challenging characteristic of this problem is the need to process data contaminated by “clutter.” Physically, such clutter can arise from a variety of inhomogeneities present in the soil including a random distribution of rocks and other discrete structures as well as the continuous, natural changes in the electromagnetic characteristics of the earth. From a signal processing perspective, we regard clutter as a correlated noise source present in the data. The correlation is evident first along the array from one sensor to the next and second along the collection track from one array position to the next. To adequately detect mines whose signals are obscured by such noise requires both careful modeling of the clutter and the development of processing methods explicitly based on these models.

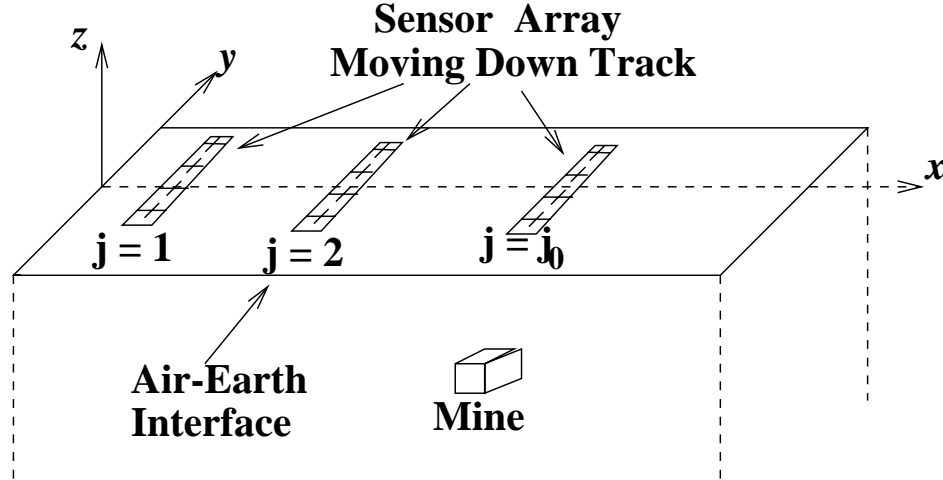


Figure 1. Data Acquisition Method for Subsurface Detection Problem

In this paper, we consider statistical signal processing methods for solving the mine detection problem outlined in the previous paragraphs. Our focus is on shallowly buried metallic anti-personal mines arising in the case of humanitarian demining. In Section 2, we describe a physical model relating the characteristics of the buried mine to the observations obtained over an array. We next use this physical model as the basis for a simple, analytical, stochastic model describing the distribution of clutter. In particular, the clutter model takes the form of a so-called “state-space” system. This system accounts both for the inter-element and inter-array correlation of the noise. In Section 4, we take advantage of the fact that the Kalman filter can be used to decorrelate (or whiten) such noise processes to obtain a method for detecting mines. At each stop of the array, we perform a series of hypothesis tests. Under the null hypothesis for each test, no mine is present. Under the alternative, there is a mine located at one of a number of positions over a coarse scale grid. The whitening properties significantly simplify the formulation of the tests while retaining the optimality property of the procedure. We stop the array and declare the presence of a mine when the sequence of tests for a mine at a hypothesized location are “sufficiently convincing.” We show in the examples in Section 5 that this process is capable of detecting mines reliably (a) up to 20cm before the array actually passes over the mine and (b) under low clutter conditions, when the mine is located up to 30 cm from one side or another of the array. Finally, conclusions and future work are described in Section 6.

2. FORWARD MODEL

To fix notation, we assume that the array is situated in the $x - y$ plane and proceeds from $x = 0$ to $x = x_{max}$ stopping at N_s equally spaced locations to obtain data. The array is composed of a single

transmitter at $y = 0$ and N_r receivers equally spaced on either side of the array. All transducers are taken to be small current loops of radii r_p located z_p above the air-earth interface carrying I A of current with an $e^{-i\omega t}$ time dependence. We define $d_{m,n}$ to be the field measured by the m th element of the array at the n th stop between $x = 0$ and $x = x_{max}$. In this paper, we assume that the mine represents a small perturbation of constant conductivity σ S/m in the electromagnetic properties of the earth so that the first Born approximation¹ can be used to obtain a tractable physical model. For the case of a single mine, it can be shown that the relationship between the physical properties of the mine and y_{mn} is given by¹

$$d_{nm} = -\frac{\sigma(\omega\mu I r_p)^2}{4} \int_V A(\mathbf{r} - \mathbf{r}_0 + \mathbf{r}_n) A(\mathbf{r} - \mathbf{r}_0 + \mathbf{r}_m) d\mathbf{r} \quad (1)$$

where the quantity A represents the Green's function for the problem of a loop over a half-space, $\mathbf{r} = [x \ y \ z]^T$ is a point in three space represented in either Cartesian or polar coordinates, \mathbf{r}_0 is the position of the center of the mine, \mathbf{r}_n (\mathbf{r}_m) is the position of the transmitter (receiver), μ is the magnetic permeability and the integral is taken over the volume, V , of the prototypical mine centered at the origin. In [1], it is shown that the Green's function A can be well approximated as

$$A(\mathbf{r}) = \frac{2}{\pi k(\mathbf{r}) \sqrt{r r_p}} \left[\left(1 - \frac{1}{2} k(\mathbf{r})^2 \right) K(k(\mathbf{r})) - E(k(\mathbf{r})) \right] \quad (2)$$

$$k(\mathbf{r}) = \frac{4 r r_p}{(z + z_p)^2 + (r + r_p)^2} \quad (3)$$

where $r = \|\mathbf{r}\|$ is the distance to the physical point, $K(x)$ is the complete elliptic integral of the first kind, $E(x)$ is the complete elliptic integral of the second kind and the parameter $k(\mathbf{r})$ is in the range 0 to 1.

For the problem of interest in this paper, we have found by direct evaluation that $k(\mathbf{r}) \in [0, 0.6]$. Under this restriction, we have developed the following highly accurate, yet simple approximation to (2):

$$A(\mathbf{r}) = \frac{2}{\pi k(\mathbf{r}) \sqrt{r r_p}} \left(\beta_4 k^4(\mathbf{r}) + \beta_6 k^6(\mathbf{r}) \right) \quad (4)$$

with $\beta_4 = 0.0882$, $\beta_6 = 0.1321$ and $k(\mathbf{r})$ as in (3). The particular values of the β parameters are obtained by a least squares fit and the use of the fourth and sixth powers of k is motivated by the Taylor series expansion for the right hand side of (2). A semilogarithmic comparison of the right hand sides of (2) and (4) is shown in Figure 2. The solid line represents the exact combination of elliptic integrals while the dashed line is the polynomial approximation. It is clear that the approximation is, for all intents and purposes, the same as the exact expression. From a computational perspective however, (4) offers definite advantages over (2) in terms of speed of evaluation and ease of manipulation.

Given this model, the goals of the processing in this paper are the detection of the presence of a mine from noisy observations of d_{ij} and determination of \mathbf{r}_0 once a detection has been declared.

3. CLUTTER MODEL

Typically, white Gaussian noise is used as the clutter model for mine detection problems due to its tractability. Unfortunately, this simplified assumption leads to a degradation in detection performance. In this work we instead explicitly model the clutter or background process, leading to more robust performance. In particular, we develop a tractable yet realistic stochastic model describing the distribution of the data obtained over our receiver in the absence of a target. This statistical model then forms the basis for a decision-theoretic procedure for the localization of buried mines discussed in the remainder of this paper.

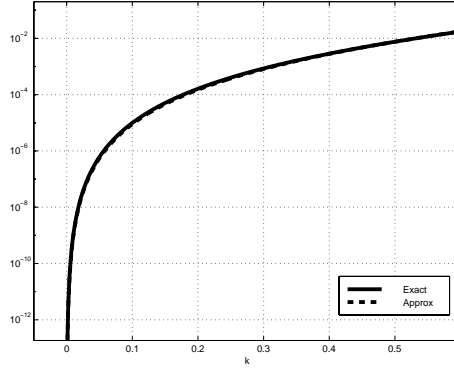


Figure 2. Comparison of exact and polynomial approximation to forward scattering model

The motivation for the mathematical form of our clutter model is actually provided by the nature of the signal observed over the array due only to a mine buried some distance away. The wavefield over the aperture of the array is rather smooth when a mine does in fact exist near the array. In Fig. 3, noiseless simulated data is shown for the case of a $.05 \times .05 \times .05$ m³ mine located at $\mathbf{r}_0 = \{x_0, y_0, z_0\} = \{2, .1, -.05\}$ m where an 8 element array of length .6m is positioned at $x \in \{0, .5, 1, 1.5\}$ m. These plots indicate the more general feature that the data over the array can be well modeled by a low order polynomial in y_i , the y position of the i th sensor in the array. In other words we can write

$$d_{m,n} = \sum_{j=0}^J \alpha_{j,n} y_m^j. \quad (5)$$

By defining the vectors $\mathbf{d}^T(n) = [d_{1,n} \ d_{2,n} \ \dots \ d_{N_r,n}]$ and $\boldsymbol{\alpha}^T(n) = [\alpha_{0,n} \ \alpha_{1,n} \ \dots \ \alpha_{J,n}]$, (5) is rewritten as

$$\mathbf{d}(n) = \mathbf{C}\boldsymbol{\alpha}(n). \quad (6)$$

where the (m, j) th element of \mathbf{C} is y_m^j .

The clutter model we employ in this work is based on the assumption that the random signal over the array in the absence of a mine is the sum of two components. The first portion is an additive zero mean, white Gaussian component of variance γ^2 representing sensor noise. The second component is intended to capture the effects due to the statistical distribution of the environmental factors not of primary interest in the detection problem. Here we assume that this signal will display the same basic smoothness as that seen in Fig. 3. That is it will be described by a low order polynomial. In the case of a mine, the coefficients of the polynomial are deterministically related to the position of the object relative to the array. For clutter however, these coefficients are taken to be random quantities. That is, $\boldsymbol{\alpha}(n)$ is a random vector. Such a model arises if we assume that the clutter represents the average effect on the sensor array of a random distribution of small scatterers. If each such scatterer results in a smooth signal over the array then so too will their sum. We note here that this choice of model certainly satisfies the criteria set forth in the Introduction that clutter essentially represents noise which is correlated from one array element to the next.

The second criteria is that the clutter should be correlated from one position of the array to the next. For example, a clutter vector at position n should be correlated to that at position $n + 1$. This implies

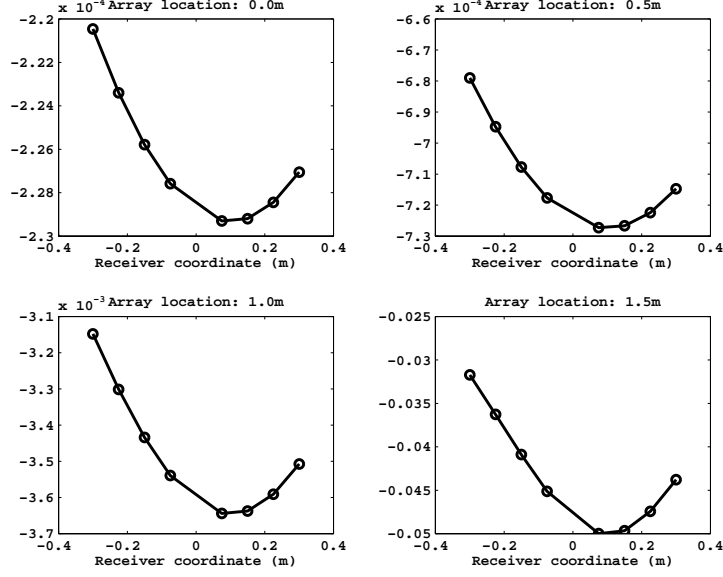


Figure 3. Data vectors, \mathbf{d}_j , obtained as array proceeds toward buried mine. Note the scale change in the plots.

the need to specify a dynamic model describing the evolution of $\boldsymbol{\alpha}(n)$. Toward this end, we employ the following model of the spatial evolution of $\boldsymbol{\alpha}(n)$:

$$\boldsymbol{\alpha}(n+1) = \boldsymbol{\alpha}(n) + \mathbf{B}(n)\mathbf{w}(n) \quad (7)$$

where $\mathbf{B} = \nu \text{diag}[\delta^{J/2}, \delta^{(J-1)/2}, \dots, 1]$, $\nu > 0$, $\delta \in (0, 1)$, and $\mathbf{w}(n)$ is a zero mean, unit variance, uncorrelated vector of Gaussian random variables. Finally, $\boldsymbol{\alpha}(0)$ is taken to be a zero mean Gaussian random vector with covariance matrix given by $\mathbf{B}^T \mathbf{B}$. Eq. (7) represents a random walk model for the vector $\boldsymbol{\alpha}(n)$. The matrix \mathbf{B} is such that variance of the additive noise is greatest for the higher order coefficients in (5) and decreases for the lower order α 's representing lower frequency components. The constant ν controls the overall power in the process and δ determines the relative variances among the coefficients.

To summarize, in the absence of a mine, the overall signal model we use for the remainder of the work is described by the following state-space model with $\boldsymbol{\alpha}(n)$ representing the state of the system:

$$\boldsymbol{\alpha}(n+1) = \boldsymbol{\alpha}(n) + \mathbf{B}\mathbf{w}(n) \quad (8)$$

$$\mathbf{d}(n) = \mathbf{C}\boldsymbol{\alpha}(n) + \mathbf{v}(n) \quad (9)$$

where $\mathbf{v}(n) \sim N(0, \gamma^2 \mathbf{I})$ is the additive white, sensor noise.

Sample paths of these clutter models are shown in Figs. 5(a) and 7(a) for an array consisting of 20 receiver elements for parameter choices described in greater depth in Section 5. In both cases, the variance of the $\mathbf{w}(n)$ is greater than that of $\mathbf{v}(n)$ so that the background “clutter to noise” ratio is assumed to be relatively large. These images indicate that our model of clutter possesses the desired features. As functions of y , they are quite smooth. As x (and hence n) increases, they display gradual changes except at a few points where the random walk model results in a sharp alteration of the “shape” of the profile.

In closing this section, we wish to emphasize that the clutter model described above is neither the only nor even necessarily the best such model for the processing of real data. Rather the precise model based on the underlying polynomial regression is intended more as a vehicle for demonstrating the efficacy of the detection algorithm described and tested in subsequent sections. Indeed, our approach towards detection is based only on the reasonable and generally unrestrictive assumption that one can describe the clutter using *some* state space model to capture the temporal correlation of the process. Any valid state space model will work and the problem of fitting field data to such a modeling structure represents an interesting and very worthwhile piece of future research.

4. DETECTION ALGORITHM

In this section, we develop our approach to the detection of buried mines from progressively obtained array EMI data. We begin by solving a simpler problem in which we assume that the mine can exist only at a known point in space and want only to determine whether the mine is present. The solution to this problem then allows for the solution to the problem where we lift the assumption of known location.

4.1. Case I: Known mine location

Assuming that the mine may only be located at position \mathbf{r}_0 , the first problem of interest is to determine whether the data we observe as the array proceeds down the track is just clutter or is comprised of clutter plus mine. From a statistical signal processing perspective, this amounts to a binary hypothesis test. Under the null hypothesis, H_0 , the data $\mathbf{d}(n)$ is comprised only of clutter and is therefore modeled using (8) and (9). Under the alternate hypothesis, the data are described by combining (8) and a modified observation equation:

$$\mathbf{d}(n) = \mathbf{s}_0(n) + \mathbf{C}\boldsymbol{\alpha}(n) + \mathbf{v}(n) \quad (10)$$

where $\mathbf{s}_0(n)$ is the signal over the array at time n arising from the mine at position \mathbf{r}_0 which can be computed using the model of Section 2.

A key difficulty in solving this binary decision problem is that the clutter is correlated. To overcome this problem, we employ a Kalman filter to “whiten” the data. In most cases, the Kalman filter is used to track the state of a stochastically evolving system. It is well known however that a bi-product of the minimum mean square error property of the filter is that it may be interpreted as a means of de-correlating a non-white noise sequence. Specifically, the Kalman filter takes as input the sequence $\mathbf{d}(n)$ and produces as output a new sequence $\mathbf{q}(n)$ termed the innovations. When there is no mine present, $\mathbf{q}(n)$ is a zero mean, unit variance, white Gaussian noise sequence. In the event that a mine is present, $\mathbf{q}(n)$ is again unit variance, white, and Gaussian, but now there is an added mean vector, $\boldsymbol{\rho}(n)$ (also known as a signature), which can be computed directly from $\mathbf{s}(n)$ and the matrices in the state space model for the noise.² Mathematically we have:

$$\text{Under } H_0 \quad \mathbf{q}(n) \sim N(0, \mathbf{I}) \quad (11)$$

$$\text{Under } H_1 \quad \mathbf{q}(n) \sim N(\boldsymbol{\rho}(n), \mathbf{I}) \quad (12)$$

That is, the detection problem now is one of determining whether or not there exists a known mean in a sequence of independent, unit covariance Gaussian random vectors.

In the case where the two hypotheses are to be distinguished based on a batch of data obtained between $n = 0$ and $n = N$, the problem stated in the previous paragraph has a well known solution given in the form of a likelihood ratio test. First, the following likelihood statistic is computed

$$l(N) = \sum_{j=0}^N \boldsymbol{\rho}^T(j)\mathbf{q}(j) - 0.5\boldsymbol{\rho}^T(j)\boldsymbol{\rho}(j). \quad (13)$$

Next, we compare $l(N)$ with a threshold $\tau(N)$. If $l(N) > \tau(N)$ we say that the mine is present otherwise, we continue to collect data. Typically, $\tau(N)$ is chosen such that the probability of false alarm P_{fa} (that is the probability of rejecting H_0 when H_0 is in fact true) is less than or equal to a predefined threshold.³ Under our Gaussian assumptions on the innovations, this threshold is³

$$\tau(N) = d(N) \operatorname{erfc}_*^{-1}(P_{fa}) - d^2(N)/2 \quad (14)$$

with

$$\operatorname{erfc}_*(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-x^2/2} dx \quad \text{and} \quad d^2(N) = \sum_{j=0}^N \boldsymbol{\rho}^T(j) \boldsymbol{\rho}(j)$$

For the demining application, we want the ability to declare the presence of a mine before the array actually passes the object. That is, we want to solve a sequential detection problem wherein at each n we decide whether to declare a mine, declare no mine, or move the array and collect more data. While well known methods exists for optimally solving this problem when $\boldsymbol{\rho}$ is independent of n , the solution to the problem for a time varying mean is less well studied. The algorithm in this paper represents a first attempt to solve this problem for the demining application.

Since we know where the mine is located, we can, in principle, take data up to but not including the time the sensor passes the mine. Thus, at each n we compute $l(n)$ as in (13) and for a fixed P_{fa} , perform the comparison to $\tau(n)$. We do not stop taking data until that point, n_0 , when the x coordinate of the array is such that at step $n_0 + 1$ the array would pass the mine. At that time we look at the sequence of decisions made for all previous n . If H_1 has been accepted more than Q times (in this paper $Q = 3$, that hypothesis is declared to be true.

We observe that we could have chosen the simpler approach of collecting data until time n_0 and then solving the fixed size binary hypothesis test. The methods described in the previous paragraph though are more easily generalized and provide greater insight into the more complex problem where we cannot wait until n_0 but must make the mine/no mine decision as rapidly as possible.

4.2. Case II: Unknown mine location

In the more realistic event that we know neither whether a mine is in a region of interest nor where it may be, we employ a “multiple model”² approach to the detection problem. As illustrated in Fig. 4, we hypothesize the existence of mines over a coarse grid extending in front of the moving array. Associated with the i th such mine location is a separate signature sequence $\boldsymbol{\rho}_i(n)$. At each n , we perform a separate binary likelihood ratio test for each location in the grid. When the array is set to pass those possible mine locations with a given x coordinate, we look to see if any of the hypotheses in that collection have been chosen more than M times. If none satisfies this criteria, we continue moving the array. If one location passes, we declare a mine. If more than one has been passed more than M times, we take as the actual mine detection that hypotheses which has exceeded its threshold by the widest margin at any time during the scan. Essentially, we take the most likely target if there are multiple possibilities.

The implementation of this procedure can be made quite efficient. Rather than keeping track of a large number of mine locations extending over the entire track of the array, we perform this procedure using a sliding window of hypothesized target positions. Let Δx be the distance the array moves from n to $n + 1$. We choose as the coarse grid of possible mine locations the points at the center of the squares on a grid of size $M\Delta x$. That is, it takes the array M steps to proceed through one column of the coarse scale grid. As shown in Fig. 4, assuming the array has been moving for a sufficiently long time so that the Kalman filter is operating in the steady state, there is a certain shift invariance to the detection problem. In particular, the mathematical forms of the hypotheses generated when the array is in column 1 and the mines in column

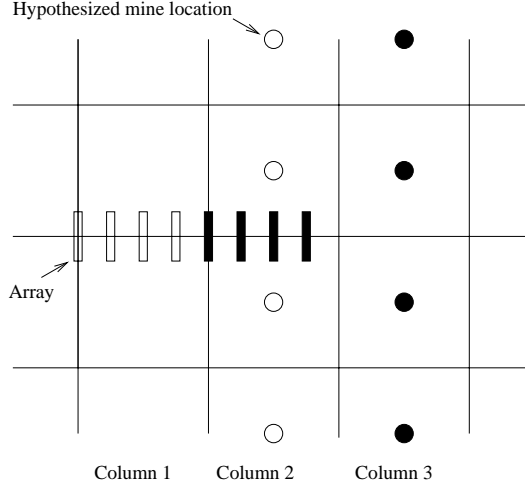


Figure 4. Relationship of moving array to coarse grid of possible mine locations. Rectangles represent positions of the array as it moves down the track. Circles represent hypothesized mine locations. There are $M = 4$ array positions per coarse grid cell.

2 (represented by the open boxes and circles) are identical to those obtained when the array is in column 2 and the mines are postulated to lie in column 3 (the darkened rectangles and circles). Clearly this holds for any array/mine combinations with identical relative spacings in the along-track direction.

We use this observation in the following manner. We begin by postulating mines over a $P \times P$ coarse grid in front of the array where P is chosen such that the response of the array when it is $PM\Delta x$ meters away is negligible. This grid (i.e. these hypotheses) are retained for the first M steps of the array. At the $M + 1$ st step, when the array is entering the second column, the grid looks very similar as in step 1 except the columns need to be renumbered. The first column, of the grid is removed since the array is leaving it. Each of the remaining columns are shifted to the left by one and a new column of hypotheses is added to the rightmost edge of the grid so that the overall array is still $P \times P$ only now looking further down track than was the case before. This procedure is repeated every M moves of the array.

This sliding window testing process is advantageous because it requires the storage of a fixed and relatively small number of signature vectors. Rather than having to maintain a $\rho(n)$ sequence for every conceivable combination of array locations and mine positions, we need only compute and store P^2M such sequences. For the EMI problems of interest here, excellent results are obtained in Section 5 for $P = 10$ and $M = 4$. The computation, storage, and manipulation of only 400 $\rho(n)$ sequences is relatively easy making this processing methodology quite amenable to real time implementation.

5. EXAMPLES

In this section, the performance of the detection approach is demonstrated and analyzed under a variety of noise conditions and mine placements. In all cases, the array is composed of a single transmitter located at $x = 0$ and 20 receivers with half equally spaced between $y = 5$ cm and $y = 50$ cm and the other half spaced between $y = -5$ cm and $y = -50$ cm. Thus, the overall length of the array is 1 meter. The transmitters and receivers are modeled as small circular loops of radius 2 cm located 10 cm above the air-earth interface and operating at 3 kHz. The array starts at $x = 0$ meters and proceeds in 75 equally spaced steps to $x = 3.7$ meters. The rolling grid of coarse scale hypothesized mine locations is composed of a 10 by 10

array of square cells stretching from about -90 cm to 90 cm in the y direction and extending about 2 meters in front of the array in the x direction. Finally, the mine-like target used in these simulations is modeled as a $5 \times 5 \times 5 \text{ cm}^3$ parallelepiped of contrast 1 S/m above the surrounding soil. Further, it is assumed that the burial depth of the mine is known *a priori*. Thus, the signature vectors are generated using this information. In practice, this assumption is not overly restrictive as range of burial depths is generally quite limited for a given demining application. Our scenario is meant to model the case of a shallowly buried anti-personnel mine, of interest for humanitarian demining.

For the low-frequency, Born-based inductive model discussed in Sec. 2, the electrical conductivity of the soil is approximated as zero. Moreover, the frequency and conductivity contrast of the mine enter the model only as multiplicative constants scaling the amplitude of the data. At higher frequencies and when an exact scattering model is employed, these various parameters will indeed play a larger role on the performance of the algorithm. For the work in this paper, however the key determination of performance is the relative strengths of the noise, clutter and target signatures. The measure used here is the signal to clutter plus noise ratio (SCNR) taken to be 20 times the base 10 logarithm of the ratio of the signal power to the noise plus clutter power taken over the entire data set obtained from $x = 0$ to $x = 3.7\text{m}$.

As a first example, we consider the detection of a mine located at $(x, y, z) = (2.2, -0.22, -0.05)$ m under a variety of SCNR. In Fig. 5(a) we display a sample data set. The mine, located in the image at $(x, y) = (2.2, -0.22)$ m, is barely discernible at this SCNR. In Fig. 5(b) the detection performance as a function of SCNR is shown. Each point represents the fraction of times the mine was detected in 100 runs of the algorithm using independent realizations of the clutter and noise processes for each run. The SCNR was changed by setting the variance of $\mathbf{v}(n)$ in (9) to 0.01 and increasing the variance of $\mathbf{w}(n)$ in (7) from 0.01 to 0.25 in steps of 0.02. It is clear, that strong performance is seen down to a SCNR of about -20 dB with a gradual decline in detection as the SCNR further degrades. At an SCNR of -15 dB the mine is detected essentially 100% of the time. Finally, since the mine was not located at one of the coarse scale grid points, a detection was declared if the estimated mine location was one of the four neighbors nearest the true mine location on the hypothesis grid.

In Fig. 6, we display the average value of the likelihood statistic less the threshold (i.e. $l - \tau$ from (13) and (14)) as a function of the position of the sensing array as the array proceeds toward the mine. The average is taken over all noise realizations and all experiments where the mine was in fact detected according to the criterion described previously. The SCNR for this problem was about 2.6 dB and the x coordinate of the mine is again 2.2 m. The plot does not reach 2.2 m because of the coarse scale gridding of the mine hypotheses, which for this problem tests for mines located at $x = 2.1$ m and $x = 2.3$ m. Fig. 6 clearly shows that the relevant decision statistic rises to a significant level above the detection threshold of zero well before the array actually passes over the buried object. In particular, these statistics become substantial when the array is located at $x = 1.95$ m which is over 25 cm from the location of the buried mine.

As a second example, we consider the problem of detecting a mine at $(x, y, z) = (2.2, -0.7, -0.05)$. This placement of the mine is interesting because, unlike the first case, the mine is located outside of the direct field of view of the EMI array. In particular, the array extends to $y = -50$ cm; however the mine is located at $y = -70$ cm. A sample data set at an SCNR of about 1.8 dB is shown in Fig. 7(a) and the detection results are given in Fig. 7(b). As in the first problem, the signature of the mine, which should be evident at around $x = 2.2$ m and $y \in (-.3, -.5)$ m, is obscured by the clutter; however the detection rate is quite high even to an SCNR of about -20dB. In Fig. 8, the plot of average likelihood statistics less threshold are shown for this problem at at SCNR of about 1. These plots show that even when the mine is located significantly off the axis of the array, we still retain a strong degree of forward looking, predictive capability. In particular, non-negligible test statistics are reliably observed for this scenario when the array is located at $x = 1.9$ m which is about 30 cm from the mine.

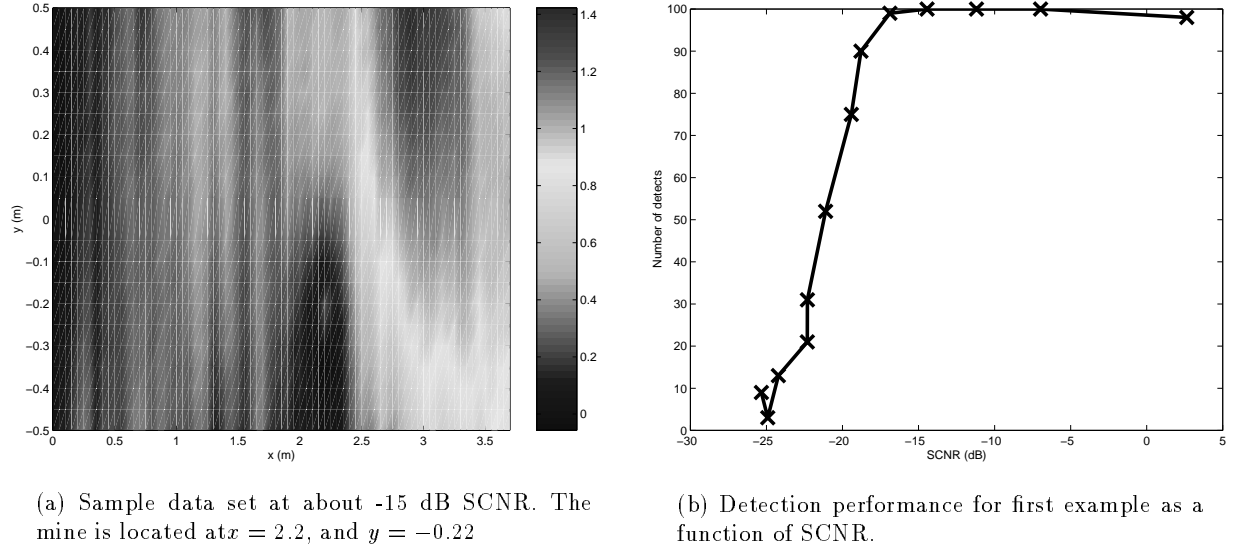


Figure 5. Sample data set and detection results for first example

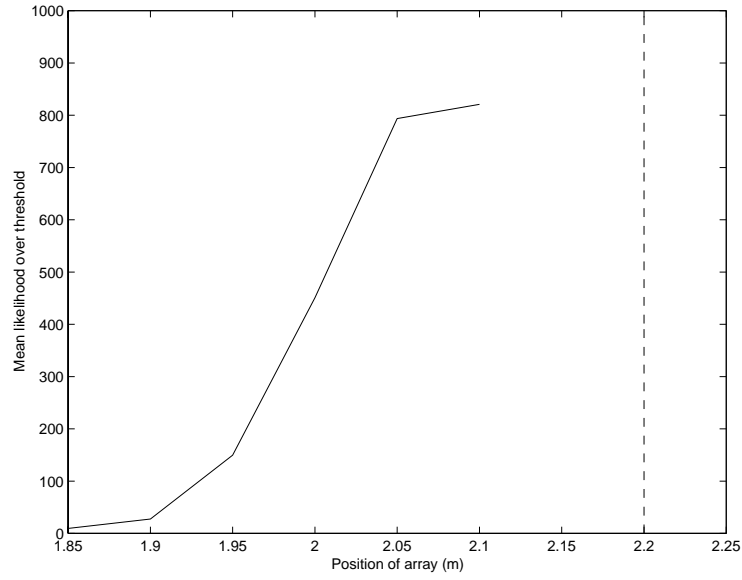


Figure 6. Average value of likelihood statistic in excess of threshold as array approaches mine location for first problem. Detection requires this value to exceed zero. The dashed line indicates the x coordinate of the mine.

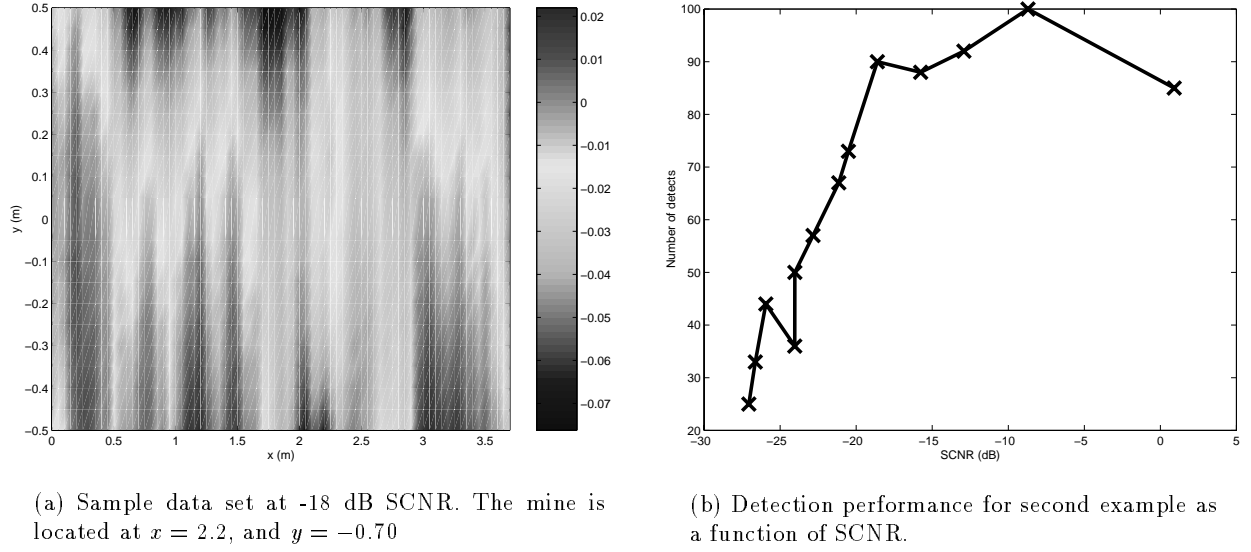


Figure 7. Sample data set and detection results for second example

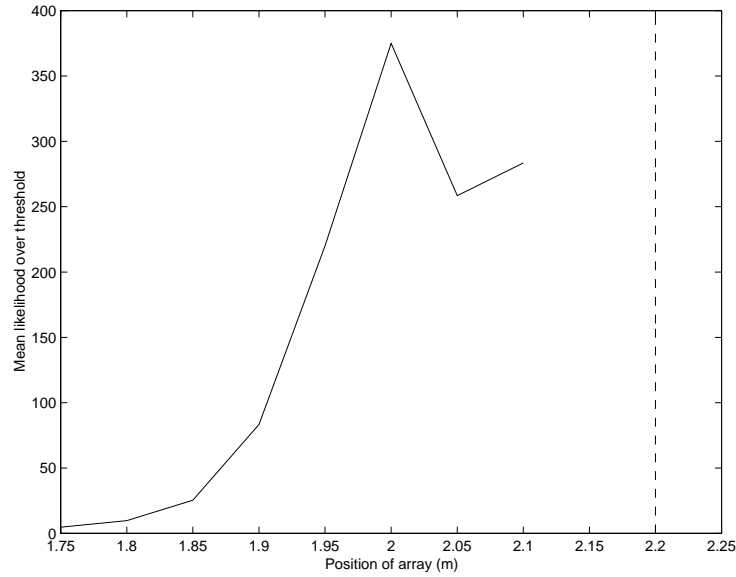


Figure 8. Average value of likelihood statistic in excess of threshold as array approaches mine location for second problem. Detection requires this value to exceed zero. The dashed line indicates the x coordinate of the mine.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented an approach for the detection and localization of buried landmines from EMI sensor data. Our statistical signal processing approach is tailored both to the sensor (an array of EMI transducers) and the data collection method (progressive acquisition as the array proceeds down track). The approach we have taken casts the detection problem in the form of a statistical hypothesis test. Using a novel state-space model for the clutter, we employ a Kalman filter to whiten the data and use a rolling window, multiple-model approach to ascertain the location of a mine. Preliminary simulations indicate that this approach can provide significant forward and side looking capability thereby increasing the area over which an EMI sensor can reliably detect buried mines.

While the work in this paper has been directed towards an EMI sensing device, the general detection methodology can be extended and applied to *any* scheme in which an array of sensors collects data in a progressive manner. All that is required is a forward model to generate the rather modest sized library of target signatures and a state-space or other recursive model for the clutter. With this in mind, future research efforts stemming from this project include the following:

1. Development of an estimation-based methods to better localize the detected object. Here we are interested in developing a non-linear least squares method which takes the coarse grid estimates generated by the algorithm in this paper and provides a more accurate determination of the mine's true position.
2. Development of methods for “discounting” detected mines from the data set. After the mine has been detected, we want to remove its effects from the data thereby allowing the array to continue down track looking for other buried objects. Failure to perform such discounting generally degrades subsequent performance.
3. Development of a full EMI scattering model for use in generation of target signatures.
4. Identification of state space type model for EMI array from real field data.
5. Extension and adaptation of these methods to ground penetrating radar (GPR) sensor.
6. Use of Kalman filtering ideas for fusion of GPR and EMI sensor data.

REFERENCES

1. W. C. Chew and R. J. Kleinberg, “Theory of microinduction measurements,” *IEEE Trans. Geoscience and Remote Sensing* **26**, pp. 707–719, Nov. 1988.
2. M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice Hall, 1993.
3. H. L. Van Trees, *Detection, Estimation and Modulation Theory: Part I*, John Wiley, New York, 1968.

Efficient T-Matrix Methods for GPR Forward Modeling

Adnan Şahin¹ Eric L. Miller

Center for Electromagnetic Research, 235 Forsyth Building
Northeastern University, Boston, MA 02115

Telephone: (617) 373-8386

Telefax : (617) 373-8627

email: adnan@cdsp.neu.edu

An important component associated with the GPR problem is the computation of the scattered fields produced by buried objects when illuminated by the radar source. The choice of technique for computing these fields is often driven by a variety of factors including computational complexity and the flexibility to handle easily a wide range of configurations of scatterers. Based on the fact that objects of interest for many GPR applications have relatively simple shapes such as cylinders and spheres, it is apparent that transition matrix (T-matrix) approach is well suited to the GPR forward problem.

The T-matrix technique is based on the expansion of the fields and the Green's function in spherical (3D) or cylindrical (2D) harmonic series. Considerable work has focused on the generation of efficient recursive T-matrix techniques for computing the scattered field from a single, electrically large, dielectric object. The approach taken in these applications is to tessellate the object into many smaller scatterers. Thus, even though the number of scatterers is large, the order of harmonics used to expand the basis functions is small thereby keeping the computational load relatively small. However, if one wants to find the scattered field from multiple, electrically large, tessellated objects, the number of scatterers grows quite quickly thereby increasing the complexity of the problem considerably. A natural alternative is to treat each object as a single scatterer (i.e. bypass the tessellation) but increase the order of the harmonic expansions for each object. Unfortunately, this approach is hindered by difficulties related to the slow convergence of the harmonic expansions.

Here, we present a modified, stable recursive T-matrix algorithm to calculate the scattered field from a heterogeneous collection of spatially separated objects based on these higher order expansions. Motivated by the GPR problem, we consider the calculation of the scattered field produced by a group of parallel cylinders buried in a lossy medium. In the most general case, these objects can be dielectric or metallic, electrically small (fraction of λ) or large ($1-2 \lambda$). The modification, which will be described in the talk accompanying this abstract, circumvents the slow convergence of the series by replacing the product of two translation matrices with product of three such matrices two of which yield the identity.

In the talk we will also discuss the verification of this algorithm against previously published results for scattered fields from multiple cylinders, compare the computational advantages of this higher order technique relative to the tessellated approach, and provide examples demonstrating the ability of this code to compute fields for GPR-relevant scattering scenarios such as mine detection and hazardous waste removal.

¹The work of this author was supported in part DOE contract DE-FC07-95ID13395, NSF Grant MIP-9623721, and by subcontract GC123920NDG from Boston University under the AFOSR MURI Program on Reduced Signature Target Recognition

Combined High-dimensional Analysis of Variance (HANOVA) and Sequential Probability Ratio Test (SPRT) to Detect Buried Mines

Xiaoyin Xu, Eric L. Miller, and Carey M. Rappaport

235 Forsyth Building
Northeastern University
Boston, MA 02115
elmiller@ece.neu.edu

ABSTRACT

We apply high-dimensional analysis of variance (HANOVA) and sequential probability ratio test (SPRT) to detect buried land mines from array ground penetrating radar (GPR) measurements. The GPR array surveys a region of interest in a progressive manner starting at a known position and moving step by step in a fixed direction. Detection method consists of two parts. Because, at each stop, path lengths are different from radar pairs to a mine target, there exists statistically significant difference among received signals when a mine target is present. At first, we use HANOVA to test for statistical significance at each stop. The HANOVA compares multiple sets of data for statistical difference. While HANOVA detects difference among received signals in time-domain, it does not incorporate new data as the GPR array moves down-track. So secondly, we resort to sequential detection to look for changes in result of HANOVA along the GPR array stops. It turns out to be a sequential probability ratio test to detect transient curved signals. This SPRT allows real-time processing as new data are obtained by the GPR array. Performance of this test is analyzed through the probability of detection and mean time between false alarms. Finally, real data and computer simulation are processed to verify the method.

Keywords: Analysis of variance (ANOVA), sequential probability ratio test (SPRT), transient signal, curved data, GPR signal

1. INTRODUCTION

The use of ground penetrating radar (GPR) arrays for detecting buried objects has received considerable attention in recent years in areas such as landmine and unexploded ordnance remediation, utility line mapping, and archaeology.¹ A typical GPR array configuration for such applications is shown in Fig. 1. Here one uniform linear array of transmitters and a second uniform linear array of receivers are simultaneously moved down a linear track. At every stop of the system, each transmitter emits a short pulse of electromagnetic energy which interacts with the surrounding medium. Based on observations of scattered fields collected by the corresponding receiver the objective of the problem is to determine if a mine is present in the field of view of the array.

Current processing methods for this problem fall into one of three categories. First, pattern matching methods² employ techniques such as fuzzy set theory and neural networks. Such methods can be fast but also require extensive training to function well. Moreover, performance analysis is limited to Monte-Carlo simulations. Second, image-then-detect techniques³ employ a beamforming or backpropagation approach to build an image of the subsurface which is then post-processed to detect objects. Such methods generally require the data from the full GPR scan to form an image and are thus not well suited to on-line computations in which information is processed sequentially as the array proceeds down track. Finally, there has been some very interesting work done in the area of statistical processing methods⁴ where one can examine quantities such as detection rates, false alarm probabilities, etc.; however the techniques in⁴ for instance are based on highly complex electromagnetic models for the GPR sensor and are thus computationally intensive.

Here we view the GPR array detection problem in a blind signal detection framework and employ statistical methods to process the GPR returns. By “blind”, we mean that the exact form of signal reflected from a mine is unknown. What we know is that this received signal becomes stronger as the GPR array moves toward a mine and then the signal becomes weaker after the array passes over the mine. Theoretical analysis and computer simulation

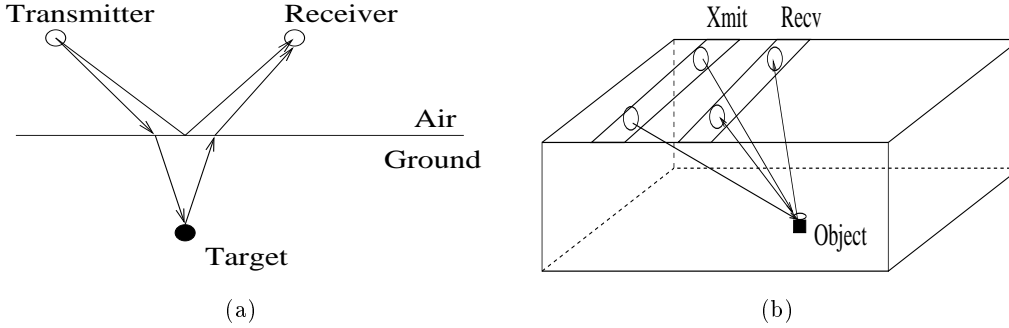


Figure 1. GPR, a) a single GPR system, b) a GPR array.

indicate that this change follows a hyperbolic curve, though parameters of this hyperbolic curve are not known beforehand. On the other hand, different T/R pair registers a different hyperbolic curve, it is so because T/R pairs are spatially separated. This information allows us to exploit two generic properties of the signal transmission process. First, for any given stop of the array, the presence of a mine close to a T/R pair results in a jump in the mean value of the observed signal relative to that seen in other pairs. Thus, we develop a high-dimensional ANOVA test⁵ to detect this change. Second, physical principles dictate that as the array moves from one stop to the next, this change will increase and then decrease; that is, the change will be transient in space domain.¹ A sequential detector is employed to track and detect this spatial up-and-down change as the GPR array moves down-track.

The paper is organized as follows. Section 2 discusses physical model of a GPR system and corresponding mathematic model. In Section 3, we present a statistical algorithm to detect mines and analyze its performance in mean time between false alarms and probability of detection. Both field data and simulation are used to show how the algorithm works. Conclusion and direction of future work is given in Section 4.

2. MODEL

2.1. GPR Model

At first, we consider a single GPR system.⁶ It consists one transmitter and one receiver. After each transmission, the receiver collects echo for a certain amount of time. There are two to three components in received signal. One is measurement noise, assumed to be white and Gaussian. The second component is reflection from ground-air interface, which we call clutter. Clutter is always present and changes as the GPR moves. We assume that it varies slowly as the array moves down the track so that it can be estimated and taken out by subtraction.

In Fig. 2 we plot the clutter subtracted observation for single T/R pair obtained from field data as a function of down-track position of the array. Each column of this image is a time-series of observation for a given stop of the array. As can be seen from this figure signal is transient in two ways. First, it appears only when the array is in the vicinity (≈ 0.5 meter) of the mine. For most systems this implies that there are on the order of tens of stops of the array when the mine can be detected out of hundreds to thousands of array positions in a typical GPR survey. Second, when the array is positioned close to the object, the effects are only seen in a portion of the time-series. Signal reflected from the mine always comes behind clutter and attenuates quickly in soil. Before proceeding, we make two observations regarding the signals in Fig. 2. First in general the precise form of the signal of interest is typically unknown due to variations caused by fluctuations in the electrical properties of the soil, unmodeled physical effects such as surface roughness, and variability in the signature caused by the unknown orientation of the mine relative to the sensor. Indeed, Fig. 2 is typical of the variability seen in practice. Second, for clarity, we have displayed signals from two mines which are relatively easy to detect. Generally, even after removing the clutter other sources of noise and interference can significantly reduce the detectability of the desired signal and increase number of false-alarms. In Fig. 3, we can see that besides signals reflected from two mines, there are other interferences as well.

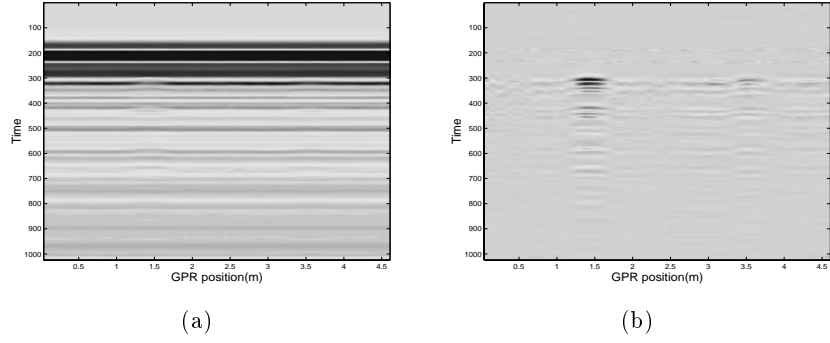


Figure 2. Signal generated from two objects buried at position 1.5m and 3.0m.

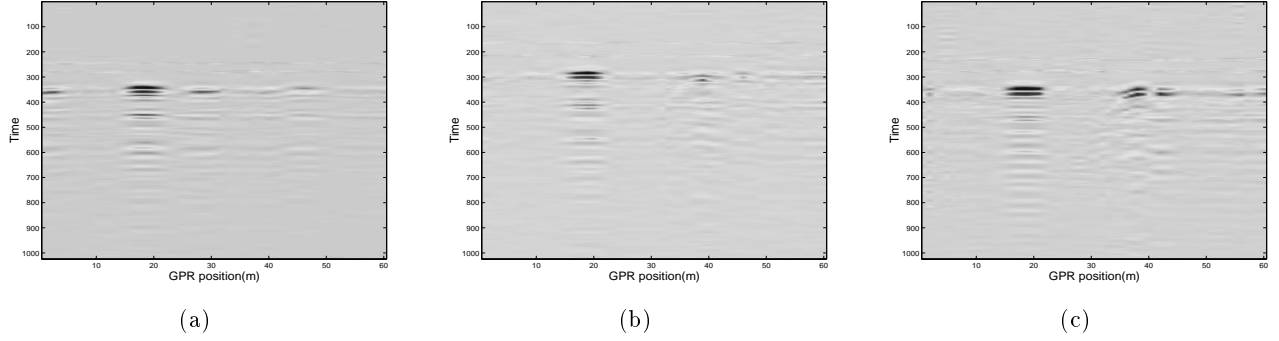


Figure 3. Signal from other three T/R pairs, from left to right, pair 2, 3, and 4.

2.2. Mathematic Model

Based on this discussion, at down-track position n of the array, after clutter has been removed, we have the classic binary hypothesis model for the GPR signal returns

$$\begin{aligned} H_0 : \quad & \mathbf{x}_m(n) = \mathbf{w}_m(n) \\ H_1 : \quad & \mathbf{x}_m(n) = \mathbf{s}_m(n) + \mathbf{w}_m(n) \end{aligned} \quad (1)$$

for $m = 1, \dots, M$, $n = 1, \dots, N$, M is number of T/R pairs, N is number of array stops. Here $\mathbf{x}_m(n)$, $\mathbf{s}_m(n)$, $\mathbf{w}_m(n)$ are column vectors of length K , representing time series of observation, signal, and noise from the i th T/R pair. The noise is assumed to be white and Gaussian with zero-mean and variance σ_m^2 . While the exact structure of \mathbf{s}_m is unknown, we see that the effects of this signal are to cause a jump in the mean of the \mathbf{x}_m for those n where a mine is in the field of view of the GPR. Thus the statistical problem of interest in this work is to detect this transient signal based on M such observation vectors at N steps.

As detailed in the remainder of this paper, we approach this problem in two steps. First, we develop a method for detecting the presence of \mathbf{s}_m for a fixed location of the GPR array. Second, we use this test in a sequential manner to process the returns as the array acquires new data.

At a given position of the array since we do not know of \mathbf{s}_m the statistical problem we pose is

$$\begin{aligned} H_0 : \quad & \boldsymbol{\mu}_m = \mathbf{0} \\ H_1 : \quad & \boldsymbol{\mu}_m \neq \mathbf{0}, \quad m = 1, \dots, M \end{aligned}$$

with $\boldsymbol{\mu}_m = E[\mathbf{x}_m]$. In many cases, analysis of variance (ANOVA) is used to solve this problem. Recently however, Fan,⁷ Fan and Lin,⁵ has noted that the performance of ANOVA suffers for problems when the signal of interest is

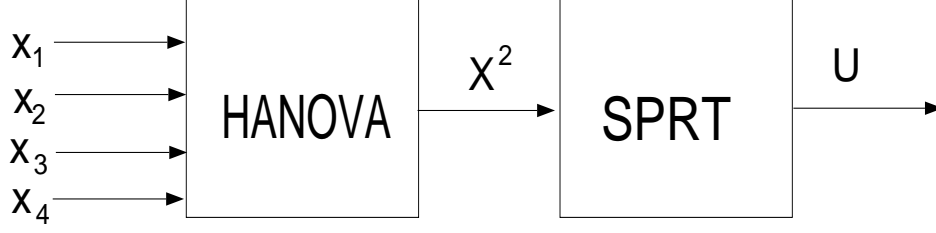


Figure 4. Flowchart, 4 vectors of input.

limited to a small number of samples in the overall observation vector as is the case for the GPR problem of interest here. Fan proposed a high-dimensional ANOVA which uses a portion of observation. Fan’s original work was limited to problems in which the first k samples were used, with k found from the data. Here we consider a generalization of Fan’s work to take into account the fact that for the GPR problem the transient object is significant over a window not generally starting with the first sample. Following the HANOVA, we turn to a sequential probability ratio test (SPRT) to make on-line detection as new data is acquired.

3. ALGORITHM

3.1. Estimation and detection

Our method consists of two steps, as shown in Fig. 4. Here the GPR array has four T/R pairs. At each stop of the GPR array, we receive four vectors of observation. These vectors are input to a high-dimensional ANOVA and the HANOVA generates one test statistic X^2 . Next, X^2 is fed into a sequential detector and produces a final test statistic $U(n)$. When $U(n)$ exceeds a preset threshold, a mine is declared.

Before describing HANOVA, it is helpful to see why in some cases ANOVA loses its discrimination power. For example, assume we have one observation vector of size $K \times 1$, $\mathbf{x}_1 \sim N(\boldsymbol{\mu}, \sigma_1^2 \mathbf{I})$ and we wish to test $H_0 : \boldsymbol{\mu} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\mu} = \boldsymbol{\mu}_1$. Standard ANOVA is to estimate $\boldsymbol{\mu}$ by \mathbf{x}_1 and use the testing procedure $X_1^2 = \|\mathbf{x}_1\|^2$. The approximate power of the standard ANOVA estimator-correlator test is

$$P_d(H_1|H_1) = Q\left(\gamma - \frac{\sum_{k=1}^K \mu_1^2(k)}{\sigma_1^2 \sqrt{2K}}\right) \quad (2)$$

where γ the test threshold and Q the complementary cumulative distribution function and strictly decreasing. If $\boldsymbol{\mu}_1$ is different from $\mathbf{0}$ for only a small number of k then as K goes large, $\sum_{k=1}^K \mu_1^2(k) \ll \sqrt{K}$. From (2), we then conclude that testing all dimensions of the data actually causes the test to lose power due to the accumulation of stochastic noise. This deterioration is reflected in the factor $1/\sqrt{2K}$. Fig. 5 shows how much target signal differs from background, it is noticeable that target signal does not appear for the first hundreds of dimensions and it quickly attenuates to zero after that. Fig. 6(a) shows that if starting from point 1, one begins to increase dimensions tested of signal of Fig. 5(c), the probability of detection at first increases, i.e., when more signal is taken into computation, and then decreases when signal runs out. On the other hand, if one fixes the end point around 500 and start from point 1, testing fewer and fewer dimensions, the probability of detection achieves its maximum around 300. Thus it provides us with an idea which portion of time series to test for statistical significance.

Based on this observation, Fan *et al.* developed an adaptive HANOVA test of the form described previously. We generalize it to GPR returns to form a two-end HANOVA, namely, testing a middle portion of time-series. Fig. 7 shows results from an ANOVA and a HANOVA test for the signals in Fig. 2. The HANOVA truncates observation vectors of length K to length L by two steps 1) discarding the first k_1 components which are consisted of ground-bounce only (the target reflected signal always comes after the ground-bounce) and 2) discarding the last k_2 elements which represent signal that has been attenuated too much to be meaningful for signal processing. Both the ANOVA and HANOVA tests easily detect the stronger signal, at position 1.5m. We want to be able to find the very weak target at 3.0m. For this target, the HANOVA shows improved performance. Specifically, the peak in the HANOVA case is more clearly visible above the nominal “noise floor.”

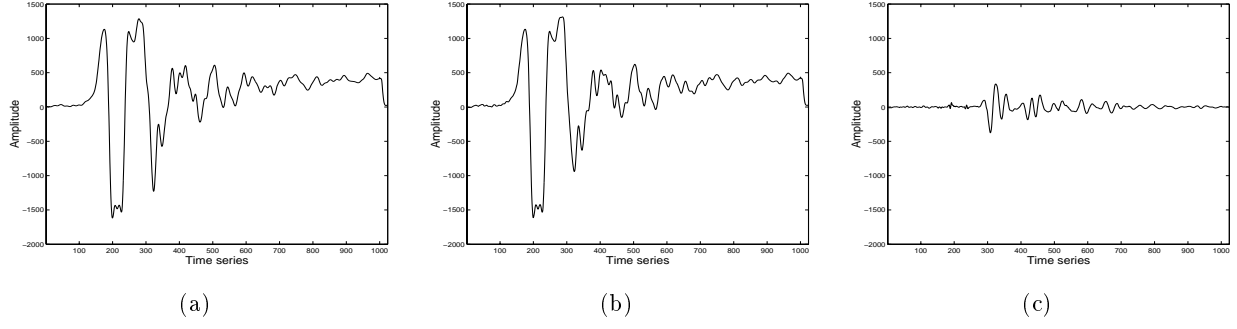


Figure 5. Received signal, a) clutter and noise, b) signal plus clutter and noise, c) signal only

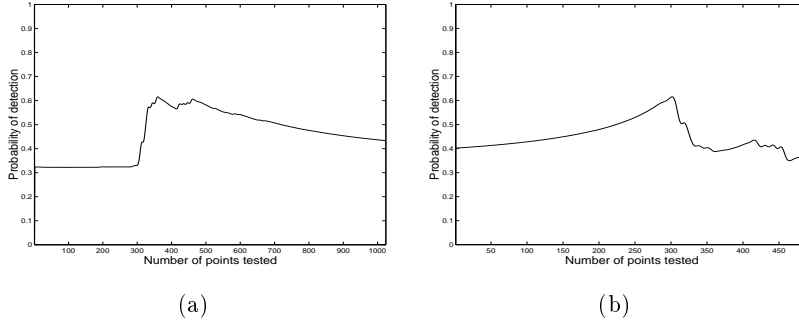


Figure 6. Probability of detection a) start from point 1 and test more and more dimensions of observations b) fix the end point and test fewer dimensions from point 1

Generalizing HANOVA to multiple observations, we build the test statistic as

$$X^2 = \sum_{m=1}^M \sigma_m^{-2} \|\mathbf{x}_m - \bar{\mathbf{x}}\|^2 \quad (3)$$

with $\bar{\mathbf{x}} = \sum_{m=1}^M \sigma_m^{-2} \mathbf{x}_m / \sum_{m=1}^M \sigma_m^{-2}$. It is easily shown that

$$X^2 \sim \chi_{ML}^2(\delta^2) \quad (4)$$

where

$$\delta^2 = \sum_{m=1}^M \sigma_m^{-2} \|\boldsymbol{\mu}_m - \bar{\boldsymbol{\mu}}\|^2 \quad (5)$$

with $\bar{\boldsymbol{\mu}} = \sum_{m=1}^M \sigma_m^{-2} \boldsymbol{\mu}_m / \sum_{m=1}^M \sigma_m^{-2}$ and $\chi_{ML}^2(\delta^2)$ is the χ^2 distribution with ML degrees of freedom and non-centrality parameter δ^2 .

While HANOVA detects difference among observations at one stop of the array, it does not capture the curvature structure seen as the array moves down-track. To improve detection performance, we employ a sequential detection scheme based on the processing of the HANOVA statistic to look for the transient signal from one stop of the GPR array to the next. Hence, the two hypotheses are

$$\begin{aligned} H_0 : & \quad X^2(n) \sim \chi_{ML}^2(0) \\ H_1 : & \quad X^2(n) \sim \chi_{ML}^2(\delta^2(n)) \end{aligned} \quad (6)$$

for $n = 1, \dots, N$ where $\delta^2(n)$ is defined in (5). At stop n , the log likelihood ratio for this problem is

$$u(n) = \ln \frac{p_n(X^2(n))}{p_0(X^2(n))} \quad (7)$$

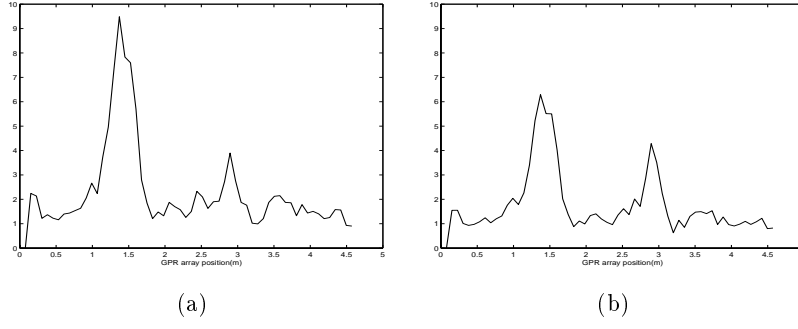


Figure 7. Test statistic, a) result of an ANOVA, using all dimensions of observations, b) result of a HANOVA, using truncated observations.

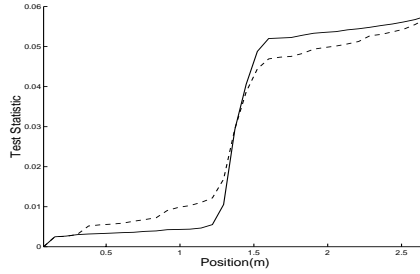


Figure 8. Test statistic $U(n)$ of sequential detection, solid line: HANOVA followed by SPRT, dash line: ANOVA followed by SPRT.

where $p_n(X^2(n)) \sim \chi_{ML}^2(\delta^2(n))$ evaluated at $X^2(n)$ and $p_0(X^2(n)) \sim \chi_{ML}^2(0)$ evaluated at $X^2(n)$. One difficulty with generating $u(n)$ is that $\delta^2(n)$ is typically not known *a priori* since the underlying μ_m are not assumed known. Here we construct a generalized log likelihood statistic. Typically, this is done by replacing $\delta^2(n)$ by its maximum likelihood estimate which for this problem is $X^2(n)$. After some experimentation in real data and computer simulation, we have found that performance can be improved by incorporating a one-step delay into the processing. Specifically, we estimate $\delta^2(n)$ as $X^2(n-1)$.

The full algorithm then is essentially a repeated SPRT⁸ and is summarized by the following steps,

- $U_1 = 0$.
- FOR $n = 2, \dots, N$
 - $\hat{\delta}^2(n) = X^2(n-1)$
 - Form $u(n)$ according to (7)
 - $U(n) = \max(0, U(n-1) + u(n))$.
 - IF $U(n) > \alpha$, declare object.
- ENDFOR

where α is a preset threshold. Fig. 8 shows the result of the algorithm applied on HANOVA output of Fig. 7. Clearly, sequential detection using results of HANOVA has a better performance.

3.2. Performance

Performance of this algorithm is studied in terms of mean time between false alarms and probability of detection. Under H_0 , the test statistic changes as a Markov chain,⁹ with update $u(n)$ at each step. Under H_1 , the update $u(n)$ is changing, and therefore, the test statistics $U(n)$ can be described as an inhomogeneous Markov chain. One method

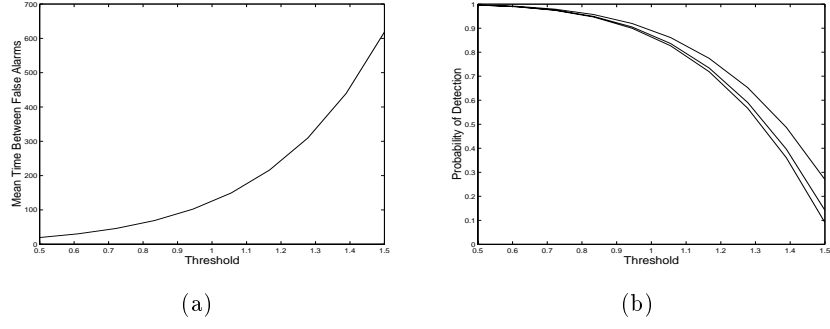


Figure 9. Performance, a) mean time between false alarms with different threshold and number of quantization levels, b) probability of detection at different signal to noise ratio, from top to bottom, SNR = 1.85dB, 2.2dB, 2.8dB.

of analyzing Markov chain is a matrix approach.⁹ First, since L is usually very large, in the order of hundreds, and therefore ML is large, χ^2 distribution can be approximated by a Gaussian distribution. We then have

$$\begin{aligned} H_0 : X^2(n) &\sim N(ML, 2ML) \\ H_1 : X^2(n) &\sim N(ML + \delta^2(n), 2ML + 4\delta^2(n)). \end{aligned}$$

Second, suppose the update $u(n)$ and the interval $[-\alpha, \alpha]$ is uniformly quantized to $2k + 1$ levels such that $v_0 = 0$, and $v_k = \alpha$. Let $\pi_{nl} = \Pr\{U_n = v_l | N > n\}$ for integer $l \in [0, k - 1]$, and $\pi_n = [\pi_{n0}, \pi_{n1}, \dots, \pi_{n(k-1)}]^T$ denote the probabilities of $U_n = v_k$ on the condition that there is no alarm until after sample time n . We can write the transition matrix $C^{(n)}$ at step n as⁹

$$C^{(n)} = \begin{bmatrix} p_n^0 & p_n^{-1} & \cdots & p_n^{-\gamma+1} \\ p_n(1) & p_n(0) & \cdots & p_n(-\gamma+2) \\ p_n(2) & p_n(1) & \cdots & p_n(-\gamma+3) \\ \vdots & \vdots & \ddots & \vdots \\ p_n(\gamma-1) & p_n(\gamma-2) & \cdots & p_n(0) \end{bmatrix} \quad (8)$$

where

$$p_n(l) = \Pr(\log \frac{f_1^{(n-1)}(X^2(n))}{f_0(X^2(n))} = v_l) \quad (9)$$

and elements in first row of Eq. 8 are $p_n^i = \sum_{l=-\infty}^i p_n(l)$. Here $f_1^{(n-1)}(X^2(n))$ denotes the PDF of $X^2(n)$ under the alternative hypothesis H_1 and using $X^2(n-1)$ to estimate $\delta^2(n)$. Under H_1 , the transition matrix $C^{(n)}$ can be built using Eq. 8 for different n , since signal is time varying. The probability of detecting a signal of length n_d is then

$$P_d(n_d) = 1 - \mathbf{1}^T \prod_{n=1}^{n_d} C^{(n)} \pi_0 \quad (10)$$

where $\mathbf{1}$ is a column vector of all ones. The mean time between false alarms \bar{T} is found to be⁹

$$\bar{T} = \mathbf{1}^T (\mathbf{I} - C^{(0)})^{-1} \mathbf{e}_1 \quad (11)$$

in which $C^{(0)}$ is the transition matrix under the signal-absent statistics and \mathbf{e}_1 is a column vector containing a one in position 1 and zeros elsewhere. With more levels of quantization, we obtain more accurate estimation of P_d and \bar{T} . Fig. 9(a) shows the mean time between false alarms under the null hypothesis H_0 , for different thresholds.

Here, we define the signal to noise ratio as

$$\text{SNR} = 10 \log_{10} \frac{\sum_{n=n_1}^{n_2} \mathbf{s}_n^T \mathbf{s}_n}{(n_2 - n_1 + 1)L\sigma_m^2} \quad (12)$$

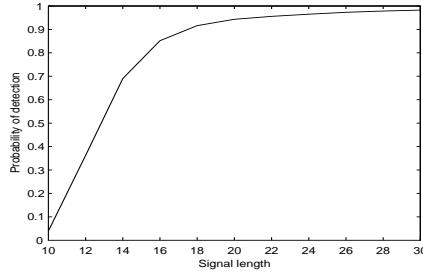


Figure 10. Probability of detection versus transient length of the signal.

where n_1, n_2 are chosen to include those GPR stops where a mine is close. As noise level reduces, probability of detection increases, as shown in Fig. 9(b). Finally, we simulated transient signals of different lengths on a computer. Fig. 10 shows the probability of detection at a fixed threshold when the length of the signal is changing. Generally, increasing signal length improves its SNR. As expected, a signal of large transient length can be detected at a higher probability.

4. CONCLUDING REMARKS

In this paper, we have proposed a sequential, high-dimensional ANOVA to process multichannel GPR returns. The method works in two directions: first it looks for statistically significant difference among multichannel observations, second, it carries out a sequential detection as new data are obtained. The HANOVA is powerful in the sense of maximizing probability of detecting statistically significant difference among sets of observations. Sequential detection enables real-time processing as new data are collected. The method has a relatively low computational complexity. We have demonstrated the performance of this technique on a sample of field data. Future research will focus on adaptive sampling of received signal to take into consideration of roughness of ground-air interface; on-line localization will be integrated to allow detect-localize-detect and thus provides higher probability of detection.

REFERENCES

1. L. Peters Jr., J. J. Daniels, and J. D. Young, "Ground penetrating radar as a subsurface environmental sensing tool," *Proceedings of the IEEE* **82**, pp. 1802–1822, Dec. 1994.
2. P. G. et al., "New results in fuzzy set based detection of landmines with gpr," in *Detection and Remediation of Mines and Minelike Targets IV*, vol. SPIE 3710, pp. 1075–1084, 1999.
3. M. B. et al., "Mine detection with a multichannel stepped-frequency ground-penetrating radar," in *Detection and Remediation of Mines and Minelike Targets IV*, vol. SPIE 3710, pp. 953–960, 1999.
4. T. Dogaru and L. Carin, "Time-domain sensing of targets buried under a rough air-ground interface," *IEEE Trans. Antennas and Propagation* **46**, pp. 360–372, Mar. 1998.
5. J. Fan and S. Lin, "Test of significance when data are curves," *Journal of American Statistical Association* **93**, pp. 1007–1021, 1998.
6. J. M. Bourgeois and G. S. Smith, "A fully three-dimensional simulation of a ground-penetrating radar: FDTD theory compared with experiment," *IEEE Trans. Geosci. Remote Sensing* **34**, pp. 36–44, Jan. 1996.
7. J. Fan, "Test of significance based on wavelet thresholding and Neyman's truncation," *Journal of American Statistical Association* **91**, pp. 674–688, 1996.
8. M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall, New York, 1993.
9. C. Han, P. K. Willett, and D. A. Abraham, "Some methods to evaluate the performance of Page's test as used to detect transient signals," *IEEE Trans. Signal Processing* **47**, pp. 2112–2127, Aug. 1999.

A Nonlinear Shape Reconstruction Algorithm For Cross-Borehole Electromagnetic Tomography Using Adjoint Fields And Level Sets

We consider the problem of recovering the shapes and locations of a collection of inclusions in an inhomogeneous medium from limited observations of scattered electromagnetic fields. In particular, we are concerned with a crosswell tomography problem arising in environmental monitoring and remediation. Here one starts with a pair of boreholes. The first is equipped with an array of electromagnetic sources and the second with an array of receivers. The medium between the two boreholes consists of a known inhomogeneous (e.g. layered) background, and some unknown localized inclusions, for example pollution plumes. The inclusions differ from the background medium in their complex electrical permittivities. The objective of the problem is the characterization of these structures from time harmonic data collected at a small number (less than 10) of frequencies.

The limited data set does not support the recovery of a fine scale collection of pixelated permittivity values. However, in many applications, including ours, it can be assumed that some prior information about the physical parameters inside the inclusions is available. For example, the permittivities inside the inclusions might be known constants. In these cases, all that remains is the recovery of the number, size, shape and locations of the inclusions.

We apply a two-step nonlinear inversion scheme to this problem with prior information. In the first step, we try to determine basic information about the number and the locations of the unknown objects by employing a recently developed nonlinear generalization of the algebraic reconstruction technique (ART) in x-ray tomography, to the given problem. The second step of the reconstruction makes use of level set methods as well as the prior information at our disposal. Introducing level set calculations based on the output of nonlinear ART allows us to recover more precise information regarding the shapes and sizes of the inclusions.

An essential feature of this inversion scheme is that the practical realizations of the two steps are closely linked with each other. In fact, the level set step simply continues with the same inversion routine as it is already used for nonlinear ART. All that has to be done numerically when switching from one step to the other is to add a post-processing routine to the calculated update formula. This yields already the correct deformation rule for the considered level sets.

In the talk accompanying this abstract we will provide a more detailed description and analysis of the various components of this inversion routine. Included in this discussion will be numerous examples demonstrating the utility of our method for problems of practical interest.

Session/topic: Inverse Scattering, Electromagnetic Signal Processing

**Author information sheet for “A Nonlinear Shape Reconstruction Algorithm For
Cross-Borehole Electromagnetic Tomography Using Adjoint Fields And Level Sets”**

Oliver Dorn
Center For Electromagnetics Research
Dept. of Electrical and Computer Engineering
Northeastern Univ., Boston MA, 02115
Tel: (617) 373-4617 FAX: (617) 373-8627
email: dorn@cer.neu.edu

Eric Miller
Center For Electromagnetics Research
Dept. of Electrical and Computer Engineering
Northeastern Univ., Boston MA, 02115
Tel: (617) 373-8386 FAX: (617) 373-8627
email: elmiller@cer.neu.edu

Carey Rappaport
Center For Electromagnetics Research
Dept. of Electrical and Computer Engineering
Northeastern Univ., Boston MA, 02115
Tel: (617) 373-2043 FAX: (617) 373-8627
email: rapapport@neu.edu

Reduced Complexity, Newton-Type Methods for Nonlinear Inverse Scattering Problems*

Eric L. Miller
235 Forsyth Building
Northeastern University
Boston, MA 02115
Tel: (617) 373-8386 Fax: (617) 373-8627
email: elmiller@cdsp.neu.edu

Prof. Andreas Mandelis and
Ms. Lena Nicolaides
Dept. of Mechanical Engineering,
Univ. of Toronto
5 King's College Road
Toronto M5S 3G8, CANADA
Phone/Fax: (416) 978-5106
e-mail: mandelis@mie.utoronto.ca

The inverse scattering problem (ISP) is characterized by a desire to determine the internal structure of a medium given scattered energy obtained along the boundary. Typical methods for performing the inversion amount to formulating the reconstruction problem as a solution to a regularized, nonlinear optimization problem and performing the optimization using an iterative algorithm. Rapid convergence often can be achieved through using a Newton-type approach where each iteration is decomposed into two steps. First, an incremental improvement is made to the estimated material properties by solving a least squares problem. Second, the new profile is used to solve for the internal field distribution over the region.

Although such an algorithm along with proper regularization can address the problem of ill-posedness, the Newton approach is known to be computationally complex. Constructing the matrices defining the least-squares problem requires the matrix representation of resolvent for the second kind integral equation relating the scattered to the incident fields for an inhomogeneous medium whose properties are defined by the current estimate of the material parameters. Direct knowledge of this resolvent also is of use in determining these fields. Explicitly determining this resolvent requires the costly inversion of a large dense matrix.

In this paper, we consider a lower complexity method which constructs an approximation to the resolvent. By linearizing the formula for the resolvent about the increment to the material properties computed in the previous least-squares iteration, we obtain an expression in which the new resolvent is equal to the old resolvent plus an increment due to the change in the material properties. Construction of this increment requires the multiplication of three matrices. To reduce the computational burden of this task, we compute only a subset of “significant” elements of this increment matrix where the notion of significance is related to the sensitivity of the cost function to small changes in each component of this matrix.

We explore the use of this method in the context of a non-destructive testing modality known as thermal wave diffraction tomography. Here, a laser is used to induce heat diffusion on the top of a small sample of material. Based on reflection or transmission measurements of the thermal field, one desires an image of the internal structure of the sample which makes evident holes or other defects. It can be shown that the thermal field obeys a scalar Helmholtz equation with a complex propagation vector so that the physics of the problem are quite similar to those of inverse electrical conductivity problems for which inverse scattering methods are commonly employed.

*The work of the first author was supported in part DOE contract DE-FC07-95ID13395, NSF Grant MIP-9623721, and by subcontract GC123920NDG from Boston University under the AFOSR MURI Program on Reduced Signature Target Recognition

Electromagnetic Modeling and Physics-based Processing Methods for Subsurface Object Characterization from Broadband Electromagnetic Induction Data

The use of broadband electromagnetic induction (BEMI) methods for the detection and characterization of buried objects has received considerable attention in recent years for applications ranging from environmental remediation to the detection and classification of buried landmines and unexploded ordinance. In this work, we describe an efficient modeling method and associated processing techniques for the extraction of target information from BEMI data. The utility of this approach is verified both on synthetic as well as experimental data collected from a fielded sensor.

We consider processing techniques based on an enhanced version of the inductive sensor model originally proposed by Das, McFee, Toews and Stuart (IEEE Trans. GRS, Vol. 28, No. 3, May 1990, pp. 278-287). The model conveniently encodes the quantities most relevant to the application at hand: a frequency dependent dipole scattering tensor used to characterize and classify the object, the spatial coordinates of the center of the target required for localization, and three Euler angles required to specify the orientation of the object. Moreover, under this model the elements of the dipole scattering tensor are independent of the orientation and location parameters and are in fact linearly related to the observed BEMI data.

Given this structure, we consider statistically optimal estimation methods for the determination of the dipole tensor, the object location, and orientation from noisy and cluttered BEMI data. The structure of the model leads to a computationally efficient, low dimensional, non-linear least squares algorithm. Specifically, we use the linearity of the model with respect to the dipole scattering tensor to eliminate this high dimensional set of unknowns from the optimization problem and concentrate on determining the location and orientation parameters: six in all. From these estimates, we can then trivially determine the optimal dipole structure.

A key element of our approach is the specification of a clutter model required for low SNR problems where the object contains little metal such as anti-personnel landmine remediation. In this paper, we develop a parametric stochastic model which captures the spatial correlation seen in typical data sets. The model is constructed to reflect the current methods in which BEMI data are collected. First, a dense set of spatial BEMI data in a known, object-free region is collected. These data along with a sparse collection of BEMI data taken in the immediate vicinity of a suspected target will be used to estimate the clutter model parameters and remove this noise source from the signal.

In the talk accompanying this abstract the utility of our processing approach and clutter model will be demonstrated using both synthetic and experimental BEMI data collected from GEM-3 sensor constructed Geophex Inc. The clutter model structure will be presented along with algorithms for estimating the parameters and performing clutter mitigation. The accuracy of the orientation and location estimates will be discussed. Finally, the use of the dipole moments for object classification will be demonstrated for both high and low metal content objects.

Session/topic: Subsurface object detection organized by Profs. Caorsi and Pichot

**Author information sheet for “Electromagnetic Modeling and Physics-based
Processing Methods for Subsurface Object Characterization from Broadband
Electromagnetic Induction Data”**

Mustafa Ozdemir (presenting author)
Dept. of Electrical and Computer Engineering
Northeastern University
Boston MA 02115
Tel: 617-373-5603
FAX: 617-373-8627
email: mozdemir@ece.neu.edu

Eric L. Miller (contact)
Dept. of Electrical and Computer Engineering
Northeastern University
Boston MA 02115
Tel: 617-373-8386
FAX: 617-373-8627
email: elmiller@ece.neu.edu

Alan Witten School of Geology and Geophysics
University of Oklahoma
100 East Boyd St., Suite 810
Norman, OK 73019
Tel: (405) 325-1563
FAX: (405) 325-3140
email: awitten@ou.edu

A Statistical Approach to Object Detection from Ground Penetrating Radar Arrays

The use of ground penetrating radar (GPR) arrays for detecting buried objects has received considerable attention in recent years in areas such as landmine and unexploded ordinance remediation, utility line mapping, and archaeology. A typical GPR configuration for such applications is comprised of one linear array of transmitters and a second array of receiving elements which are simultaneously moved down a linear track. At every stop of the system, each transmitter emits a short pulse of electromagnetic energy which interacts with the surrounding medium. Based on observations of scattered fields collected by the receivers the objective of the problem is to determine if an object is present in the field of view of the array.

From a signal processing perspective a number of factors make this problem quite challenging. First, the attenuation associated with the scattering of the GPR signal from the ground and target results in useful signal only over receivers located closest to the firing transmitter. In many functioning systems in fact only a single receiver is employed per transmitter. Thus, detection methods based on beamforming, MUSIC, or ESPRIT which require array-based observations cannot easily be employed. In essence then for GPR we have a multichannel (one channel per transmit/receiver pair), as opposed to an array, processing problem. Even if full array data were available, the assumptions underlying e.g. MUSIC (far field propagation, plane wave or at worst spherical wave structure) are violated. Indeed, to obtain a precise description of the received GPR signal one must resort to computationally intensive electromagnetic modeling methods to capture the complex physics associated with near field propagation and scattering of the transmitted GPR waveform off a buried object in the presence of a random rough air-ground interface. While such modeling can be done in principle, from a practical perspective such computationally intensive methods are not appropriate for use in a functioning GPR system.

Given these issues, here we view the problem in the framework of multichannel blind signal detection and employ statistical methods to process the GPR returns. This approach allows us to exploit two generic properties of the signal transmission process. First, for any given stop of the array, the presence of the mine close to a transmitter-receiver (T/R) pair results in a jump in the mean value of the observed signal relative to that seen in other pairs. Thus, we develop a widowed analysis of variance (ANOVA) test to detect this change. Second, physical principles dictate that as the array moves from one stop to the next, the magnitude of this jump first increases as the array approaches the object and then decreases as the sensors move past the location of the target. This behavior is exploited in the synthesis of a sequential detection strategy designed to recursively process the output of the ANOVA test as the GPR system moves down the track. The use of this sequential methods allows us to determine a priori important performance statistics for the GPR system, such as mean time between false alarms and average delay in declaring the presence of a target. In the talk accompanying this abstract, we provide a more detailed description and analysis of our algorithms as well as the physical models on which they are based. We will demonstrate the performance of this method both on synthetic data obtained from computational electromagnetic codes and on actual field data collected by a GPR system constructed by EG&G Inc.

Session/topic: Electromagnetic Signal Processing, Subsurface Object Detection, Remote Sensing

**Author information sheet for “A Statistical Approach to Object Detection from
Ground Penetrating Radar Arrays”**

Eric L. Miller (contact)
Dept. of Electrical and Computer Engineering
Northeastern University
Boston MA 02115
Tel: 617-373-8386
FAX: 617-373-8627
email: elmiller@ece.neu.edu

Xiaoyin Xu (presenting author)
Dept. of Electrical and Computer Engineering
Northeastern University
Boston MA 02115
Tel: 617-373-5603
FAX: 617-373-8627
email: xxu@ece.neu.edu

Multiscale, adaptive methods for reduced order inverse scattering

The objective of an electromagnetic inverse scattering problem (ISP) is to use observations of scattered wavefields obtained on the boundary of a medium to ascertain the spatial distribution of permittivity internal to the region of interest. Typically, the solution to an ISP is approached by formulating and then solving a high dimensional non-linear optimization problem to recover the permittivity values in a fine scale discretization of the underlying region. The optimization functional usually contains one term encouraging fidelity to the data and one or more regularizers required to combat the inherently ill-posed nature of the inverse problem. In addition to the high computational complexity associated with this technique, in previous work we shown that a uniform, fine scale discretization of the permittivity does not represent the best use of the data. More precisely, using information theoretic methods it is possible to show that the level of uncertainty in the reconstructed permittivity grows with the distance of the pixel from the sources and receivers. Motivated by this observation, here we consider an inversion approach designed to obtain a low order representation of the permittivity in which fine scale information is adaptively distributed in a more rational manner.

Our approach is based on the use of a two dimensional frame of quadratic B-splines to represent the unknown permittivity. This family of functions has the convenient property that coarse scale members can be defined as a superposition of finer scale B-splines. We use this nesting property in the synthesis of a multiscale inversion routine. We begin by representing the permittivity as a superposition of a few, coarse scale spline functions. After determining the optimal expansion coefficients, the remainder of the procedure is a loop in which we alternate between adding degrees of freedom to improve the reconstruction, and removing unnecessary detail to control the complexity of the estimate.

We have found that this approach to inversion is useful in a number of ways. First, by adaptively placing fine scale information only where it is really required, the dimensionality of the problem is kept low (hundreds rather than thousands of unknowns) thereby leading to a significant reduction in computational complexity relative to a single, fine grid approach. Second, a common problem with all non-linear ISPs is convergence of the problem to a local minimum of the cost function. The multiscale nature of our approach can in fact lead to convergence to a lower cost minimum than would otherwise be the case. Finally, we can use the same information theoretic techniques previously mentioned to quantify the reduction in uncertainty associated with our lower order method.

The talk accompanying this abstract will provide a more detailed description and analysis of our approach. We will demonstrate its utility using as an example a cross-borehole radar tomography problem motivated by applications in environmental remediation and monitoring. As an example of the typical performance of our approach for this application, the multiscale algorithm adaptively converges to an estimate of the profile containing on the order of 300 B-spline functions when the underlying discretization contains 4096 pixels representing a saving of about 92%. Despite the smaller number of unknowns, the low order reconstruction is of comparable quality to its fine scale counterpart.

Session/topic: Inverse Scattering, Electromagnetic Signal Processing

**Author information sheet for “Multiscale, adaptive methods for reduced order
inverse scattering”**

Eric L. Miller (contact and presenting author)
Dept.of Electrical and Computer Engineering
Northeastern University
Boston MA 02115
Tel: 617-373-8386
FAX: 617-373-8627
email: elmiller@ece.neu.edu

A Unified Statistically-Based Approach to the Modeling and Processing of EMI Time and Frequency Domain Sensor Data

In the past decade, significant effort has been devoted to the processing of electromagnetic induction (EMI) data for purposes of buried object detection and characterization. Current EMI sensors operate either in the frequency domain using a stepped frequency approach to probe the subsurface or directly in the time domain. Time domain systems generally measure the transient response of the earth and any embedded objects to pulsed input. The data collected by such sensors take the form either of samples of the time response or integrals of the response over a collection of time windows. In this paper, we describe a unified approach to both the modeling and the processing of data from these apparently dissimilar sensors.

One popular approach to EMI sensor modeling is to describe the response of the sensor in terms of a superposition of decaying exponentials in time, or equivalently, one pole transfer functions in frequency. The decay rates/pole positions can then be used to characterize the object. While useful, this model fails to explicitly take into account spatial information which may be obtained as the sensor collects data over a region above and surrounding the object. To overcome this shortcoming, we propose a hybrid model. The spatial structure is captured using the technique in Das *et. al.*, (*IEEE. Trans. Geoscience and Remote Sensing*, **28**, 278-287, May, 1990) while the temporal/Fourier content of the data are represented using the exponential/single pole structure. The result is a closed form model for the spatial-temporal and spatial-Fourier response of an EMI sensor parameterized directly in terms of a small number of quantities, nine, which directly describe the location, orientation, and class of the object under investigation.

Using this model, we will present a statistical signal processing framework for object characterization and sensor optimization. First, the Maximum likelihood estimator for the parameter vector will be derived for swept frequency, sampled time, and integrated time sensors. We will describe the implementation of the non-linear search routine required to determine these estimates. A Bayesian classifier will be employed to map the ML parameter estimates to object class. Processing results will be provided using both synthetic data as well as real sensor data. Finally, using the asymptotic properties of our estimator and standard decision-theoretic methods, we will discuss how the overall modeling/processing approach can be of use to optimize sensor design. For example, our approach provides a mechanism for determining the “best” frequencies or time gates to discern one object from another.

Efficient T-Matrix Methods for GPR Forward Modeling

Adnan Şahin¹ Eric L. Miller

Center for Electromagnetic Research, 235 Forsyth Building
Northeastern University, Boston, MA 02115

Telephone: (617) 373-8386

Telefax : (617) 373-8627

email: adnan@cdsp.neu.edu

An important component associated with the GPR problem is the computation of the scattered fields produced by buried objects when illuminated by the radar source. The choice of technique for computing these fields is often driven by a variety of factors including computational complexity and the flexibility to handle easily a wide range of configurations of scatterers. Based on the fact that objects of interest for many GPR applications have relatively simple shapes such as cylinders and spheres, it is apparent that transition matrix (T-matrix) approach is well suited to the GPR forward problem.

The T-matrix technique is based on the expansion of the fields and the Green's function in spherical (3D) or cylindrical (2D) harmonic series. Considerable work has focused on the generation of efficient recursive T-matrix techniques for computing the scattered field from a single, electrically large, dielectric object. The approach taken in these applications is to tessellate the object into many smaller scatterers. Thus, even though the number of scatterers is large, the order of harmonics used to expand the basis functions is small thereby keeping the computational load relatively small. However, if one wants to find the scattered field from multiple, electrically large, tessellated objects, the number of scatterers grows quite quickly thereby increasing the complexity of the problem considerably. A natural alternative is to treat each object as a single scatterer (i.e. bypass the tessellation) but increase the order of the harmonic expansions for each object. Unfortunately, this approach is hindered by difficulties related to the slow convergence of the harmonic expansions.

Here, we present a modified, stable recursive T-matrix algorithm to calculate the scattered field from a heterogeneous collection of spatially separated objects based on these higher order expansions. Motivated by the GPR problem, we consider the calculation of the scattered field produced by a group of parallel cylinders buried in a lossy medium. In the most general case, these objects can be dielectric or metallic, electrically small (fraction of λ) or large ($1-2 \lambda$). The modification, which will be described in the talk accompanying this abstract, circumvents the slow convergence of the series by replacing the product of two translation matrices with product of three such matrices two of which yield the identity.

In the talk we will also discuss the verification of this algorithm against previously published results for scattered fields from multiple cylinders, compare the computational advantages of this higher order technique relative to the tessellated approach, and provide examples demonstrating the ability of this code to compute fields for GPR-relevant scattering scenarios such as mine detection and hazardous waste removal.

¹The work of this author was supported in part DOE contract DE-FC07-95ID13395, NSF Grant MIP-9623721, and by subcontract GC123920NDG from Boston University under the AFOSR MURI Program on Reduced Signature Target Recognition

Simultaneous multiple regularization parameter selection by means of the L-hypersurface with applications to linear inverse problems posed in the wavelet transform domain

M. Belge, M. Kilmer and E. L. Miller

235 Forsyth Building
Northeastern University
Boston, MA 02115

ABSTRACT

In this paper, we introduce the L-hypersurface method for use in linear inverse problems. The new method is intended to select multiple regularization parameters simultaneously. It is a multidimensional extension of classical L-curve method and hence does not require any specific knowledge about the noise level or signal semi-norm. We give examples of the L-hypersurface method applied to the linear inverse problems posed in the wavelet domain and evaluate the performance of the new method on a signal restoration experiment.

Keywords: Inverse problems, wavelets, regularization parameters

1. INTRODUCTION

The term inverse problems refers to the estimation of an unknown quantity, called an object, from its noise corrupted functionals. Such problems arise in a variety of fields including image processing, medical imaging and geophysical prospecting. For example, a common image processing problem which we explore later in this paper is that of deconvolution where one observes a blurred and noise corrupted version of the original image and seeks to recover the original. In this work, we consider the following discrete linear inverse problem

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n} \quad (1)$$

where \mathbf{H} is a known matrix representing the degradation, \mathbf{g} is the vector holding the data, \mathbf{f} is the desired object to be estimated and \mathbf{n} denotes zero mean white Gaussian noise. The problem of estimating \mathbf{f} from \mathbf{g} is said to be ill-posed if the operator \mathbf{H} is not invertible or its inverse is unstable.^{1,2} In such a case direct inversion is not advisable and a unique and stable estimate $\hat{\mathbf{f}}$ is sought by incorporating some *a priori* information about unknown image \mathbf{f} . Such a technique is called regularization. The prior information is usually specified in the form of a smoothness constraint which is just some (semi)norm of \mathbf{f} . The regularized solution is found by minimizing the following expression with respect to \mathbf{f}

$$J(\mathbf{f}, \boldsymbol{\alpha}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \Phi(\boldsymbol{\alpha}, \mathbf{f}) \quad (2)$$

where Φ expresses the prior for \mathbf{f} and is dependent on a set of parameters $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_M]^T$ (called regularization or smoothing parameters). A typical example is $\Phi(\boldsymbol{\alpha}, \mathbf{f}) = \alpha \|\mathbf{L}\mathbf{f}\|_2^2$ with \mathbf{L} being a regularization operator (usually a differential operator). In this case, α governs the tradeoff between the fidelity to the data term and smoothness of the solution term. Once the form of prior information (form of Φ in (2)) has been determined, the overall quality of the estimate is affected only by the choice of regularization parameters and therefore the choice of regularization parameters $\boldsymbol{\alpha}$ is a crucial part of the inversion process.

Other author information:

M.B.: E-mail: belge@cdsp.neu.edu; WWW: <http://www.cdsp.neu.edu/info/students/belge/>

M.K.: E-mail: mkilmer@cdsp.neu.edu

E.L.M.(correspondence): E-mail: elmiller@cdsp.neu.edu; WWW: <http://claudius.cdsp.neu.edu/elhome/>; Telephone: (617) 373-8386;

FAX: (617) 373-8627

The problem of choosing a suitable regularization parameter which would yield an estimate as close to the original object as possible has received considerable attention in the past and is still an open problem. There are methods which rely on prior information such as a bound on the (semi)norm of the signal or the noise level. A well known example is the Morozov's discrepancy principle.² On the other hand, methods such as Generalized Cross Validation (GCV)³ and the L-curve^{4,5} do not require any side information. We note that all of the methods mentioned above deal with choosing a *single* parameter. As more sophisticated inversion algorithms are developed, the need for using multiple regularization parameters is becoming increasingly important. For example, in the solution of image restoration problems posed in the wavelet domain one needs multiple regularization parameters to capture the scale or orientation dependent structure of the underlying image.⁶ Another example is the inverse problem of electrocardiography⁷ where one needs both temporal and spatial regularization constraints to obtain a useful solution.

Extension of methods developed for a single parameter to choose multiple regularization parameters is often a non-trivial task. In this paper, we introduce the L-hypersurface method for use in linear inverse problems which require multiple regularization parameters. The L-hypersurface is based on the classical L-curve method. It is an M dimensional function of the regularization parameters and provides significant information about the likelihood of a particular parameter set being optimum. In the next section, we provide some background information on the L-curve method and then introduce the L-hypersurface method in Sect. 3. In Sect. 4, we present a brief introduction to the solution of discrete linear inverse problems posed in the wavelet domain. In Sect. 5 we present result obtained by applying the L-hypersurface to the linear inverse problems posed in the wavelet domain and then conclude the paper in Sect. 6.

2. THE L-CURVE METHOD

A convenient tool for choosing a single regularization parameter which does not require any side information is the L-curve method popularized by Hansen^{4,5} *et. al.*. The L-curve is simply a logarithmic plot of residual norm (the first term on the r.h.s. of (2)) versus the log of the reconstruction (semi)norm (second term on the r.h.s. of (2)) for a set of admissible regularization parameters. In this way, the L-curve displays the compromise between the minimization of these two quantities. It has been argued and numerically shown that the so called "corner" of the L-curve, defined as the point with maximum curvature, corresponds to a point where regularization and perturbation errors are balanced.⁴

In Fig. 1 we plotted a typical L-curve along with its curvature and the error between the original and restored signals for a range of regularization parameters. The experiment for which the L-curve was computed was the reconstruction of a 1-D signal using Tikhonov regularization with identity.¹ The circle indicates the corner of the L-curve defined to be the point with maximum curvature. The part of the L-curve to the left of the corner contains a region where the regularization parameter is getting smaller and the error between the original and reconstructed signals is dominated by the perturbation errors. The solution (semi)norm is very sensitive to small changes in the regularization parameter indicating a noisy solution. On the other hand, the part of the L-curve to the right of the corner is a region where the regularization parameter is gradually increasing and the residual norm is the most sensitive to the changes in the regularization parameter. In this region, the restored signal is excessively smooth. These two regions are clearly separated by the corner point, hence the corner corresponds to a point where regularization and perturbation errors are approximately balanced. Supporting this observation, the corner of the L-curve in this example indicates a point which is very close to the point where the error between the original and restored signals is a minimum (bottom figure in Fig. 1).

3. L-HYPERSURFACE: A NEW METHOD FOR SIMULTANEOUS DETERMINATION OF MULTIPLE REGULARIZATION PARAMETERS

As mentioned in the introduction, in order to obtain a useful solution for the inverse problem of electrocardiography,⁷ one needs both temporal and spatial regularization constraints. Facing the problem of choosing two or more parameters simultaneously Brooks⁷ *et. al.* tried to extend the idea of L-curve by drawing the residual norm against two side constraint norms. They named the resulting plot the "L-surface" (L-hypersurface). Although, they enjoyed some success in choosing the regularization parameters, it was apparent that the interpretation of such a surface was very difficult.

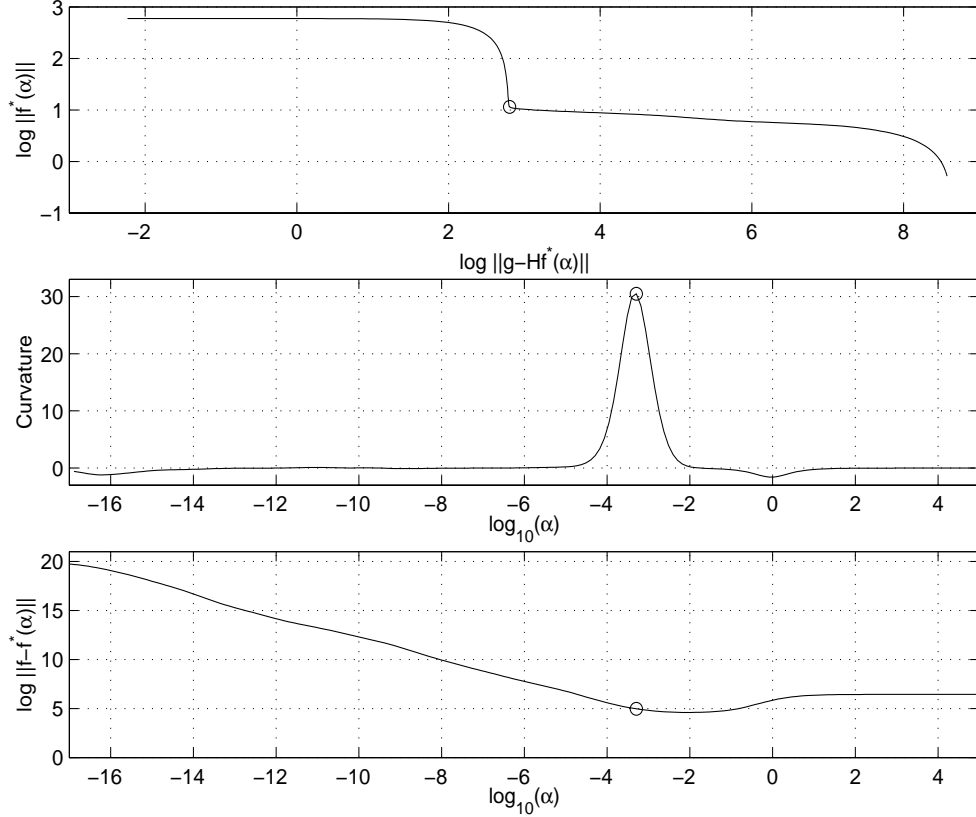


Figure 1. A typical L-curve, its curvature and error between original and restored signals. Circle indicates the point with maximum curvature. $\mathbf{f}^*(\alpha)$ is the estimate of \mathbf{f} using α as the regularization parameter.

Considering these difficulties, we propose using the *Gaussian curvature* of the L-hypersurface as a means to choose multiple regularization parameters. To clarify this idea, consider the following inversion scheme:

$$\mathbf{f}^*(\alpha) = \arg \min_{\mathbf{f}} \left\{ \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \sum_{i=1}^M \alpha_i \|\mathbf{R}_i \mathbf{f}\|_p^p \right\}, \quad (3)$$

where $\|\mathbf{R}_i \mathbf{f}\|_p^p$, $1 \leq i \leq M$, are the constraints on the solution (semi)norm and α_i are the corresponding regularization parameters and $1 \leq p \leq 2$. The cost function in (3) represents a multiply constrained least squares problem and includes many popular image restoration schemes as its special cases. For example, by taking $\mathbf{R}_i = \mathbf{D}^{(i)}$ (i. e. the i th order differentiation) and $p = 2$ we obtain the classical Sobolev regularization. The wavelet domain inversion algorithm introduced by the authors is obtained if the problem is posed in the wavelet domain and the \mathbf{R}_i are the operators extracting the desired portions of the wavelet transform of the object \mathbf{f} .⁶ For example, we can take \mathbf{R}_1 as the operator extracting the scaling coefficients and \mathbf{R}_2 as the operator extracting the wavelet coefficients for a doubly constrained problem.

To construct the L-hypersurface, we first introduce the following quantities

$$\mathbf{f}^*(\alpha) = \arg \min_{\mathbf{f}} J(\mathbf{f}, \alpha) \quad (4)$$

$$z(\alpha) = \log \|\mathbf{g} - \mathbf{H}\mathbf{f}^*(\alpha)\|_2^2 \quad (5)$$

$$x_j(\alpha) = \log \|\mathbf{R}_j \mathbf{f}^*(\alpha)\|_p^p, \quad j = 1, \dots, M \quad (6)$$

With the above definitions, the L-hypersurface is defined as a subset of \mathcal{R}^{M+1} associated with the map $\mathcal{S}(\alpha) : U \rightarrow$

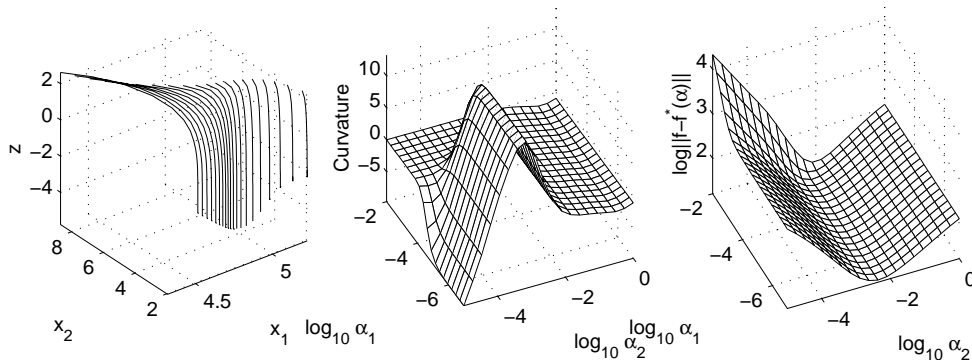


Figure 2. A typical L-hypersurface, its curvature and the error between the original and restored signals.

\mathcal{R}^{M+1} , $\alpha \in U$, $U = \bigcup_{j=1}^M [a_j, b_j]$ such that

$$\mathcal{S}(\alpha) = (x_1(\alpha), \dots, x_M(\alpha), z(\alpha)). \quad (7)$$

In simple terms the L-hypersurface is a plot of the residual norm as a function of the constraint norms drawn in log scale for a range of regularization parameters.

If the surface \mathcal{S} defined above is smooth with no self intersections and has a definite tangent plane at all points it is said to be regular.⁸ A regular surface possesses an interesting geometrical quantity called Gaussian curvature which measures how much the surface is warped at a point. A plane has 0 Gaussian curvature while a sphere has constant Gaussian curvature at all points. The Gaussian curvature reduces to the regular curvature for a 1-D curve.

Instead of considering \mathcal{S} , we consider its Gaussian curvature field as a function of regularization parameters. This approach is motivated by the analogous results for the L-curve method. If we consider the L-curve and its curvature in Fig 1., we observe that the curvature plot provides the location of turning points which are practically the only interesting points on the L-curve. Given the curvature plot we can easily determine the correct regularization parameter. Motivated by this fact, we will omit displaying the L-hypersurface itself. The reason is that the interpretation of the curvature plot is much easier than the L-hypersurface and also the curvature plot provides an automatic way of locating the regularization parameters if we define the L-hypersurface selection of the regularization parameters as the point with the maximum curvature on the Gaussian curvature plot. This is more easily understood by looking at a typical L-hypersurface, as displayed in Fig. 2. In the leftmost plot in Fig. 2, we display the L-hypersurface for a least squares problem with first and second order derivatives of the object as constraints. The middle and the rightmost plots are the curvature of the L-hypersurface and the norm of the error between the original and estimated objects. We observe that the points on the curvature plot where the curvature achieves a local maximum seems to track the local minimum of the estimation error surface. Note that, locating the correct regularization parameters by just examining the shape of the L-hypersurface is more difficult.

By using elementary calculus, one can show that the L-hypersurface for the model problem in (3) is regular. The Gaussian curvature of S can be easily computed given the first and second order partial derivatives of $z(\alpha)$ with respect to $x_i(\alpha)$, $1 \leq i \leq M$, and is given by the following expression⁹

$$\kappa(\alpha) = \frac{(-1)^M}{w^{M+1}} |\mathbf{P}| \quad (8)$$

where $w^2 = 1 + \sum_{i=1}^M (\frac{\partial z}{\partial x_i})^2$, $\mathbf{P}_{i,j} = \frac{\partial^2 z}{\partial x_i \partial x_j}$ and derivatives are evaluated at $q = (x_1(\alpha), \dots, x_M(\alpha), z(\alpha))$. In the next section, we present the application of the L-hypersurface method to the discrete linear inverse problems posed in the wavelet domain.

4. WAVELET DOMAIN ADAPTIVE EDGE-PRESERVING IMAGE RESTORATION

Recently, there has been much interest in the wavelet domain formulation and solution of linear inverse problems, especially in the area of image processing. Part of the reason is the collection of mathematical results indicating that

wavelets are especially useful for representing functions that contain discontinuities. The wavelet transform¹⁰ of an object produces a set of scaling coefficients and a set of wavelet coefficients. The scaling coefficients are obtained by first filtering the object by a low pass filter and then down sampling the resulting signal. The wavelet coefficients are obtained by first filtering the object by a high pass filter and then down sampling. Wavelet coefficients obtained in this way are close to zero at regions where the object is smooth and significant only at regions where there is a high intensity change. On the other hand, the scaling coefficients appear to be a low resolution representation of the original object.

The wavelet domain formulation of a linear inverse problem presents an interesting challenge since the scale-space structure of the object may be highly irregular. A direct result of this phenomena is that one typically needs more than one regularization parameter to adapt to the changes in the scale-space structure of the object. We can represent the problem in (1) in the wavelet domain as⁶

$$\begin{aligned}\mathcal{W}_d \mathbf{g} &= (\mathcal{W}_d \mathbf{H} \mathcal{W}_h^T) \mathcal{W}_h \mathbf{f} + \mathcal{W}_d \mathbf{n} \\ \hat{\mathbf{g}} &= \hat{\mathbf{H}} \hat{\mathbf{f}} + \hat{\mathbf{u}},\end{aligned}\tag{9}$$

where \mathcal{W}_d and \mathcal{W}_h are wavelet transform matrices with possibly different scaling functions, $\hat{\mathbf{g}}$, $\hat{\mathbf{f}}$ and $\hat{\mathbf{n}}$ are the vectors holding the scaling and wavelet coefficients of the data, the original image and the disturbance, \mathbf{H} is the wavelet domain representation of our linear degradation operator \mathbf{H} and $\mathcal{W}_h^T \mathcal{W}_h = I$ follows from the orthogonality of the wavelet transform. Note that, since wavelet transform is orthogonal $\hat{\mathbf{u}}$ is again Gaussian with zero mean and variance σ^2 .

In previous work,⁶ the authors introduced a wavelet domain image restoration scheme where an estimate of \mathbf{f} is found by minimizing the following cost function

$$J(\hat{\mathbf{f}}, \alpha) = \|\hat{\mathbf{g}} - \hat{\mathbf{H}} \hat{\mathbf{f}}\|_2^2 + \sum_j \alpha_j \|\hat{\mathbf{R}}_j \hat{\mathbf{f}}\|_p^p,\tag{10}$$

where $\hat{\mathbf{R}}_j$ are the operators extracting the desired portions of the wavelet transform (such as scaling coefficients) and $1 \leq p \leq 2$. The l_p norm of the wavelet coefficients are used as the regularization function since the recent results in the area of image coding and denoising suggest that the distribution of wavelet coefficients are notably non-Gaussian hence using $p < 2$ is more appropriate.¹¹

There are a variety of methods for structuring the regularizer through the choice of the matrices $\hat{\mathbf{R}}_j$ in (10). For example, we may employ a Besov regularization scheme for which the fundamental assumption is that the regularization parameter increases exponentially with scale as we go to finer scales in the wavelet domain. That is,

$$\alpha_j = \alpha_1 2^{-\lambda(j-j_0)}, \quad j_0 \leq j \leq J-1,\tag{11}$$

where j_0 is the lowest (i.e. coarsest) and $J-1$ the highest scales of interest. In this case we need to determine two regularization parameters, α_0 and α_1 for the scaling and coarsest scale wavelet coefficients respectively (assuming that λ and p are pre-specified). Correspondingly, we have two $\hat{\mathbf{R}}_j$ matrices: $\hat{\mathbf{R}}_1 \mathbf{f}$ is the vector containing scaling coefficients and $\hat{\mathbf{R}}_2 \mathbf{f}$ is the vector of wavelet coefficients. On the other hand, a level dependent regularization scheme can be developed. In this case we need to choose $M+1$ parameters simultaneously and we have $M+1$ $\hat{\mathbf{R}}_j$ matrices: $\hat{\mathbf{R}}_1 \mathbf{f}$ is the vector of scaling coefficients and each of the vectors $\hat{\mathbf{R}}_j \mathbf{f}$, $2 \leq j \leq M+1$ contains the wavelet coefficients at scale j . We note that an efficient optimization method exists to compute the estimate $\hat{\mathbf{f}}^*$ once the regularization parameters are determined.⁶

Both of the above cases can be handled by the means of the appropriate L-hypersurfaces. For the implementation, we only need a set of derivatives as given in (8). Formulas for these quantities are provided in appendices.

5. SIMULATION STUDY

In this section, the performance of the L-hypersurface method for choosing two regularization parameters, $\alpha = [\alpha_1, \alpha_2]$, will be measured by the classical efficiency E_L ¹²

$$E_L = \frac{\min_{\alpha} \|\mathbf{f}^*(\alpha) - \mathbf{f}\|_2}{\|\mathbf{f}^*(\alpha_L) - \mathbf{f}\|_2},\tag{12}$$

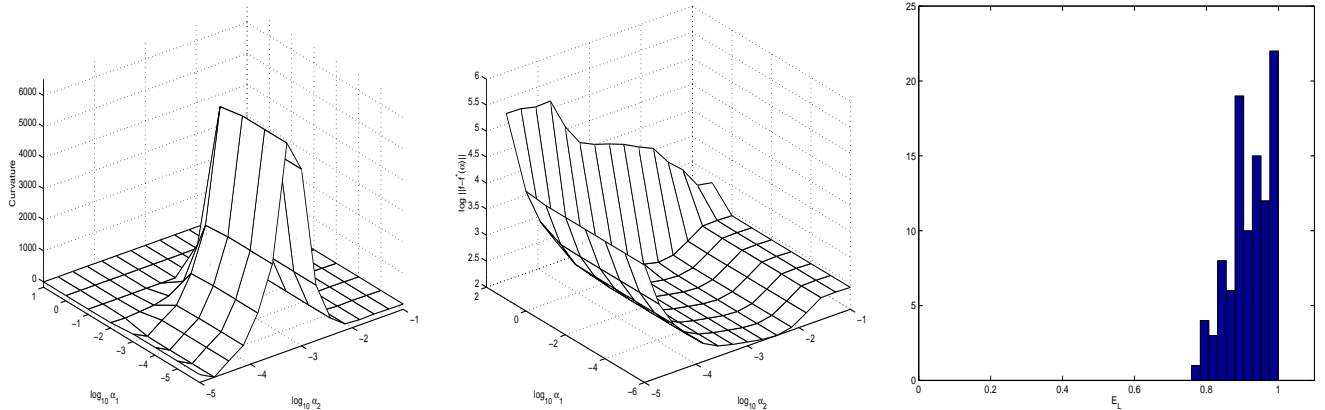


Figure 3. From left to right: A typical L-hypersurface for the model experiment, the error between original and restored signals and the histogram of E_L values over 100 experiments.

where α_L is obtained by the L-hypersurface method and the vector of optimal regularization parameters α are those minimizing the error between the original and the estimated objects. A value close to one obtained for E_L indicates a performance close to optimal while a value for E_L close to zero indicates bad performance. The global performance of the L-hypersurface method was evaluated by observing the values of E_L obtained on 100 experiments which differed only by simulated noise. These 100 efficiency values were then used to create a histogram by partitioning the range $[0, 1]$ into ten intervals of equal length. We believe that such a histogram is a good summary of the performance of the L-surface method.

The signal to be reconstructed is the 256 point “Blocks” sequence extracted from Donoho’s Wavelab software package.¹³ It was degraded by a Gaussian convolutional kernel with $\sigma = 2.0$. Zero mean white Gaussian noise was added to obtain a SNR of 30dB. The inversion scenario in (10) was implemented with \mathbf{R}_1 being the operator extracting the scaling coefficients and \mathbf{R}_2 being the operator extracting the wavelet coefficients. We used $p = 1.0$ norm and set $\lambda = 1.0$. The signal was decomposed by a 5-level wavelet decomposition with Haar wavelets. For the calculation of Gaussian curvature, we used the derivatives obtained analytically.

Fig. 3 summarizes the results obtained in this experiment. The top leftmost plot shows the Gaussian curvature plot for a typical experiment. In the middle, we have the surface obtained by plotting the norm of the error between original and reconstructed signals. There is an extended ridge along the α_1 direction in this plot. This ridge seems to track the points where the error between the original and restored signals is locally minimum. Such behavior here is not specific to this experiment but rather general. We performed numerous experiments by changing the signal, degradation or noise level and obtained exactly the same results.

In the rightmost figures, we plot the histograms of E_L values obtained on 100 experiments. From this plot, we observe that the regularization parameters obtained by the L-hypersurface method produce restorations which are almost as good as the ones obtained by the optimal method.

Figure 4 shows an example of the L-hypersurface method applied to an image restoration experiment. As seen in Fig 5, the original Mandrill image was blurred by a Gaussian kernel of width 2 pixels and then zero mean white Gaussian noise was added to the blurred image to set the SNR at 30dB. We restored the degraded image by our multiscale image restoration scheme with $p = 1.0$ norm. However, in this case we utilized a 3-level wavelet decomposition and assigned a different regularization parameter for each scale in the wavelet domain. For scaling coefficients we used $\alpha_0 = 10^{-5}$ as the regularization parameter. Figure 4 shows the L-hypersurface obtained for this experiment. The L-hypersurface is a function of three regularization parameters. Therefore, each figure in Figure 4 shows a slice of the L-hypersurface with the regularization parameter corresponding to the lowest scale being constant. This example shows that, the L-hypersurface tracks well the minima in the RMSE ($RMSE = \sqrt{\frac{1}{N} \|\mathbf{f} - \mathbf{f}^*(\alpha)\|_2^2}$, N =number of pixels in the image \mathbf{f}) surface along which the restoration error is close to being minimum. Higher dimensional L-hypersurfaces can be obtained by just augmenting the parameter set.

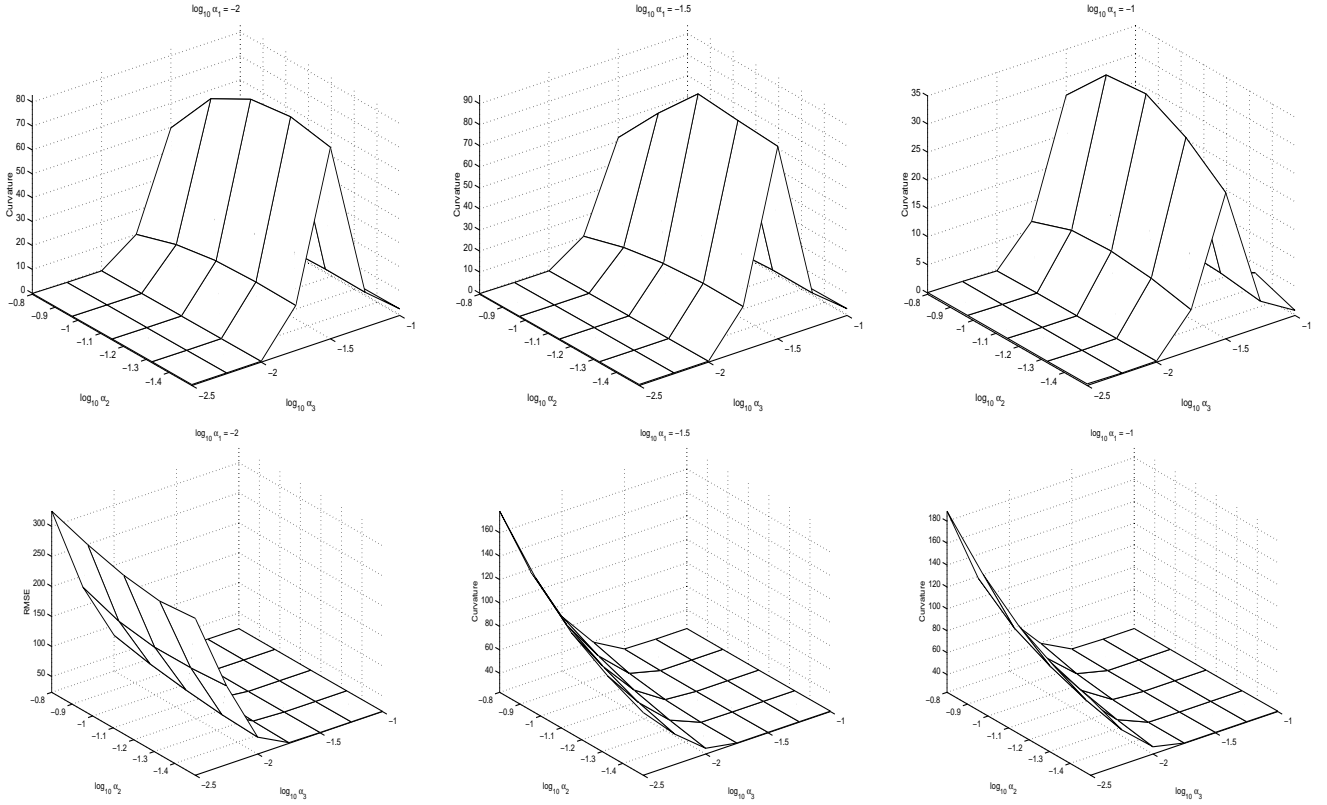


Figure 4. L-hypersurface cuts and corresponding RMSE plots for an image restoration experiment.

6. CONCLUSIONS AND DISCUSSION

In this paper, we presented a new method for choosing multiple regularization parameters simultaneously without using any side information. The method is called the L-hypersurface and is based on the popular L-curve method. Instead of directly looking at the surface obtained by plotting the residual norm against the side constraint norms, we look at its Gaussian curvature. From numerical examples, it was seen that the points where Gaussian curvature reaches a local maxima appear to be closely tied to the local minima of the mean square error surface. We choose the regularization parameters as those maximizing the Gaussian curvature of the L-hypersurface. Monte Carlo simulations show that this selection method is indeed very good and produces results comparable to the optimal method.

In the context of image restoration, multiple parameter regularization problems arise quite naturally. For example, a simple restoration strategy which minimizes the norm of the gradient of an image as the side constraint could in fact use two parameters corresponding to the gradients in the horizontal and vertical directions because those might be quite different (for example the image under consideration might be a layered in horizontal direction while being smooth in vertical direction). Another example would be a wavelet transform domain restoration algorithm where each subband is regularized differently as described in Sect. 4. We hope that the L-hypersurface method will be actually useful in these cases.

As the future work, we will consider the computational efficient implementations of the L-hypersurface and investigate the theoretical properties of the L-hypersurface.



Figure 5. From left to right: Original, blurred and the restored MANDRILL images. The regularization parameters were obtained by the L-hypersurface method.

APPENDIX A. FORMULAS FOR THE IMPLEMENTATION OF THE L-HYPERSURFACE FOR THE INVERSE PROBLEMS POSED IN THE WAVELET DOMAIN

A.1. Characterization of Solutions

In this section, we present the necessary formulas for the implementation of the L-hypersurface method for wavelet domain multiscale object restoration algorithm (10). The formulae are given for the case of a 1-D signal but extension to 2-D signals is straightforward. We begin with defining the solution $\mathbf{f}^*(\boldsymbol{\alpha})$.

$$\mathbf{U} = \text{diag} \left[\frac{\alpha_i}{\left(|\hat{f}_i^*|^2 + \beta \right)^{1-p/2}} \right], \quad \alpha_i = \alpha_j \text{ if } i \in I_j \quad (13)$$

$$\hat{\mathbf{H}}^T \left(\hat{\mathbf{H}}\mathbf{f}^* - \hat{\mathbf{g}} \right) + \frac{p}{2} \mathbf{U}\mathbf{f}^* = 0, \quad (14)$$

where I_j is the index set for the wavelet coefficients at scale j , \hat{f}_i^* is the i th element of the vector \mathbf{f}^* and $\beta > 0$ is a small stabilization constant. Note that, the equations (14) do not represent an exact solution to (10). The l_p norm term at the right hand side of (10) has been slightly perturbed to avoid indifferentiability of the cost function. Equation (14) give the conditions that must be satisfied by \mathbf{f}^* . An iterative algorithm based on (14) can be developed to approximate the solution \mathbf{f}^* .⁶

A.2. Formulas for the L-hypersurface

By (8), the only quantities that we need to compute the Gaussian curvature of the L-hypersurface at a point $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_M)$ are the first and second order partial derivatives of the residual norm $z(\boldsymbol{\alpha})$ with respect to M constraint norms $x_i(\boldsymbol{\alpha})$, $1 \leq i \leq M$. Since $z(\boldsymbol{\alpha})$ is not explicitly defined in terms of the constraint norms, we obtain the necessary derivatives by implicit differentiation. We first define the following quantities:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial x_1}{\partial \alpha_1} & \cdots & \frac{\partial x_M}{\partial \alpha_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial \alpha_M} & \cdots & \frac{\partial x_M}{\partial \alpha_M} \end{bmatrix}, \quad \mathbf{D}_{\boldsymbol{\alpha}}^2 = \begin{bmatrix} \frac{\partial^2}{\partial \alpha_1^2} & \cdots & \frac{\partial}{\partial \alpha_1 \partial \alpha_M} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial \alpha_M \partial \alpha_1} & \cdots & \frac{\partial^2}{\partial \alpha_M^2} \end{bmatrix}, \quad \mathbf{d}_{\boldsymbol{\alpha}} = \begin{bmatrix} \frac{\partial}{\partial \alpha_1} \\ \vdots \\ \frac{\partial}{\partial \alpha_M} \end{bmatrix}. \quad (15)$$

Note that, $\mathbf{d}_{\boldsymbol{\alpha}}$ and $\mathbf{D}_{\boldsymbol{\alpha}}^2$ are operators which represent first and second order differentiation with respect to the variables $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_M]^T$. By the definitions above, the partial differentials of z with respect to \mathbf{x} are given by:

$$\frac{\partial z}{\partial x_i} = \mathbf{d}_{\boldsymbol{\alpha}}^T \mathbf{J}^{-1} \mathbf{e}_i$$

$$\frac{\partial^2 z}{\partial x_i \partial x_j} = \mathbf{e}_j^T \mathbf{J}^{-T} \mathbf{D}_{\alpha}^2 \mathbf{J}^{-1} \mathbf{e}_i - \mathbf{d}_{\alpha}^T \mathbf{J}^{-1} \begin{bmatrix} \mathbf{e}_j^T \mathbf{J}^{-T} \mathbf{D}_{\alpha}^2 x_1 \\ \vdots \\ \mathbf{e}_j^T \mathbf{J}^{-T} \mathbf{D}_{\alpha}^2 x_M \end{bmatrix} \mathbf{J}^{-1} \mathbf{e}_i, \quad (16)$$

where \mathbf{e}_i denotes the unit vector with all zero entries except for the i th one and $1 \leq i, j \leq M$. Remembering the formulas $\mathbf{P}_{i,j} = \frac{\partial^2 z}{\partial x_i \partial x_j}$ and $w^2 = 1 + \sum_{i=1}^M (\frac{\partial z}{\partial x_i})^2$ we realize that the curvature at the point α can be readily computed by just plugging in the values $\frac{\partial z}{\partial \alpha_i}$, $\frac{\partial^2 z}{\partial \alpha_i \partial \alpha_j}$, $\frac{\partial x_m}{\partial \alpha_i}$, $\frac{\partial^2 x_m}{\partial \alpha_i \partial \alpha_j}$ into the equations in (16). These partial derivatives can be calculated in a straightforward but tedious way from (14).

REFERENCES

1. A. Tikhonov and V. Arsenin, *Solution of Ill-Posed Problems*, Wiley, New York, 1977.
2. V. A. Morozov, *Methods for solving incorrectly posed problems*, Springer-Verlag, New York, 1984.
3. P. Craven and G. Wahba, "Smoothing noisy data with spline functions," *Numer. Math.* **48**, pp. 671–698, 1986.
4. P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Review* **34**, pp. 561–580, 1992.
5. P. C. Hansen, "The use of l-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.* **14**, pp. 1487–1503, 1993.
6. M. Belge, E. L. Miller, and M. Kilmer, "Wavelet domain image restoration with adaptive edge-preserving regularization," *submitted to IEEE Trans. Img. Proc.* .
7. G. F. Ahmad, D. Brooks, G. Maratos, and R. MacLeod, "Augmented inverse problem of electrocardiography: combined time and space regularization," *Proc. 16th Ann. Int. Conf. IEEE EMBS* **v. 15**, pp. 773–774, 1993.
8. M. P. DoCarmo, *Differential Geometry of Curves and Surfaces*, Prentice Hall, New Jersey, 1976.
9. H. Flanders, *Differential Forms with Applications to the Physical Sciences*, Dover, New York, 1989.
10. I. Daubechies, *Ten Lectures on Wavelets*, SIAM Press, New York, 1992.
11. E. P. Simoncelli and E. H. Adelson, "Noise removal via bayesian wavelet coring," in *ICIP'96, Proceedings IEEE Int. Conf. on Image Processing* **3**, pp. 185–188, 1996.
12. L. Desbat and D. Girard, "Regularization parameters: some more efficient methods and their application to deconvolution problems," *SIAM J. Sci. Comput.* **16**, pp. 1387–1403, 1995.
13. J. Buckheit and D. Donoho, "Wavelab and reproducible research," in *Wavelets and statistics*, A. Antoniadis, ed., pp. 53–71, Springer-Verlag, 1995.

High resolution processing algorithms for near field object detection: Performance bounds and sensitivity analyses¹

Adnan Şahin Eric L. Miller

Center for Electromagnetic Research, 235 Forsyth Building

Northeastern University, Boston, MA 02115

Telephone: (617) 373-8386

Telefax : (617) 373-8627

email: adnan@cdsp.neu.edu

A common problem in many application areas is the detection and localization of targets with known structure based on observations of scattered electromagnetic or acoustic fields. In recent years, there has been considerable work in methods for solving such problems in a manner which bypasses the need to solve a large, ill-posed inverse scattering problem. These processing methods extract from the data a small number of geometric parameters describing the target distribution. In previous work, we have considered algorithms of this type for locating targets in the nearfield of a linear receiver array. Initial results based on the use of high resolution array processing techniques indicate that these methods provide highly accurate localization of arbitrary collection of buried metallic and dielectric objects.

In this paper, we examine issues of performance and sensitivity analysis associated with these processing methods. Because our techniques are statistical in nature, we are able to develop explicit expressions for the error variance (or bounds on this quantity) which provide hard limits on the best performance achievable from the use of these array processing techniques. We show that these bounds are easily and elegantly obtained when one employs a T-matrix model to describe the scattering processes.

A key assumption underlying the success thus far of our methods is that we know the number of targets for which we are looking as well as their material properties. In this paper, we explore the sensitivity of our localization techniques to show how our detection schemes behave when the actual parameters are different from the ones used in the algorithm. For example, we determine the degradation in performance when the actual shape of the target differs from the one the algorithm assumes. Of particular interest are sensitivity to changes in target shape and material properties. We establish analytical bounds for sensitivity against changes in these parameters, and verify these bounds with Monte-Carlo runs.

In addition, we determine "detectability" of targets for a given geometry. For example, we find the minimum radii of objects for a certain soil type or minimum distance between two objects to achieve a particular detection probability for a fixed false alarm rate. In the talk, we will show examples of sensitivity analyses and detectability problems for multiple mine-like and drum-like objects.

¹This work was supported by the US Army Research Office MURI grant #DAAG55-97-1-0013 and the US Dep. of Energy Grant DE-FG07-97ID13566

High resolution processing algorithms for near field object detection: Performance bounds and sensitivity analyses¹

Adnan Şahin Eric L. Miller

Center for Electromagnetic Research, 235 Forsyth Building

Northeastern University, Boston, MA 02115

Telephone: (617) 373-8386

Telefax : (617) 373-8627

email: adnan@cdsp.neu.edu

A common problem in many application areas is the detection and localization of targets with known structure based on observations of scattered electromagnetic or acoustic fields. In recent years, there has been considerable work in methods for solving such problems in a manner which bypasses the need to solve a large, ill-posed inverse scattering problem. These processing methods extract from the data a small number of geometric parameters describing the target distribution. In previous work, we have considered algorithms of this type for locating targets in the nearfield of a linear receiver array. Initial results based on the use of high resolution array processing techniques indicate that these methods provide highly accurate localization of arbitrary collection of buried metallic and dielectric objects.

In this paper, we examine issues of performance and sensitivity analysis associated with these processing methods. Because our techniques are statistical in nature, we are able to develop explicit expressions for the error variance (or bounds on this quantity) which provide hard limits on the best performance achievable from the use of these array processing techniques. We show that these bounds are easily and elegantly obtained when one employs a T-matrix model to describe the scattering processes.

A key assumption underlying the success thus far of our methods is that we know the number of targets for which we are looking as well as their material properties. In this paper, we explore the sensitivity of our localization techniques to show how our detection schemes behave when the actual parameters are different from the ones used in the algorithm. For example, we determine the degradation in performance when the actual shape of the target differs from the one the algorithm assumes. Of particular interest are sensitivity to changes in target shape and material properties. We establish analytical bounds for sensitivity against changes in these parameters, and verify these bounds with Monte-Carlo runs.

In addition, we determine "detectability" of targets for a given geometry. For example, we find the minimum radii of objects for a certain soil type or minimum distance between two objects to achieve a particular detection probability for a fixed false alarm rate. In the talk, we will show examples of sensitivity analyses and detectability problems for multiple mine-like and drum-like objects.

¹This work was supported by the US Army Research Office MURI grant #DAAG55-97-1-0013 and the US Dep. of Energy Grant DE-FG07-97ID13566

IGARSS 2000 ABSTRACT SUBMISSION TEMPLATE

CORRESPONDING AUTHOR: Eric L. Miller

AFFILIATION: Northeastern University

ADDRESS: 235 Forsyth Building

CITY: Boston

STATE: MA

POSTAL CODE: 02115

COUNTRY: USA

TELEPHONE: (617)373-8386

FACSIMILE: (617)373-8627

EMAIL: elmiller@ece.neu.edu

ABSTRACT TITLE:

Detection and Localization of Buried Mines from GPR Array Measurement: A Statistical Approach

AUTHOR(S):

List names only...if subsequent authors have different mailing addresses, please use a second sheet to provide data

Xiaoyin Xu, Eric L. Miller, Gary Sower, and Tom Broach

ABSTRACT TEXT:

The use of ground penetrating radar (GPR) for mine detection is complicated by clutter associated with a rough air-ground interface and sensor noise. One way of improving detector performance is to employ array measurements as a means of providing data diversity which can increase detection probabilities and decrease false alarm rates. Here we develop a decision-theoretic, sequential detection algorithm for processing such GPR data. Our approach takes into account the interface, the nature of the mine signature seen in the data, and most importantly, signal diversity from array measurements. It is a low complexity technique specifically designed to optimally process the data as they are collected by the sensor and is thus applicable to real-time applications. The algorithm consists of a three step procedure. First, a windowed analysis of variance (ANOVA) procedure is used to process the array-data at each stop of the system. Second, a sequential, generalized likelihood ratio test is developed to recursively take the ANOVA output and generate a final test statistic as the array proceeds down the track. Finally, comparing this statistic to a threshold at each stop of the array indicates if and when a detection is declared. Field data from EG&G is used to demonstrate the validity of our algorithm. Experimental results and theoretical analysis show that the algorithm exhibits robust performance and a high detection probability for a given false alarm rate.

Text block boundaries are fixed. Abstract cannot exceed boundaries.

TOPIC PREFERENCE: A.4

Image Restoration with $1/f$ -Type Fractal Models and Statistical Estimation of the Model Parameters*

R. M. Dufour and E. L. Miller

Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115

rdufour@cdsp.neu.edu, elmiller@cdsp.neu.edu

Abstract

A common problem in image processing is recovery of an image given noisy linear functionals of the original. While it has been shown that in certain situations, models possessing a $1/f$ -type power spectrum perform well as regularizers to stabilize these ill-posed inverse problems, the optimal parameters for the model are rarely known a priori. Previously, it was demonstrated that the expectation maximization (EM) algorithm can satisfactorily perform the estimation of the model parameters in the unblurred one-dimensional case. In this paper, we extend this analysis to the situation of two-dimensional objects and an environment which includes blurring. We show that again the EM algorithm performs well. In addition, we examine performance in terms of the variance of the estimates and bounds on these quantities.

1. Introduction

A common image processing problem is the reconstruction of an image given data which is a noisy and blurred representation of the original. Examples of such problems are found in application areas ranging from medical imaging and geophysical exploration to non-destructive testing and radar signal processing. Here we consider a discrete version of such linear inverse problems which takes the mathematical form

$$\mathbf{y} = \mathbf{T}\mathbf{g} + \mathbf{n} \quad (1)$$

where \mathbf{g} is a vector of coefficients representing the image to be recovered, \mathbf{T} is the blurring kernel, \mathbf{n} is

additive noise and \mathbf{y} is the data. For simplicity, the noise vector, \mathbf{n} , is taken to be zero mean, white, and Gaussian (i.e. $\mathbf{n} \sim (0, \sigma^2 \mathbf{I})$). We assume that \mathbf{g} is a zero mean, Gaussian vector with an unknown covariance matrix, $\mathbf{P}_{\mathbf{g}}$. As we discuss below, this model for \mathbf{g} constitutes a prior distribution and serves to aid in the regularization of the recovery of \mathbf{g} from \mathbf{y} .

For most practical problems, \mathbf{T} is highly ill-conditioned so that even small perturbations in the data cause large high frequency artifacts in the reconstruction. To counteract the ill-conditioning, some form of regularization is required. Generally, the regularization imposes a low-pass or smoothness constraint upon the solution. In particular, we consider a regularized least squares estimator which defines the reconstructed image, $\hat{\mathbf{g}}$, as the solution to the following optimization problem:

$$\hat{\mathbf{g}} = \arg \min_{\mathbf{g}} \|\mathbf{y} - \mathbf{T}\mathbf{g}\|_{\mathbf{R}^{-1}}^2 + \|\mathbf{g}\|_{\mathbf{P}_0^{-1}}^2, \quad (2)$$

where $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}$, the matrix $\mathbf{R} = \sigma^2 \mathbf{I}$ is the noise covariance, and \mathbf{P}_0^{-1} is used to stabilize the solution typically by limiting the amount of high frequency information in $\hat{\mathbf{g}}$.

It can be shown that the minimum mean square error (MSE) of this estimation scheme is achieved when the regularization matrix \mathbf{P}_0 is equal to the covariance matrix of \mathbf{g} . As we assume that $\mathbf{P}_{\mathbf{g}}$ is not known *a priori*, it is necessary to estimate simultaneously both the matrix as well as \mathbf{g} . In this work, we assume the availability of only a single data vector \mathbf{y} so that sample covariance matrix estimation schemes are inapplicable. Rather, we employ a low-order, parametric model for $\mathbf{P}_{\mathbf{g}}$ such that this matrix can be completely described by a small vector of hyperparameters.

To describe \mathbf{g} we employ a family of models which have been shown to accurately represent stochastic processes with $1/f$ -type power spectra. These

*This work was supported by a grant from Textron Systems Division, NSF CAREER Grant MIP-9623721, and a sub-contract GC123920NDG from Boston University under the AFOSR MURI Program on Reduced Signature Target Recognition.

model have been developed and used in characterizing signals such as Brownian motion and ocean height data [1]. Also, they have been shown to work well as regularizers for least squares-type reconstruction schemes like (2) primarily because the $1/f$ structure essentially imposes a lowpass filtering operation to the solution, penalizing against high frequency artifacts. These models are most naturally described in the wavelet transform domain where the covariance matrix is diagonal (i.e. the wavelet coefficients are uncorrelated for these processes) and the values of these diagonal elements are specified in terms of a vector of three hyperparameters [2]. For this model, it was shown by Wornell [3] that in the non-degraded (i.e. $\mathbf{T} = \mathbf{I}$) one dimensional case, an efficient estimation routine can be arrived at using the EM algorithm for jointly determining both this vector as well as \mathbf{g} itself. In this paper we extend the work to both the blurred case and to two dimensional image data.

The wavelet coefficients for \mathbf{g} will be specified as $\gamma_{\mathbf{k}}^{\mathbf{s}}$, where the two element vector \mathbf{s} is the scale indices in the horizontal and vertical directions, with smaller indices corresponding to coarser scales, and \mathbf{k} is the shift indices. The $1/f$ prior model is now defined in the transform domain as

$$\text{var}(\gamma_{\mathbf{k}}^{\mathbf{s}}) = \kappa 2^{-\alpha^T \mathbf{s}} = \kappa 2^{-\alpha_x s_x - \alpha_y s_y}, \quad (3)$$

where α and κ are the hyperparameters which define the model. The vector α contains the fractal parameters which specify the smoothness of the process in both the horizontal and vertical directions, and κ sets the overall power of the process. Upon selecting a unique ordering for the two scale and shift indices, we can reduce the notation to a single index as

$$\text{var}(\gamma_j) = \kappa 2^{-\alpha^T \mathbf{s}_j} = \kappa 2^{-\alpha_x s_j^x - \alpha_y s_j^y}. \quad (4)$$

These variances then form the diagonal elements the covariance matrix which we write as

$$\mathbf{P}_0 = \kappa \mathbf{F}(\alpha). \quad (5)$$

where \mathbf{F} is a diagonal matrix whose j th element is $2^{-\alpha_x s_j^x - \alpha_y s_j^y}$. To simplify notation, the dependence of \mathbf{F} upon α will be often assumed but not written.

The objective of this paper is the development of an efficient wavelet domain estimation method for determining both the hyperparameters of the model and the wavelet coefficients of the object. This is performed using an expectation-maximization algorithm.

2. EM Algorithm

Normally the large size of the matrices arising in image restoration problems would make an EM approach highly computationally complex. By working in the wavelet transform domain however, we are able to take advantage of the fact that many classes of both convolutional and non-convolution blurring matrices have very sparse transform domain representations [4] thereby lowering the cost associated with the EM algorithm. Significant additional savings are obtained through the use of the diagonal $1/f$ regularization matrix. Specifically, it was demonstrated in [3] that for the unblurred, one-dimensional case, operating in the wavelet transform domain and using the $1/f$ models allows the maximization step to be reduced to a fast root finding algorithm. In this paper and the accompanying talk, we extend this technique to the case of blurred data and to two dimensional image data. We demonstrate that the root finding algorithm is applicable to the solution of multiple fractal parameters by successively iterating between them. The complexity of the expectation step is reduced by applying a recently developed algorithm which simultaneously provides for (a) rapidly inverting the information matrix associated with (2) and (b) preconditioning the conjugate gradient algorithm used to estimate γ for a given set of hyperparameters [5].

The log-likelihood function for α and κ , derived from the multivariate Gaussian, is

$$\begin{aligned} l(\alpha, \kappa; \boldsymbol{\eta}) &= -\frac{1}{2} \boldsymbol{\eta}^T (\kappa \boldsymbol{\Theta} \mathbf{F} \boldsymbol{\Theta}^T + \mathbf{R})^{-1} \boldsymbol{\eta} \\ &\quad - \frac{1}{2} \log |(\kappa \boldsymbol{\Theta} \mathbf{F} \boldsymbol{\Theta}^T + \mathbf{R})|. \end{aligned} \quad (6)$$

Here, the entire problem is transformed to the wavelet transform domain. The matrix $\boldsymbol{\Theta}$ represents the transform domain version of \mathbf{T} while $\boldsymbol{\eta}$ contains the transform of \mathbf{y} . Finally, the noise covariance matrix is \mathbf{R} . The Expectation step of the EM algorithm for the Gaussian case consists of constructing the estimate of the error covariance matrix for the current estimate of the parameter set, and computing the MAP estimate of $\hat{\gamma}$ as

$$\mathbf{C}^{(p)} = \left(\boldsymbol{\Theta}^T \mathbf{R}^{-1} \boldsymbol{\Theta} + \frac{1}{\kappa^{(p)}} (\mathbf{F}^{(p)})^{-1} \right)^{-1} \quad (7)$$

$$\hat{\gamma}^{(p)} = \mathbf{C}^{(p)} \boldsymbol{\Theta}^T \mathbf{R}^{-1} \boldsymbol{\eta}. \quad (8)$$

For convenience, we will define an additional vector, \mathbf{t} , which has the elements

$$t_j^{(p)} = \hat{\gamma}_j^{(p)2} + \mathbf{c}_j^{(p)}. \quad (9)$$

The maximization step consists of maximizing the following function

$$\begin{aligned} \begin{bmatrix} \boldsymbol{\alpha}^{(p+1)} \\ \kappa^{(p+1)} \end{bmatrix} &= \arg \max_{\boldsymbol{\alpha}, \kappa} \left(-N \log \kappa - \log |\mathbf{F}| \right. \\ &\quad \left. - \frac{1}{\kappa} \sum_j (\mathbf{F}^{-1} \mathbf{t}^{(p)})_j \right). \end{aligned} \quad (10)$$

Taking the derivative with respect to κ and setting it equal to zero gives the update solutions for κ as

$$\kappa^{(p+1)} = \frac{1}{N} \sum_j ((\mathbf{F}^{(p+1)})^{-1} \mathbf{t}^{(p)})_j. \quad (11)$$

To solve for $\boldsymbol{\alpha}$ we will use the relation that

$$\frac{\partial \mathbf{F}}{\partial \alpha_x} = -(\log 2) \mathbf{S}_x \mathbf{F} \quad (12)$$

where \mathbf{S}_x is a diagonal matrix of the scale coefficients associated with α_x . The derivative with respect to α_y is identical, thus we will show only the x results. Differentiating (11) and setting the result equal to 0 gives

$$0 = -(\log 2) \sum_j (\mathbf{F}^{-1} (\mathbf{S}_x - \frac{\text{tr} \mathbf{S}_x}{N} \mathbf{I}) \mathbf{t}^{(p)})_j. \quad (13)$$

Since \mathbf{F} , \mathbf{S}_x , and \mathbf{I} are diagonal matrices, the equation can be rewritten as

$$0 = \sum_j 2^{\alpha_x s_j^x} 2^{\alpha_y s_j^y} (s_j^x - \frac{\text{tr} \mathbf{S}_x}{N}) (t_j^{(p)}) = P(2^{\alpha_x}). \quad (14)$$

Which is a polynomial in 2^{α_x} with one sign change. The maximization with respect to α_x is therefore computed as the unique real positive root of this polynomial. Likewise, the counterpart equation in α_y is

$$0 = \sum_j 2^{\alpha_y s_j^y} 2^{\alpha_x s_j^x} (s_j^y - \frac{\text{tr} \mathbf{S}_y}{N}) (t_j^{(p)}) = P(2^{\alpha_y}). \quad (15)$$

It can be shown that (10) with respect to $\boldsymbol{\alpha}$ is a concave surface on the positive values and has a unique maximum. Further, solving (14) and (15) always increases the value of (6). Iterating between the equations thus continually moves towards the maximum until it converge to this maximum point. This is therefore the update solution for $\boldsymbol{\alpha}$.

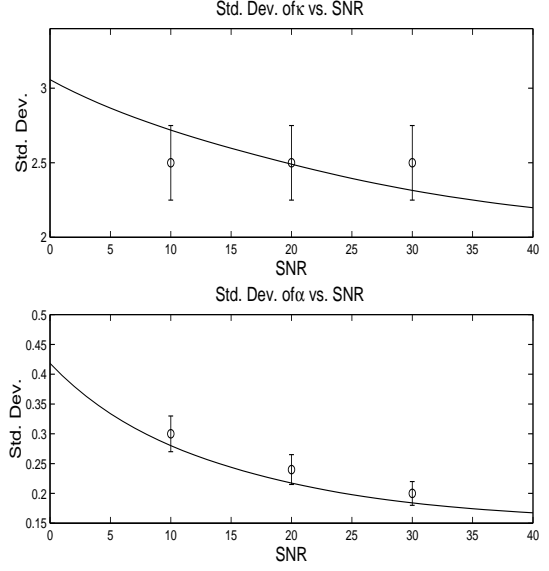


Figure 1: Standard deviation of the parameter estimates.

The entire algorithm is now as follows:

1. Provide an initialization of the model parameters: κ and $\boldsymbol{\alpha}$.
2. Calculate the expectations using these parameters in equations (7), (8) and (9).
3. Solve for the updated $\boldsymbol{\alpha}$ by iterating on (14) and (15) until they converge.
4. Solve for κ using (11).
5. Repeat steps 2 thru 4 until $\|\hat{\boldsymbol{\gamma}}^{(p+1)} - \hat{\boldsymbol{\gamma}}^{(p)}\| < \epsilon$.

3. Estimation Variances

The performance bounds on the estimation of the model parameters is given by the Cramer-Rao bounds. The bounds are computed as the diagonal elements of the inverse of the Fisher information matrix which is given as [6]

$$\mathbf{J} = \begin{bmatrix} -E \frac{\partial^2 l}{\partial \kappa^2} & -E \frac{\partial^2 l}{\partial \kappa \partial \alpha_x} & -E \frac{\partial^2 l}{\partial \kappa \partial \alpha_y} \\ -E \frac{\partial^2 l}{\partial \kappa \partial \alpha_x} & -E \frac{\partial^2 l}{\partial \alpha_x^2} & -E \frac{\partial^2 l}{\partial \alpha_x \partial \alpha_y} \\ -E \frac{\partial^2 l}{\partial \kappa \partial \alpha_y} & -E \frac{\partial^2 l}{\partial \alpha_x \partial \alpha_y} & -E \frac{\partial^2 l}{\partial \alpha_y^2} \end{bmatrix}. \quad (16)$$

The inverse of this matrix provides the bound on the error covariance matrix (ECM). Extensive analysis of the parameter estimation performance can be achieved using these formulas. We will present here an examination of the performance versus SNR.

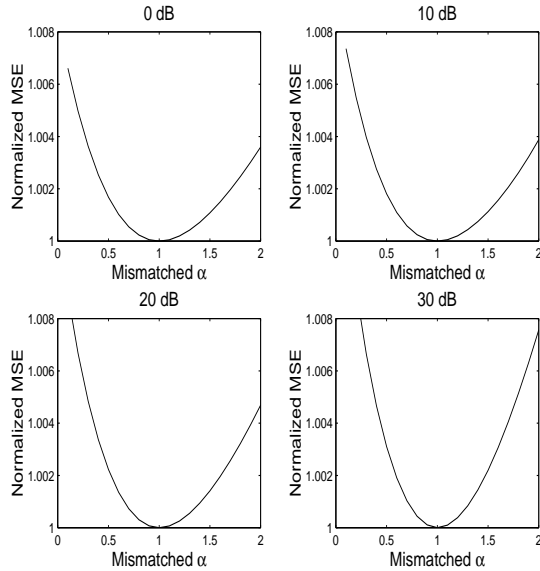


Figure 2: Normalized MSE in estimate of γ for a mismatched model. Here model energy is kept constant to provide a constant level of regularization.

Fig. 1 shows the standard deviations in the parameter estimates as a function of the SNR. The solid line is the CRB and is calculated from the inverse of (16). The results of Monte Carlo simulations are also shown to confirm the analysis. The true model parameters for this case are a κ of 10, and α_x and α_y of 1. We can see here that, as expected, the parameter estimates, specifically of α , improve as the SNR increases.

Next, we define the Normalized Mean Square Error (NMSE) as the ratio of the MSE in the estimate of \mathbf{g} obtained using arbitrary values for the hyperparameters in \mathbf{P}_0 to that of an estimator which knows the true values. It is easily shown that the NMSE is always greater than one and equals one exactly when the true model parameter values are used. In Fig. 2, the NMSE for this situation is shown where α_x is mismatched. The sensitivity to the mismatched parameter increases as SNR increases; however, the results of Fig. 1 indicate that the large SNR case is exactly where we can achieve accurate model estimates. Conversely, at low SNRs, the NMSE plots of Fig. 2 show a only a moderate degradation from optimal performance. Thus, we conclude from Figs. 1 and 2 that the EM algorithm provides accuracy exactly where such performance is needed in the recovery of \mathbf{g} and the effects of degraded model estimates at low SNR do not translate into large relative losses in our ability to recover \mathbf{g} .

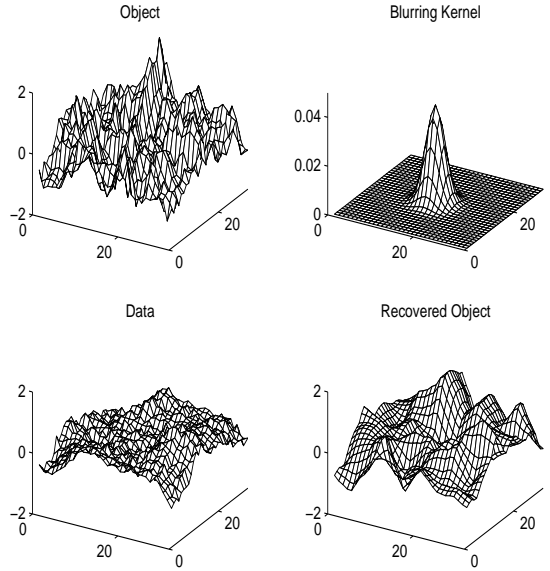


Figure 3: The reconstruction of an image. The blurring function is a Gaussian kernel of $\sigma = 2$ and the SNR is 10.8 dB.

Similar analysis can be performed versus the degree of blurring imposed by \mathbf{T} . Again in this situation, we find that the parameter estimation performs well in precisely the same situations in which object reconstruction is sensitive to a closely matched model. Where the parameter estimation is poor, model mismatch is not as much of a concern.

4. Example

In Figs. 3 and 4, we show the quality of the reconstructions available with the $1/f$ model and the convergence properties of the EM algorithm. The upper left plot is of a 32×32 pixel image of a fractal with $\alpha_x = 1.5$, $\alpha_y = 1.9$ and $\kappa = 10$. The image is blurred with a Gaussian kernel of $\sigma = 2$ which has significant energy content in approximately a 8×8 pixel area, and then corrupted by additive noise. The signal to noise ratio is 10.8 dB. The recovery is shown after 6 iterations of the EM algorithm. The EM algorithm moves quickly to the vicinity of the minimum of the cost function and further iteration does not significantly lower the cost function nor effect the reconstruction. In the case of this experiment, the algorithm converged to hyperparameters of $\hat{\alpha}_x = 1.2$, $\hat{\alpha}_y = 2.4$, and $\hat{\kappa} = 6.8$.

The lower plot in Figure 4 shows the error between the reconstruction and the original at each iteration

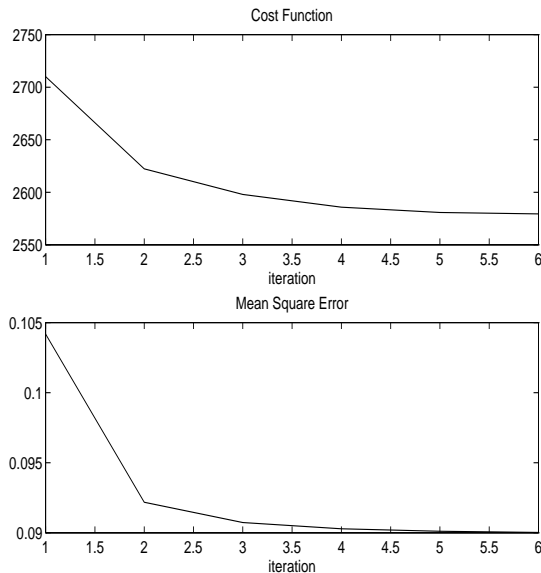


Figure 4: The convergence of the cost function and the error between the object and the reconstruction.

of the algorithm. As was shown in [7] the recovery is robust to small mismatches in the fractal parameters α (and to mismatches up to an order of magnitude upon the regularization level κ), therefore the discrepancy between the algorithm's hyperparameters and the true hyperparameters does not translate to a large error in the reconstruction. The MSE obtainable with the matched model was .086. The coarse scale features of the object (i.e. the large peaks) have been recovered quite well both in terms of their locations and amplitudes.

5. Conclusions and Future Work

Previous work has shown that the $1/f$ models produce satisfactory results for estimating one dimensional signals in noise and that the structure of the models in the wavelet transform domain allows for accurate and efficient estimation of the model parameters. In this paper, we extended this analysis to two-dimensional image problems and an environment which includes blurrings. By revising the expectation step to operate upon a statistic calculated from the estimates of the object and the error covariance matrix, we again achieve the computational simplicity of the unblurred case. Further, we show that multiple fractal parameters can be estimated by iterating upon the polynomials. Lastly we showed a method to evaluate the performance of the parameter estimates

and their effect upon the estimates of the object.

The algorithm presented assumes a Gaussian distribution to the object and noise and a cost function that minimizes the mean squared error. However the structure of the model should be exploitable with other distributions and cost functions associated with other criteria such as edge preservation [8]. By pursuing a similar approach to that presented here, one could develop efficient EM algorithms for a variety of problems.

References

- [1] P.W. Fieguth, A.S. Willsky, and W.C. Karl. Multiresolution stochastic imaging of satellite oceanographic altimeter data. In *IEEE International Conference on Image Processing*, volume 2, 1994.
- [2] G.W. Wornell. A karhunen-loeve-like expansion for $1/f$ processes via wavelets. *IEEE Transactions on Information Theory*, 36:859–61, July 1990.
- [3] G.W. Wornell and A.V. Oppenheim. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on Signal Processing*, 40:611–23, March 1992.
- [4] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms, i. *Commun. on Pure Appl. Math.*, 44:141–183, 1991.
- [5] Eric L. Miller. Efficient statistical inversion in the wavelet transform domain. Technical Report TR-CDSP-97-41, Northeastern University, Boston, MA, January 1997.
- [6] Louis L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley, Reading, MA, 1991.
- [7] R.M. Dufour and E.L. Miller. Statistical estimation of $1/f$ -type prior models: Robustness to mismatch and efficient model determination. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1996.
- [8] M. Belge, E. L. Miller, and M. Kilmer. Wavelet domain image restoration with adaptive edge-preserving regularization. *In preparation for submission to IEEE Trans. Img. Proc.*

Electromagnetic Inversion for Multi-Bistatic Ground Penetrating Radar

Peter M. Johansen¹, Carey M. Rappaport, Anthony J. Devaney, and Eric L. Miller
Center for Electromagnetics Research
Forsyth Building
Northeastern University, Boston, MA 02115

In recent years there has been an increasing interest in use of ground penetrating radar (GPR) for detection of buried land mines or other objects located in the ground. To achieve an accurate image it is important that the inversion algorithm be based on an accurate forward model relating the buried objects to the measured electric field. Such model must at least include a model of the antennas and take into account the presence of the air/soil interface.

In this talk we present an electromagnetic inversion scheme for GPR imaging based on a forward model in which both the antennas and the air/soil interface are explicitly described. The presented forward model is an extension of the model in [1] where the air/soil interface is neglected. Consequently, the present model more accurately describes the physical GPR-configuration, and a better reconstructed image is expected in practice.

The GPR is assumed to be operated in a stepped-frequency mode and the data are collected in a multi-bistatic survey (multiple independent transmitter and receiver antennas). The forward model is based upon the electric dyadic Green's function for a lossy dielectric half space. This dyadic Green's function is constructed from the Green's function for free space using the Fresnel reflection and transmission coefficients. The antennas are modeled by applying Kern's scattering matrix formulation [1]. Also, the dispersion of the soil is accounted for by including a realistic soil model. To analytically calculate the pseudo inverse, the forward model is linearized using the Born approximation. Various regularization schemes are employed to stabilize the analytic inversion. Numerical examples are provided to illustrate the impact of the air/soil interface and the differences among the various regularization schemes.

- [1] R. W. Deming and A. J. Devaney, "Diffraction tomography for multi-monostatic ground penetrating radar imaging," *Inverse Problems*, vol. 13, no. 1, pp. 29–45, Feb. 1997.

¹Peter M. Johansen is a visiting Post Doc. from the Department of Electromagnetic Systems at the Technical University of Denmark. The Danish Technical Research Council is acknowledged for supporting the work of this author. This work is partially supported by the Army Research Office MURI grant # DAAG55-97-1-0013.

On the detection of buried objects from inductive arrays

Eric L. Miller
Dept. of Electrical
and Computer Engineering
Northeastern University,
Boston, MA 02215
Telephone: (617) 373-8386
Telefax : (617) 373-8627
email: adnan@cdsp.neu.edu

W. Clem Karl
Dept. of Electrical Engineering
Boston University
Boston, MA 02115

In recent years there has been considerable work in the development of inversion methods which, rather than forming an image of a medium, extract directly from the data information relevant to the distribution of desired targets located in the region under investigation. One such application of these methods is the detection and localization of buried landmines from low frequency electromagnetic induction (EMI) measurements. In this work we consider the development of inversion techniques adapted to the processing of data from an EMI array which surveys a region in a progressive manner starting at a known position and moving in a straight line to a final location. The use of space diversity coupled with advanced signal processing techniques will be shown to improve detection/false alarm performance and localization accuracy relative to a more typical mono-static EMI system.

We begin by developing a forward model relating the observed data directly to the pertinent physical characteristics of the targets of interest including size, shape, orientation, and electromagnetic properties. Our model is obtained using a reciprocity argument and the full three dimensional Maxwell's equations. For simplicity, we restrict our attention to a linearized model obtained under the first Born approximation. The final measurement model expresses the data at the receiver array as a linear combination of physically derived target signature vectors. The vectors carry information about the geometric structure of the buried objects. The weights represent the relative contrast of the objects and the number of vectors is equal to the number of mines in the receiver field of view.

This forward model lends itself particularly well to the analysis of the mine detection/localization problem and to the synthesis of adaptive algorithms for identifying buried mines. We will show how the above model can be used to provide quantitative results on the manner in which detection and false alarm rates and localization accuracy is dependent on the size, orientation, and position of the mines relative to the sensor array. We demonstrate how this analysis and the scattering model can be used as the basis for the synthesis of algorithms for performing mine detection and localization. Because the array is sensitive to mines in a full 3D volume, we can obtain detection and localization information before the array passes directly over or adjacent to the mine. We will discuss algorithms which provide this forward looking information in conjunction with accuracy estimates both of which are progressively refined and updated as the EMI array moves toward the buried object.

Efficient methods for the solution and analysis of statistical linear inverse problems in the wavelet transform domain*

Eric L. Miller

Department of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115
email: elmiller@cdsp.neu.edu

March 10, 1997

Abstract

We present efficient, wavelet domain algorithms for computing the error variances and pre-conditioning the normal equations for a large class of linear inverse problems. An iterative method based on the optimization of a diagonal dominance criteria is introduced to determine an optimal ordering of the rows and columns of the so-called information matrix. This ordering, combined with the judicious use of the block matrix inversion formula allows for the easy extraction of the error variances in a manner which does not require the full inversion of the information matrix. We demonstrate that this algorithm leads naturally to the specification of a left-preconditioner for a conjugate gradient solution to the normal equations associated with the MAP problem.

1 Introduction

In this paper, we consider computational issues associated with the use of maximum *a posteriori* (MAP) estimation methods for linear restoration (i.e. inverse) problems posed in the wavelet transform domain. An MAP approach provides a framework for performance analysis via the examination of second-order error statistics. Such information can be especially valuable for quantitatively understanding how issues such as the structure of space varying blurring operators or sparse and irregular sampling patterns impact the accuracy of the reconstruction. Additionally, wavelet-based methods are useful for a number of reasons. Large classes of both convolutional and non-convolutional blurring operators have sparse matrix representations in the wavelet transform domain thereby leading to efficient inversion methods based on conjugate-gradient type algorithms. A second advantage of a multiscale approach is the availability of a useful class of so-called $1/f$ statistical models/regularizers specified directly in the transform domain. Finally, wavelet-domain error variance analysis

have proved useful for rigorously capturing the nature of the ill-posedness in these inverse problems [1–3].

An important difficulty with such variance analysis is computational complexity. The error variances are the diagonal elements of the error covariance matrix (ECM). The quantities available from the problem statement allow only for the direct specification of the *inverse* of the ECM, called the information matrix. Thus, obtaining the error-variances requires the inversion of the information matrix a task requiring a costly order N^3 operations with N the number of samples in the signal or pixels in an image. In this work, we develop and demonstrate a wavelet-domain algorithm which exploits both the sparsity of the operator matrices and the covariance structure of the $1/f$ -type prior models to lower significantly this computational burden for a wide class of linear inverse problems.

2 Problem Formulation

We consider inverse problems of the form

$$\mathbf{y} = \mathbf{T}\mathbf{g} + \mathbf{w} \quad (1)$$

where \mathbf{g} is a lexicographically ordered vector of pixel values of the image to be recovered, \mathbf{w} is additive white Gaussian noise (i.e. $\mathbf{w} \sim \mathcal{N}(0, r\mathbf{I})$), \mathbf{y} is the data vector and \mathbf{T} is a matrix representing the blurring operation. As is discussed in [2], transformation of (1) to a multiscale domain is achieved by defining two orthonormal wavelet transform matrices, \mathbf{W}_y and \mathbf{W}_g , and applying them to (1) as

$$\mathbf{W}_y \mathbf{y} = [\mathbf{W}_y \mathbf{T} \mathbf{W}_g^T] [\mathbf{W}_g \mathbf{g}] + \mathbf{W}_y \mathbf{w}$$

or equivalently

$$\boldsymbol{\eta} = \boldsymbol{\Theta} \boldsymbol{\gamma} + \boldsymbol{\nu}. \quad (2)$$

In (2), \mathbf{W}_g takes \mathbf{g} into a vector $\boldsymbol{\gamma}$ comprised of all the wavelet coefficients and coarsest scale scaling coefficients in a two dimensional wavelet transform of \mathbf{g} . A similar interpretation holds for \mathbf{W}_y , $\boldsymbol{\eta}$, and $\boldsymbol{\nu}$. The matrix $\boldsymbol{\Theta}$ is the standard form wavelet transform of the matrix \mathbf{T} . As \mathbf{W}_y is orthonormal, $\boldsymbol{\nu} \sim \mathcal{N}(0, r\mathbf{I})$. Finally, $\boldsymbol{\gamma}$ is assumed to be distributed as a $1/f$ -type process and is thus zero mean with a diagonal covariance matrix denoted by \mathbf{P}_0 [2].

*This work was supported in part by NSF CAREER Grant MIP-9623721, subcontract GC123920NDG from Boston University under the AFOSR MURI Program on Reduced Signature Target Recognition, and DOE contract DE-FC07-95ID13395

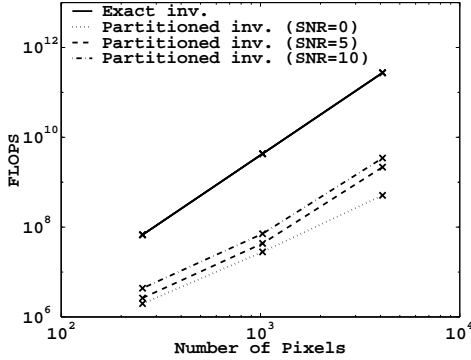


Figure 1: FLOP Comparison

For the problem defined by (2), the MAP estimate of γ given η is given as the solution to the normal equations:

$$(\Theta^T \mathbf{R}^{-1} \Theta + \mathbf{P}_0^{-1}) \hat{\gamma} = \Theta^T \mathbf{R}^{-1} \eta \quad (3)$$

where $\mathbf{L} = \Theta^T \mathbf{R}^{-1} \Theta + \mathbf{P}_0^{-1}$ is the information matrix. The problem of interest here is computing the error variances for this linear restoration problem. For the model in (2), the error covariance matrix, \mathbf{P} , is equal to \mathbf{L}^{-1} . The major difficulty with obtaining the error variances lies in the fact that even for small images \mathbf{L} is a large matrix making direct inversion impractical.

3 Outline of Inversion Method

Our method for solving this problem is motivated by the results in [3] indicating that for ill-posed problems $\hat{\gamma}$ can be divided into two sub-vectors: $\hat{\gamma}_1$ and $\hat{\gamma}_2$ corresponding to those coefficients for which the data do and do not provide “significant” information. Given this decomposition, the rows and column of (3) are re-ordered to obtain a permuted form of the normal equations:

$$\begin{bmatrix} \tilde{\mathbf{L}}_{11} & \tilde{\mathbf{L}}_{12} \\ \tilde{\mathbf{L}}_{12}^T & \tilde{\mathbf{L}}_{22} \end{bmatrix} \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} \quad (4)$$

where $\mathbf{v} = \Pi \Theta^T \mathbf{R}^{-1} \eta$, $\tilde{\mathbf{L}} = \Pi \mathbf{L} \Pi^T$, and Π is the permutation matrix that takes the vector $\hat{\gamma}$ to a vector with $\hat{\gamma}_1$ as the first collection of elements and $\hat{\gamma}_2$ as the second. It was observed in [3] that $\tilde{\mathbf{L}}_{11}$ was small and dense, $\tilde{\mathbf{L}}_{22}$ was large but well approximated as diagonal, and $\tilde{\mathbf{L}}_{12}$ was rectangular and sparse. It was shown that $\tilde{\mathbf{L}}^{-1}$ could be found with an algorithm whose complexity was dominated by the inversion of the small matrix $\tilde{\mathbf{L}}_{11}$. The primary problem was that Π was built based on knowing the error variances which meant that $\tilde{\mathbf{P}}$ had already been calculated.

In the talk accompanying this paper, we describe an algorithm for constructing Π in a manner which does not require prior knowledge of $\tilde{\mathbf{P}}$. The algorithm is composed of the following three stages:

1. **Building L.** The Schwartz inequality is used to construct a subset of the elements of $\Theta^T \mathbf{R}^{-1} \Theta$ without performing the matrix-matrix products.
2. **Permuting L.** We determine a permutation of

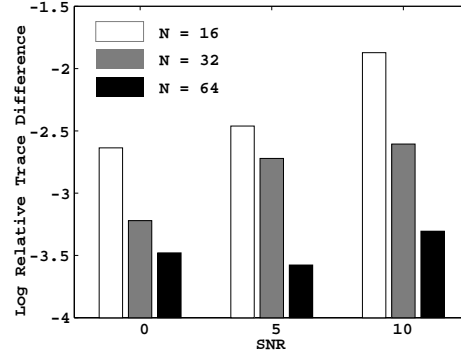


Figure 2: Relative trace difference

the rows and columns which addresses the tradeoff between having a large $\tilde{\mathbf{L}}_{22}$ and a loss in accuracy associated with its approximation as diagonal. An efficient, iterative method based on the sequential “local” optimization of a diagonal dominance criteria is introduced to determine a family of permutation matrices corresponding to various size $\tilde{\mathbf{L}}_{22}$.

3. **Inverting L.** Using the block-matrix inversion formula, we extract the error variances.

A by-product of this method is a matrix which serves as a left-preconditioner for the normal equations resulting in improved convergence of the CG algorithm at a fraction of the original computational cost.

As an example of the performance of this approach, in Figures 1 and 2, we compare the FLOP counts and accuracy results for our algorithm against direct inversion of the information matrix for a space-varying problem where T corresponds to convolution with a Gaussian window whose width varies sinusoidally in both spatial dimensions. Accuracy is measured as the 1-norm of the error in the variance calculation due to our method divided by the 1 norm of the exact variances. It is evident that over a wide range of SNRs, about three orders of magnitude in computational savings can be obtained with a lose in accuracy of well under 1%. In the talk accompanying this paper, we will provide a detailed description and analysis of this computational approach, demonstrate its effectiveness on a variety of problems, and consider the issue of pre-conditioning.

References

- [1] Mark R. Luetgten, W. Clem Karl, and Alan S. Willsky. Efficient multiscale regularization with applications to the computation of optical flow. *IEEE Trans. Image Proc.*, 3(1):41–64, 1994.
- [2] Eric L. Miller and Alan S. Willsky. A multiscale approach to sensor fusion and the solution of linear inverse problems. *Appl. Comput. Harmonic Anal.*, 2:127–147, 1995.
- [3] Eric L. Miller and Alan S. Willsky. Wavelet-based methods for the nonlinear inverse scattering problem using the Extended Born Approximation. *Radio Science*, 31(1):51–67, Jan–Feb 1996.

Title: **EM and Radar Signal Processing
Methods for Buried Object
Identification**

Presenter: **Prof. Eric Miller
Dept. Electrical & Computer Eng.
Northeastern University, Boston MA**

Sponsor: **ARO ODDR&E MURI**

Mission area: **Humanitarian Demining**

User Req's Supported: **4, 5, 13**

Motivations and Goals

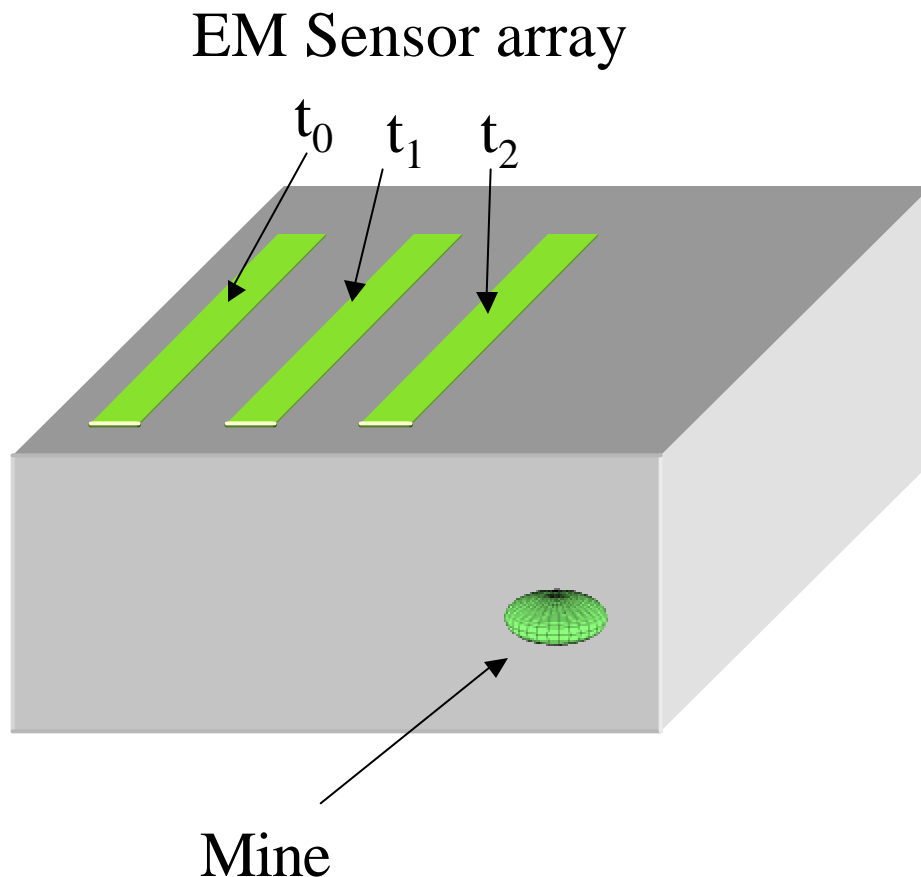
Development of detection, localization, and characterization methods for buried UXO from array-based EM and GPR sensing systems

- System types
 - Multi-monostatic such as GEM-3
 - Multi-multistatic such as FAR
- Algorithm characteristics
 - Statistically-based
 - Exploit spatial, spectral, and temporal diversity
 - Recursive implementation
 - Employ simplest possible physical model

Technical approach: Overview

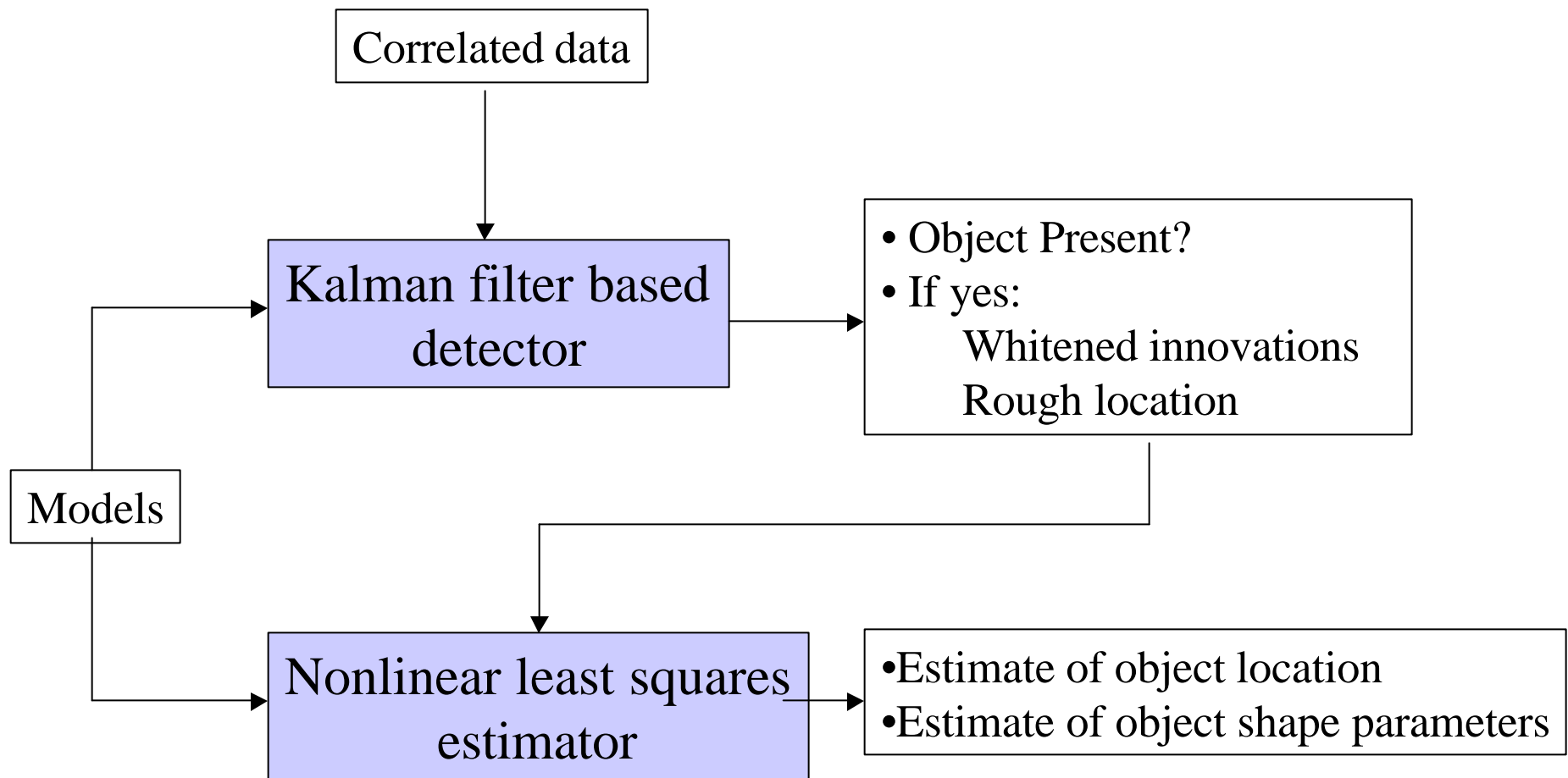
- EM projects
 - Recursive detection and localization
 - Spatial/spectral analysis for orientation estimation and target characterization
- Radar projects
 - Matched field methods for multi-target detection and localization
 - Detection and characterization from FAR data

EM Project 1: Recursive, Model-Based Detection and Localization



- Collect data one stop at a time
- Algorithm Goals
 - Detect as soon as possible
 - Localize
- Methods & Models
 - Physical target models
 - Recursive, stochastic clutter models
 - Statistical methods for detection, localization

Processing flow



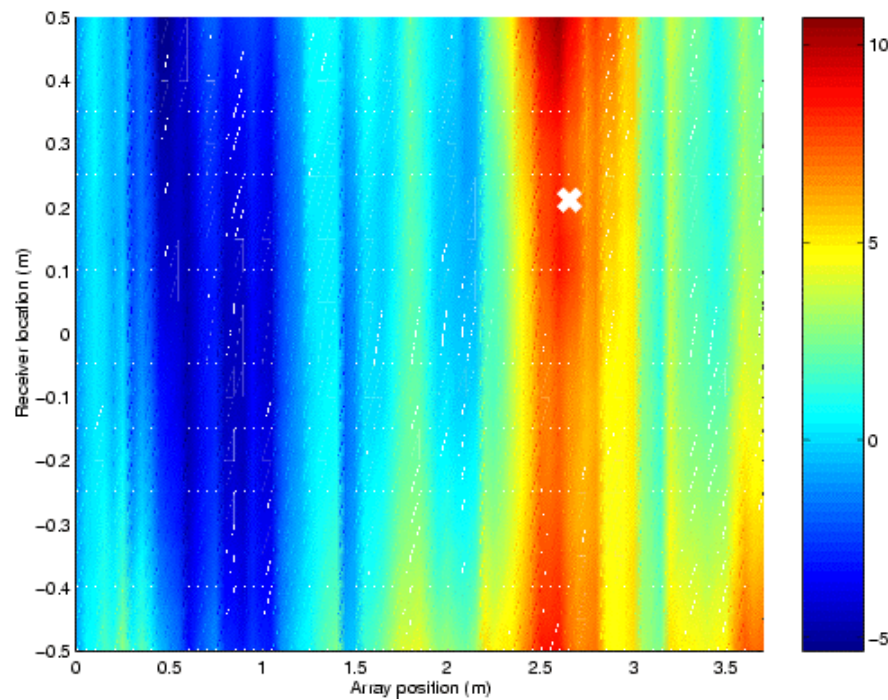
Example

- System
 - Single transmitter
 - Ten receivers on either side (\oplus 50cm)
 - Scan 75 positions between 0 and 3.7 m
- Randomize over mine location and size

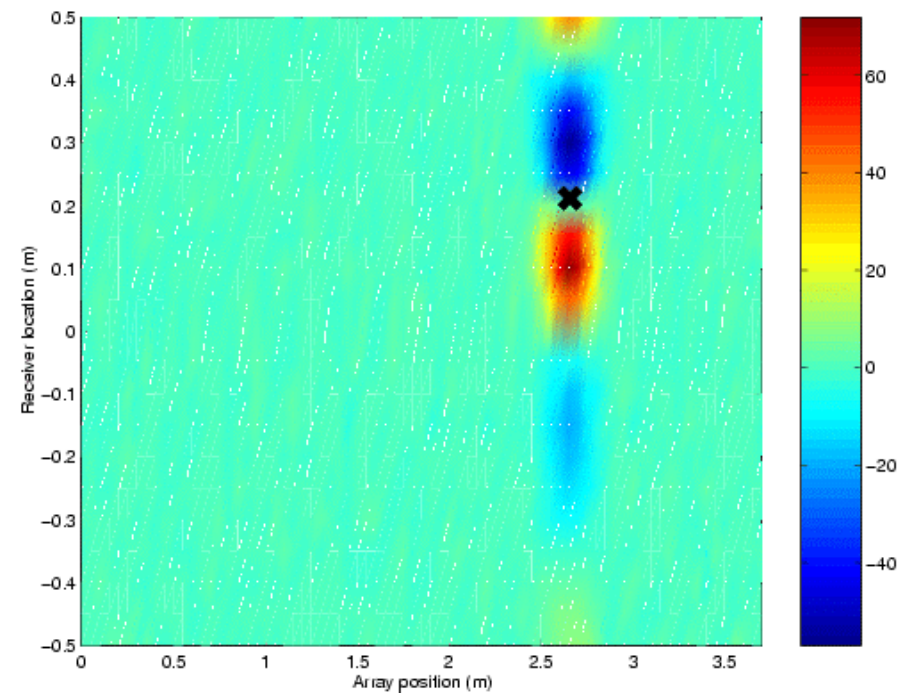
Example (cont)

- Detection
 - Based on single, nominal mine size and burial depth
 - Coarse localization: 20 cm x 20 cm spacing
- Localization / classification
 - Used detection results to generate initial position guess

Why Whiten?



Raw Data

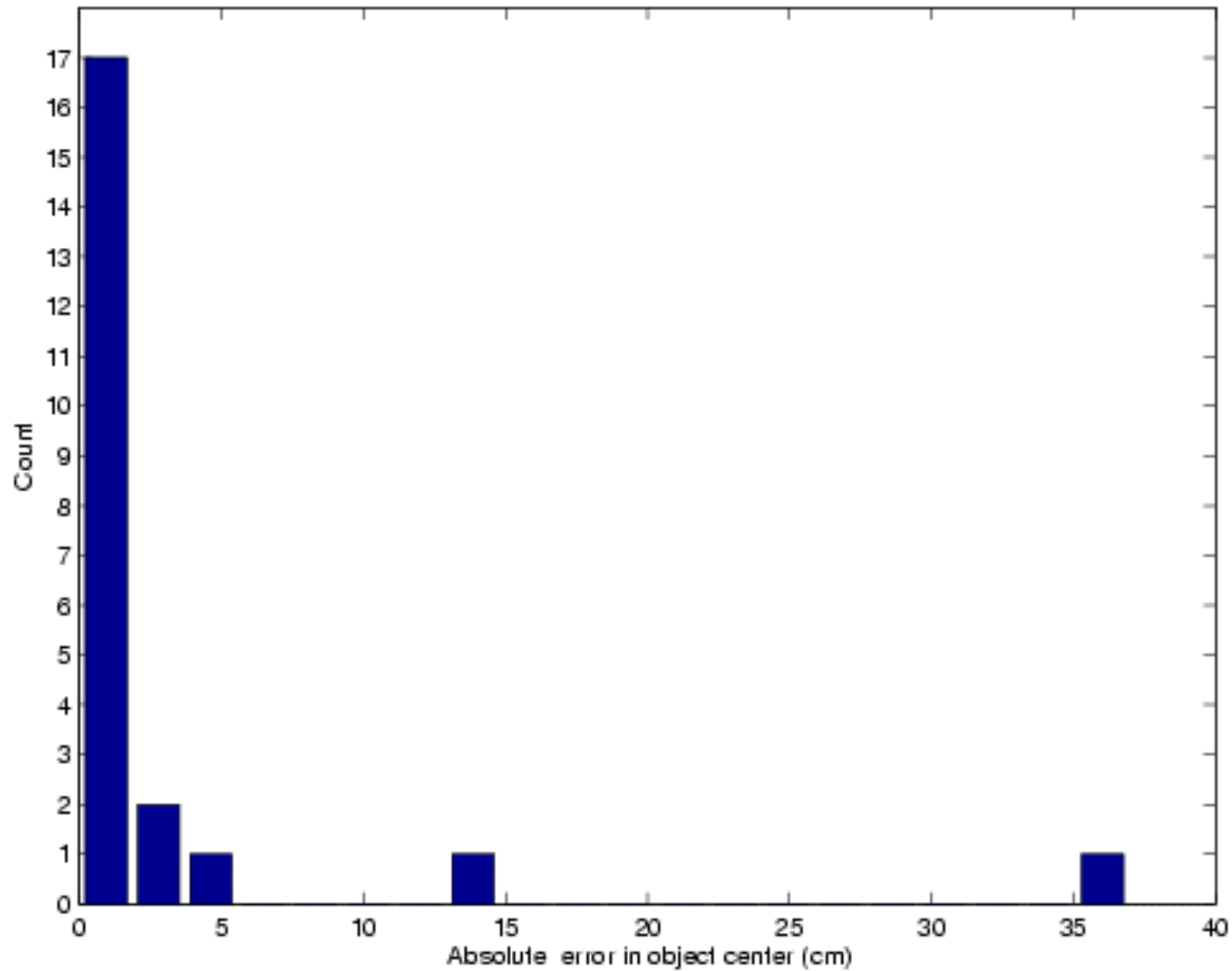


Filtered Innovations

Performance on this case

Parameter	True Value (cm)	Estimate (cm)
Major Axis Length	10.98	10.93
Minor Axis Length	4.33	4.39
Down Track Position	265.83	265.83
Cross Track Position	21.16	21.15
Depth Below Array	18.63	18.64

Errors in center estimates



EM Project 2:

UXO Orientation Estimation and Classification

- Goals
 - Estimation of target orientation and spectral dipole moment functions from multi-frequency spatial EM data
- Hypotheses/Approach
 - Classification from orientation-independent dipole moment spectrum
 - Estimate orientation explicitly
 - Entirely model based
- Team: NU, OU, Geophex

The model

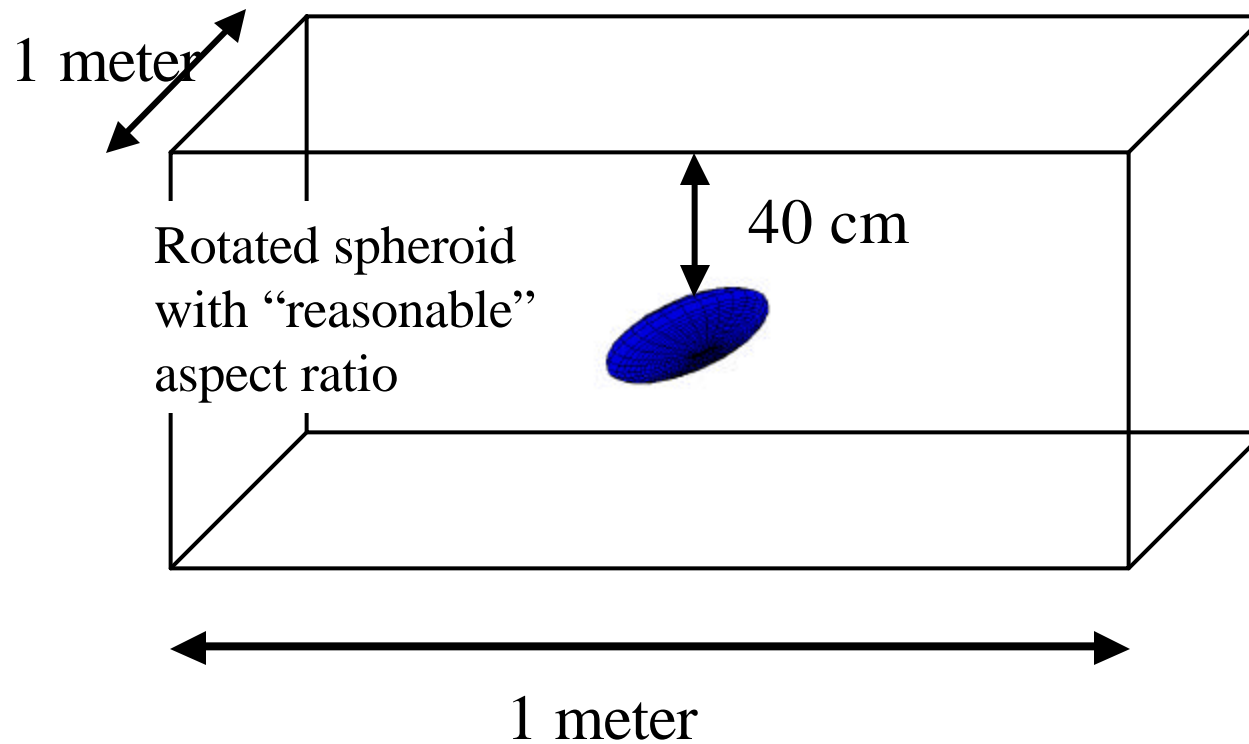
$$y = \mathbf{d}^T [\mathbf{R}^T \mathbf{D} \mathbf{R}] \mathbf{x}$$

- y = datum at location r and frequency ω
- \mathbf{x}, \mathbf{d} = 3×1 vectors of (x, y, z) field components dependent on positions of transmitter, receiver and mine
- \mathbf{R} = 3×3 rotation matrix encoding orientation of object
- \mathbf{D} = diagonal matrix holding frequency dependent dipole moments

The Processing

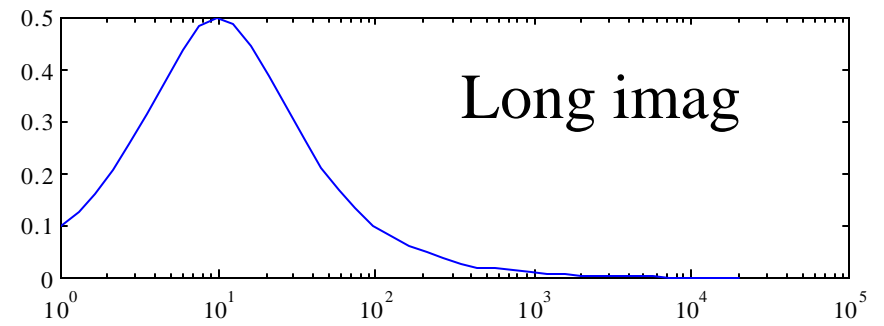
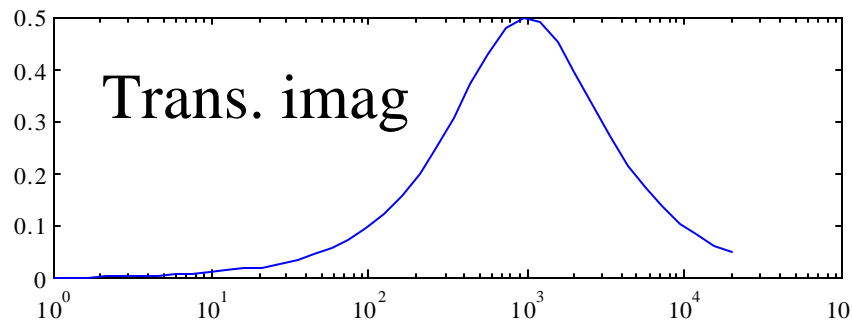
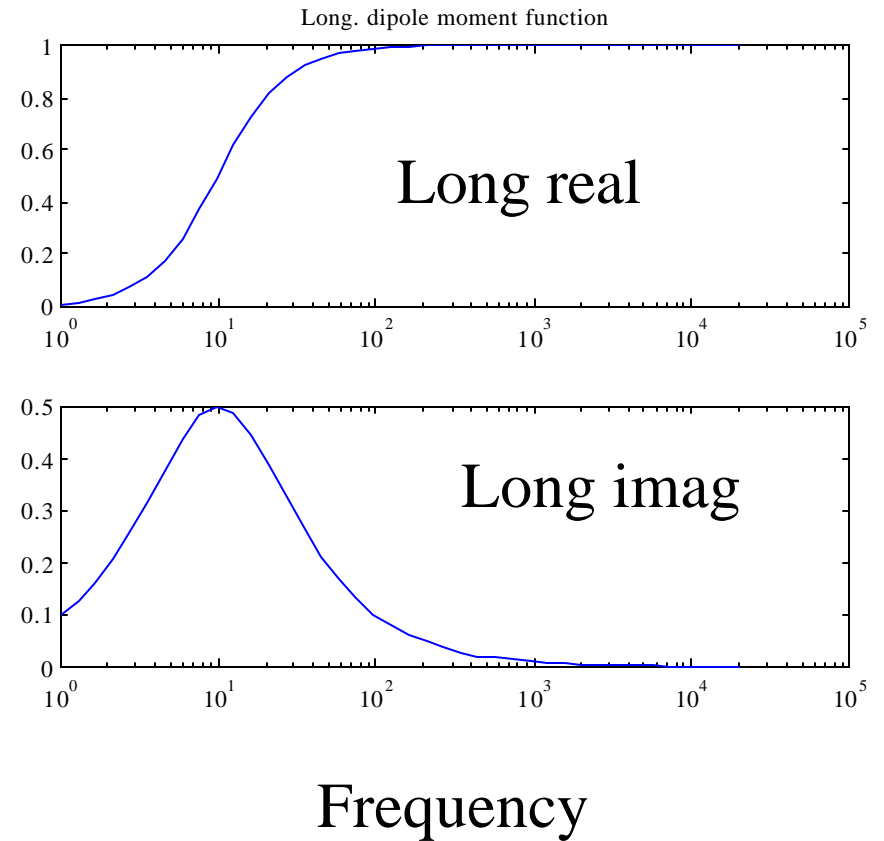
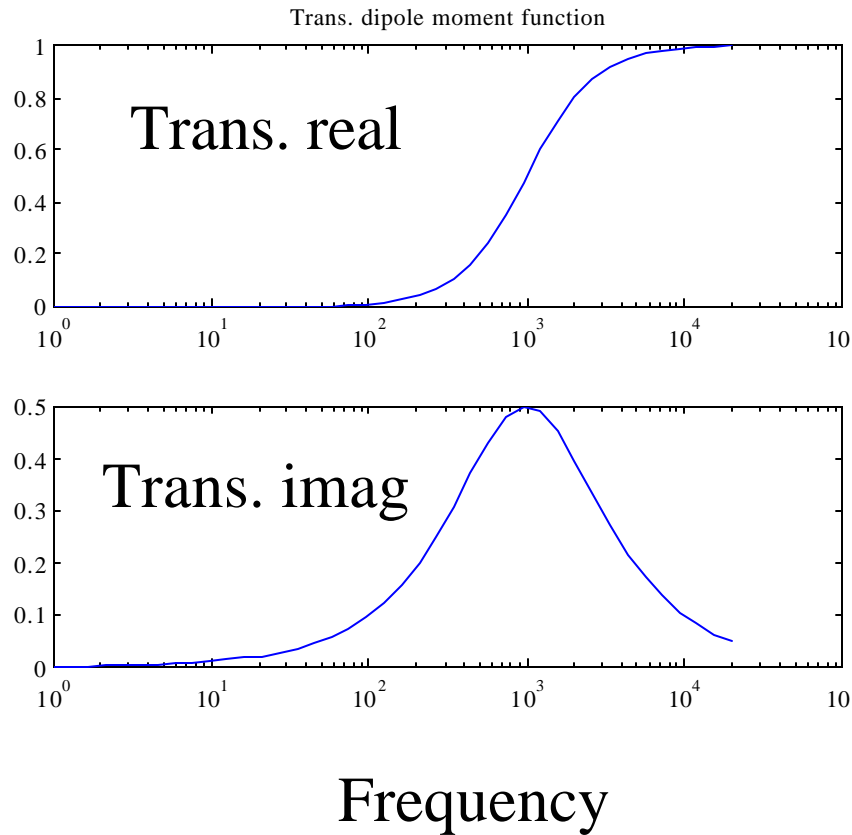
- Goals:
 - Determine two unique elements per frequency of \mathbf{D}
 - Determine two orientation angles buried in \mathbf{R}
- Method:
 - Estimation of dipole moments = linear MAP problem
 - Estimation of angles = nonlinear ML problem
 - Performance analysis: CRB readily available
 - Lots of structure.
 - Looking into efficient implementation
 - Also looking into sub-optimal, still faster methods

Example

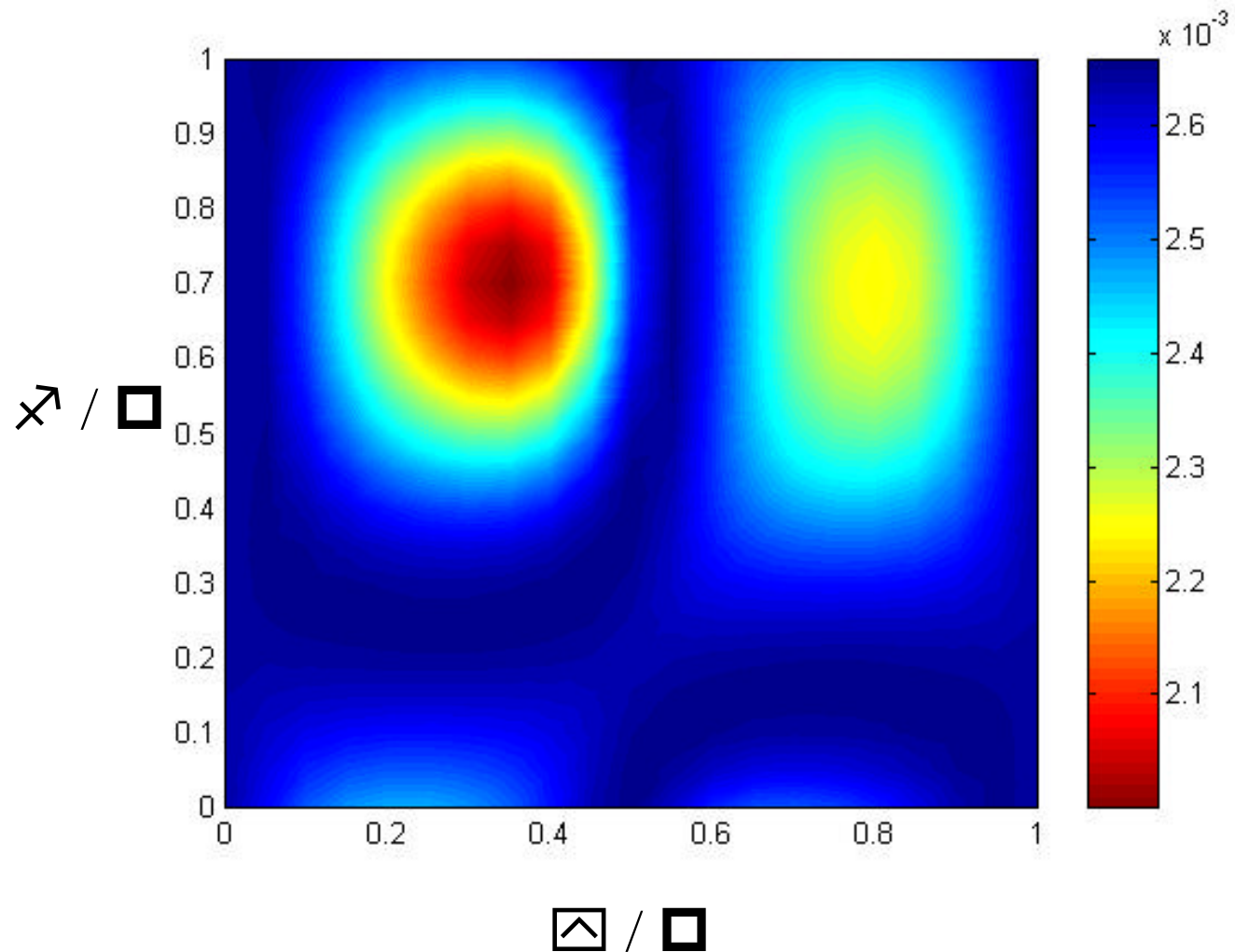


- 20dB SNR
- 10 x 10 grid of measurements
- Non-permeable
- 20 log. Spaced frequencies between 1 Hz and 20kHz
- Assume position previously determined

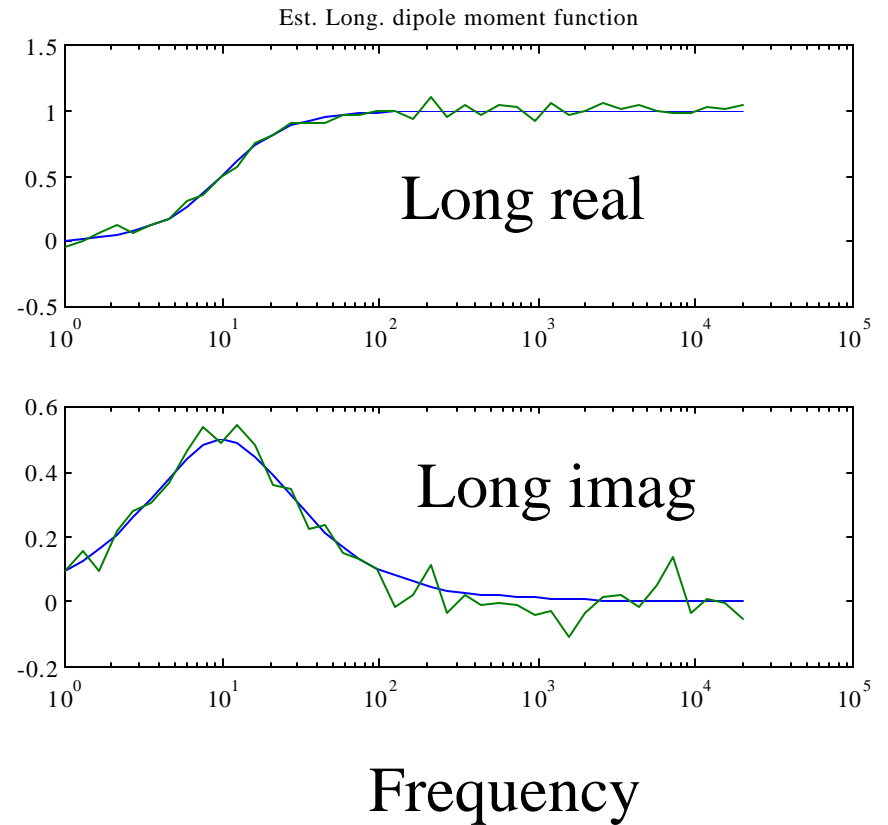
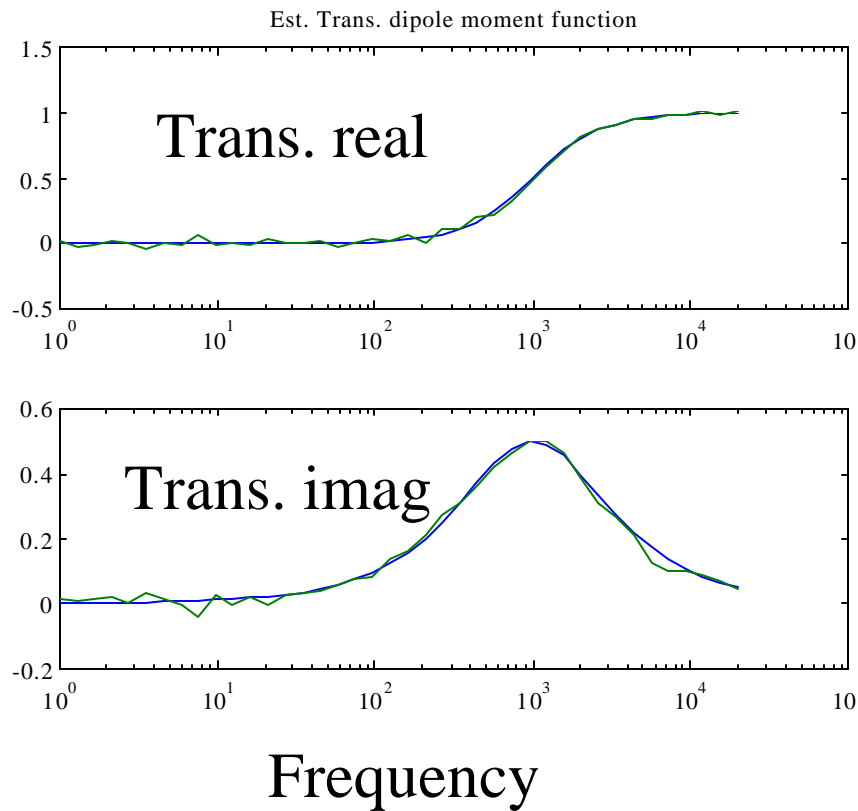
Dipole moments



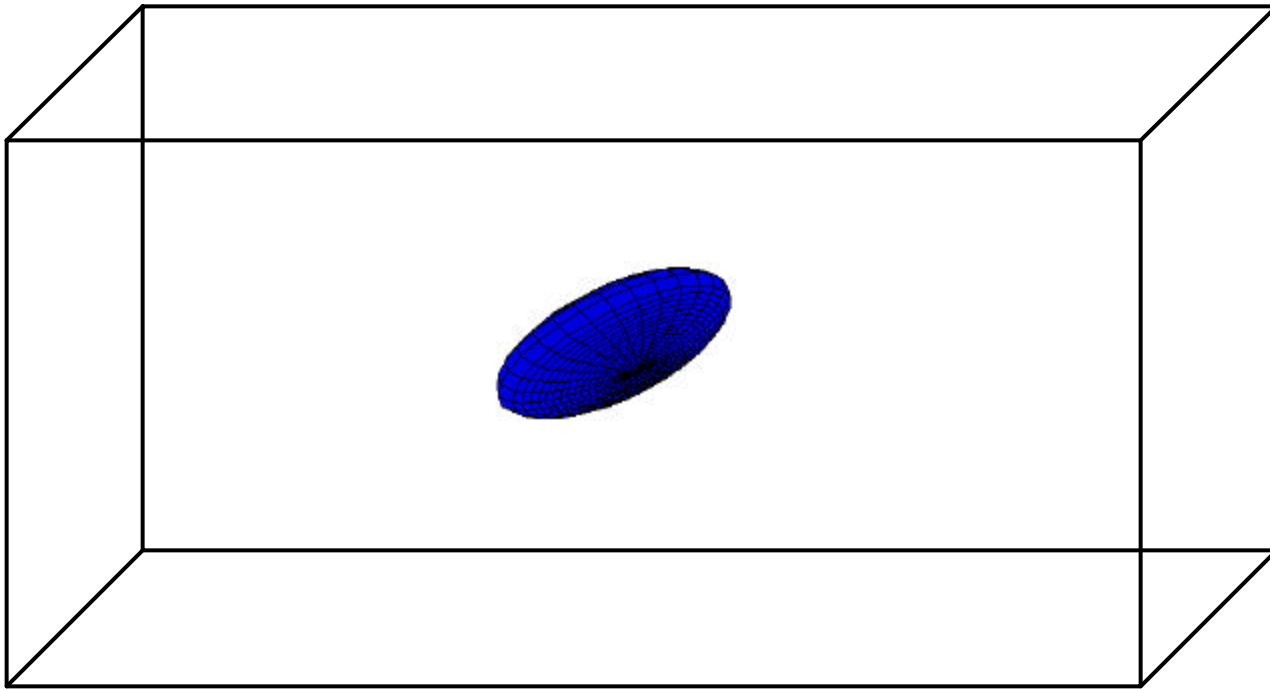
Cost versus rotation angles



Estimated dipole moments



Reconstructed object

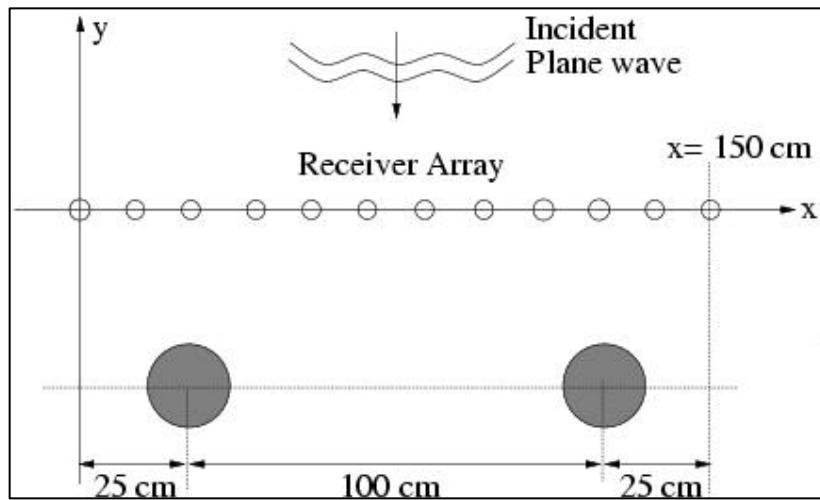


Radar Project 1:

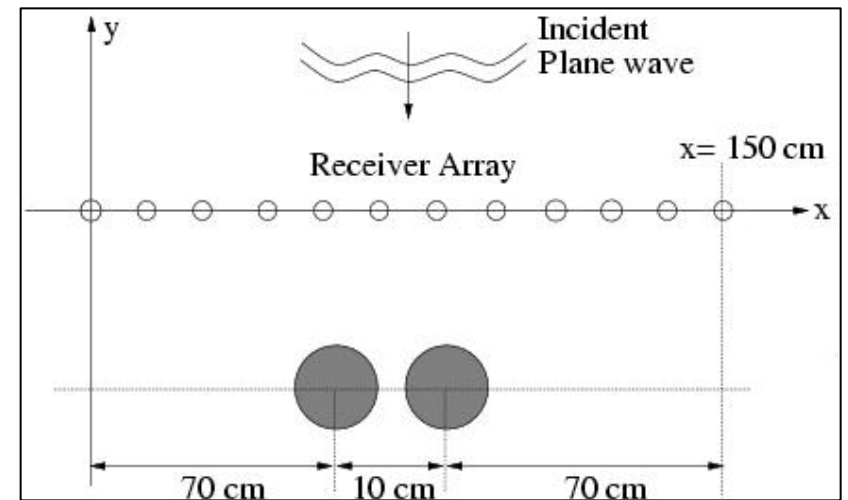
Array Processing for Multi-Object Near-Field Detection

- Goals
 - Employ high resolution array processing methods to detect multiple objects in near field of antenna array
- Hypotheses/approach
 - Use of model-based array processing would allow for discrimination of closely spaced objects
 - Matched field processing with various forward solvers assuming homogeneous medium
 - Closed form CRB analysis
- Team: NU

Experimental Setup

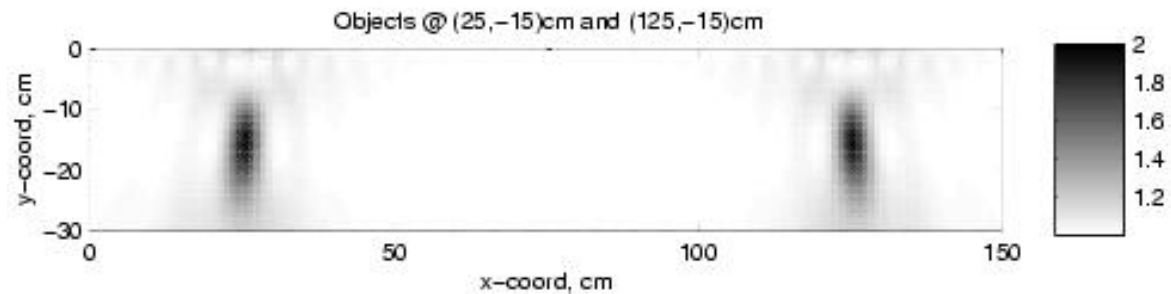


Well separated metal objects

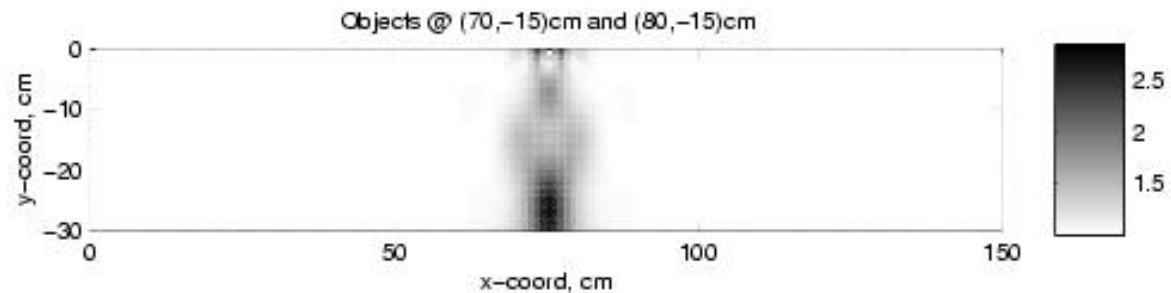


Closely spaced metallic objects

Simple physical model

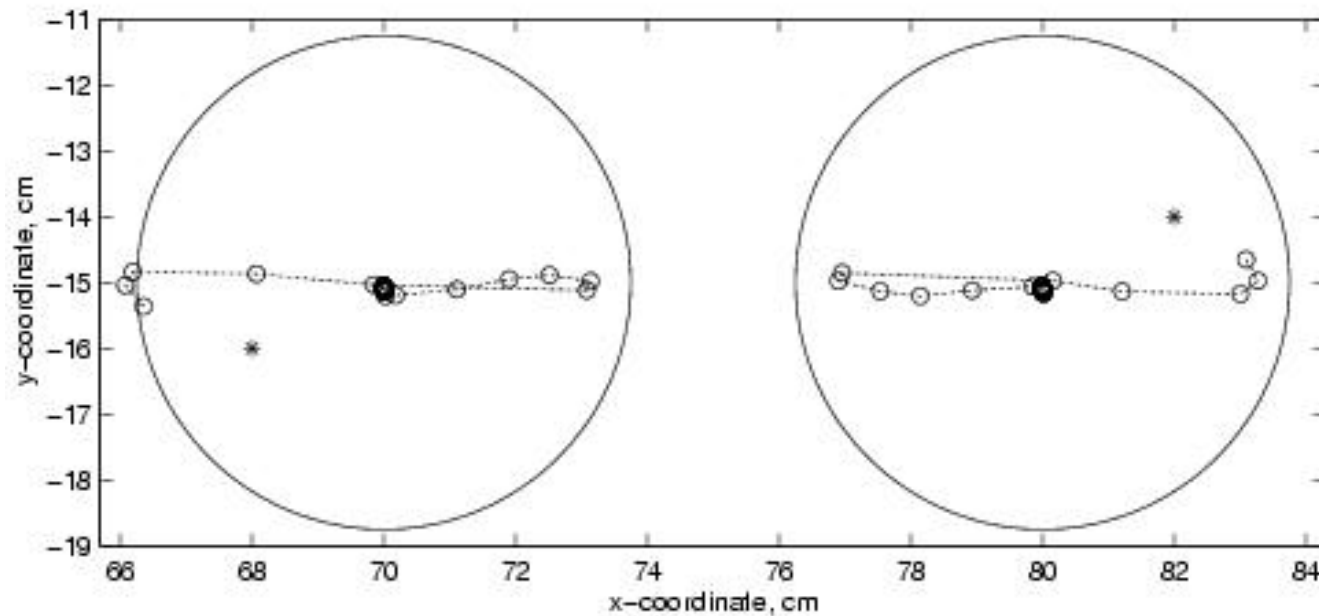


Well separated: Easily locates both scatterers



Closely spaced: Cannot see separate objects

Advanced model, close objects

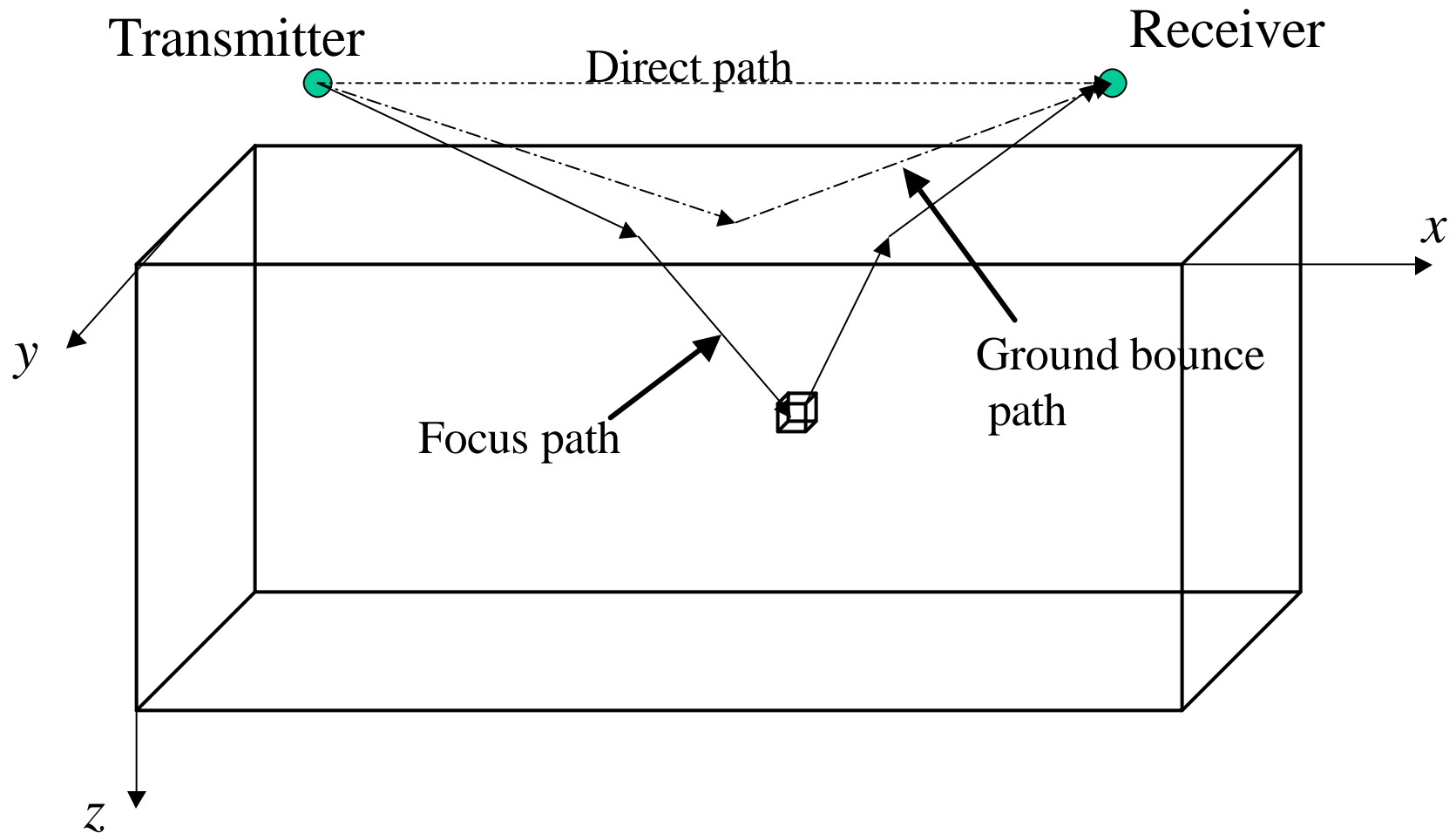


Both objects well localized

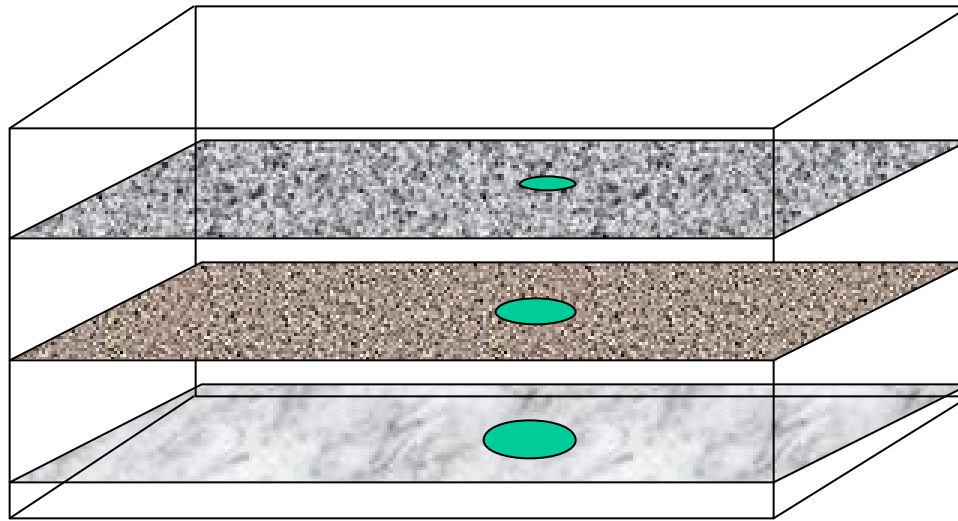
Radar Project 2: FAR Processing

- Goals
 - Detection and localization of buried objects from FAR-type data
- Hypotheses/approach
 - Detection can be done using stochastic, recursive change detection methods
 - Localization using delay estimation based on hybrid ray-optics/FDTD model
- Team: NU and GeoCenters

FAR System Overview



Layer-based detection



- **Target:**
 - Elliptical anomaly with correlated structure across layers
- **Approach:**
 - Statistical model for background texture/clutter
 - Recursively whiten data and identify cross-layer anomalies




Localization and classification

- Model
 - Received signal superposition of delayed and scaled base signal with delays coming from half-space ray-optics approximation
 - Base signal is target dependent and calculated from FDTD simulation
- Processing
 - Time delay estimation gives target location
 - Multiple model approach gives target type

Conclusions

- EM Projects
 - Demonstrated utility of model based, statistical methods for detection and localization
 - Near real time implementation possible
- Radar projects
 - Demonstrated feasibility for finding multiple objects in near field of array
 - Implementable approach to FAR detection, localization, and classification

Next Steps and Transition

- EM
 - Project 1: Verify using data from array-based system
 - Project 2: Continued work using GEM-3 data
 - Goal: merge the two for overall approach to EM ATR
- Radar
 - Project 1:
 - Many extensions (3D models, include interface, clutter models, unknown object characteristics ...).
 - Requires 's and  / 
 - Project 2: Just getting off the ground

The Processing

- Estimate unique elements of the symmetric matrix
 $\mathbf{B} = \mathbf{R}^T \mathbf{D} \mathbf{R}$.
 - Small (6x6) linear least squares estimation problem
 - Closed form performance analysis
- From estimate of \mathbf{B} , determine dipole moments (2) and rotation angles (2)
 - Can write dipole moments as function of angles
 - Therefore solution = 2D non-linear least squares estimation problem for angles.
 - CRB analysis easily available
 - Also quick

QMR-BASED PROJECTION TECHNIQUES FOR THE SOLUTION OF NON-HERMITIAN SYSTEMS WITH MULTIPLE RIGHT-HAND SIDES*

MISHA KILMER[†], ERIC MILLER[‡], AND CAREY RAPPAPORT[‡]

Abstract. In this work we consider the simultaneous solution of large linear systems of the form $Ax^{(j)} = b^{(j)}$, $j = 1, \dots, K$, where A is sparse and non-Hermitian. We describe single-seed and block-seed projection approaches to these multiple right-hand side problems that are based on the QMR and block QMR algorithms, respectively. We use (block) QMR to solve the (block) seed system and generate the relevant biorthogonal subspaces. Approximate solutions to the nonseed systems are simultaneously generated by minimizing their appropriately projected (block) residuals. After the initial (block) seed has converged, the process is repeated by choosing a new (block) seed from among the remaining nonconverged systems and using the previously generated approximate solutions as initial guesses for the new seed and nonseed systems. We give theory for the single-seed case that helps explain the convergence behavior under certain conditions. Implementation details for both the single-seed and block-seed algorithms are discussed and advantages of the block-seed algorithm in cache-based serial and parallel environments are noted. The computational savings of our methods over using QMR to solve each system individually are illustrated in two examples.

Key words. QMR, projection, Krylov subspace, iterative methods, block Krylov

AMS subject classifications. 65F10, 65N22

PII. S1064827599355542

1. Introduction. In many applications one desires the solution of multiple linear systems of the form

$$(1.1) \quad Ax^{(j)} = b^{(j)}, \quad j = 1, \dots, K,$$

involving the same $N \times N$ coefficient matrix A but K different right-hand sides $b^{(j)}$, all of which are available simultaneously. Such problems arise, for instance, in the numerical solution of frequency-domain electromagnetic wave scattering; here, the right-hand sides might correspond to incident fields over the scatterer induced either by plane waves at various angles of incidence or by excitation sources at different locations.

Systems involving large, sparse matrices make good candidates for solution by iterative Krylov subspace methods since storage is kept to a minimum and matrix-vector products can be done efficiently. However, the naive approach of solving each of the K linear systems independently using a Krylov subspace method does not take advantage of the fact that the $b^{(j)}$'s, and hence the $x^{(j)}$'s, may be *closely related* due to the underlying physical nature of the problem. By closely related, we mean that the solution to the j th system has large components in the initial directions of the k -dimensional ($k \ll N$) Krylov subspace generated from one of the other systems. Projection-type techniques for both the Hermitian and non-Hermitian cases,

*Received by the editors May 4, 1999; accepted for publication (in revised form) September 29, 2000; published electronically August 15, 2001. This work was supported by the Army Research Office Demining MURI under grant DAAG55-97-1-0013.

<http://www.siam.org/journals/sisc/23-3/35554.html>

[†]Department of Mathematics, Tufts University, Medford, MA 02155 (mkilme01@tufts.edu).

[‡]Electrical and Computer Engineering Department, Northeastern University, Boston, MA 02115 (elmiller@ece.neu.edu, rappaport@neu.edu).

discussed in more detail below, that specifically exploit such shared information have been proposed (e.g., see [5, 26, 24] and the references therein).

Another alternative is to use a block Krylov subspace algorithm to solve the systems simultaneously [16, 23, 4, 8, 15]. Essentially these methods seek solutions in block Krylov subspaces, or some deflated version thereof, generated by the matrix A and the $N \times K$ matrix $R = B - AX_0$; here the columns of B are the $b^{(j)}$ and the columns of X_0 are the initial estimates for each of the systems. However, this approach can be more expensive in terms of storage than projection techniques because the length of the recurrences to update the iterates depends on the number of right-hand sides, or, in the case of deflation [8], the number of right-hand sides corresponding to the deflated Krylov sequences. Also, if a deflation technique is used, a deflation tolerance must be specified in advance, and we have found in experiments that the performance and convergence of the systems depend in a somewhat unpredictable manner on this value. On the other hand, block Krylov subspace algorithms have the advantage that they are better suited to parallelism [17, 14, 15] and make use of higher level BLAS [2]. Therefore, in this paper we develop new single and block-seed projection approaches based on the QMR and block QMR algorithms, respectively; our block-seed method exploits the best properties of the block QMR algorithm while preserving the basic properties of our sequential projection technique. To our knowledge, ours is the first block-seed projection algorithm for non-Hermitian linear systems with multiple right-hand sides; in particular, as our algorithm is built around the BL-QMR algorithm of [8], it incorporates a deflation strategy.

Specifically, the idea of a projection technique is to first select one of the systems as “seed” and solve it by an iterative Krylov subspace method. As the relevant subspaces are generated, the approximations to the other systems are simultaneously updated by projecting the residual onto a particular subspace and by either enforcing a Galerkin-type condition [11, 25] or minimizing the projected residual [24]. Such methods are sometimes referred to as Lanczos–Galerkin approaches [21].

Smith [25] and Joly [11] both consider a projection approach based on BiCG for nonsymmetric A . In [11] a similar approach for CGS is also given. However, the BiCG-projection approach can exhibit the potentially erratic convergence behavior observed when applying BiCG to a single linear system (see the results in [24]). Simoncini and Gallopoulos [24] also present an approach to solving (1.1) when A is nonsymmetric.

Our single-seed projection algorithm is most similar in concept to the project-minimize approach in [24]. However, as a result of the underlying unsymmetric Lanczos process, we do not need to store the basis vectors, we do not need to predetermine a subspace dimension, and we show that the approximate solutions and residuals to the nonseed systems are cheaply computed and available at every stage of the algorithm because they are updated with short-term recurrences. Since we are minimizing quantities rather than enforcing Galerkin conditions, the convergence behavior should be less erratic than methods based on the latter. As noted and as we illustrate in Theorem 3.4, the success of our single-seed method depends on the closeness, in the sense described above, among the right-hand sides. The block-seed approach we introduce here is more efficient when the right-hand sides are not all close.

This paper is organized as follows. In section 2, we give the necessary background on the QMR approach. We give an outline of our single-seed projection approach and discuss its convergence in section 3. Background on the block QMR algorithm is presented in section 4. The block variant of our QMR-projection algorithm is reported in section 5, and related computational issues are discussed in section 6. Section 7

gives numerical results, and section 8 reports conclusions and future work.

2. Background: The QMR algorithm. The quasi-minimal residual (QMR) algorithm was introduced by Freund and Nachtigal in [9]. The original algorithm was based on three-term recurrences. In [10], the authors proposed a mathematically equivalent algorithm which employed coupled two-term recurrences. Since the latter variant has been found to be more numerically stable for solving linear systems, in numerical experiments we use this implementation. However, to simplify the notation in this section and in section 3, and to be consistent with the notation in section 5, we will follow the notation in [9]. Further, for simplicity, we consider only the version without look-ahead but note that our algorithm could be adapted to account for look-ahead.

In the remainder of the paper, the notation $\|\cdot\|$ always refers to the Euclidean norm $\|\cdot\|_2$. The superscript T is used to denote the transpose operation, and superscript $*$ is used to denote the conjugate transpose operation.

A Krylov subspace of dimension k generated by a matrix G and a vector q is defined according to

$$K_k(G, q) = \text{span}\{q, Gq, G^2q, \dots, G^{k-1}q\}.$$

The QMR algorithm is a Krylov-subspace-based iterative method that can be used to solve non-Hermitian linear systems of the form $Ax = b$, $A \in \mathcal{C}^{N \times N}$. At the k th iteration, the current solution estimate has the form

$$(2.1) \quad x_k = x_0 + V_k z_k,$$

where x_0 denotes the initial guess and $V_k = [v_1, v_2, \dots, v_k]$ is an $N \times k$ matrix whose columns are basis vectors for $K_k(A, v_1)$ with $v_1 = r_0/\|r_0\|$ and $r_0 = b - Ax_0$. The length k vector z_k is chosen as the solution to a particular minimization problem, as discussed below. Those basis vectors are generated via the nonsymmetric Lanczos process (see [22]) and are constructed to be biorthogonal to vectors $w_i, i = 1, \dots, k$, which form a basis for $K_k(A^T, w_1)$.¹ The columns of the $N \times k$ matrix W_k are the w_i .

From biorthogonality it follows that

$$(2.2) \quad W_{k+1}^T V_{k+1} = D_{k+1}, \quad D_{k+1} = \text{diag}(\delta_1, \dots, \delta_{k+1}), \quad \delta_i \neq 0.$$

Also as a result of the Lanczos algorithm we obtain the relation

$$(2.3) \quad AV_k = V_{k+1} \bar{T}_k,$$

where \bar{T}_k is a $(k+1) \times k$ tridiagonal matrix. Using (2.1), (2.2), and (2.3), and setting $\beta = \|r_0\|$, at the k th iteration the residual, $r_k = b - Ax_k$, is given by [22]

$$(2.4) \quad r_k = V_{k+1}(\beta e_1 - \bar{T}_k z_k),$$

where e_1 denotes the first Cartesian unit vector. Since the columns of V_{k+1} are not orthonormal,

$$\|r_k\| \leq \|V_{k+1}\| \|\beta e_1 - \bar{T}_k z_k\|.$$

¹Here we always take $w_1 \equiv v_1$, but note that other choices are possible. A version of the algorithm is also possible using $K_k(A^*, w_1)$ for the second Krylov subspace.

The QMR algorithm determines z_k by minimizing the norm of the quasi-residual term [9]; that is,

$$z_k = \arg \min_{z \in \mathbb{C}^{k \times 1}} \|\beta e_1 - \bar{T}_k z\|.$$

We make the following alternate observation. From (2.4) and using biorthogonality,

$$\|D_{k+1}^{-1} W_{k+1}^T r_k\| = \|\beta e_1 - \bar{T}_k z_k\|.$$

Thus, the z_k which defines the k th QMR iterate can also be thought of as the one that minimizes the norm of the residual projected onto a smaller dimensional subspace. We will make use of this alternate definition of the QMR iterates in subsequent sections.

3. The single-seed QMR-projection algorithm. In this section we describe a single-seed QMR-projection algorithm for solving linear systems of the form (1.1). Our algorithm proceeds as follows. First, we select one system, say, system j , to serve as “seed” and apply QMR (without look-ahead) to the seed system. In the following, we use $r_0^{j,l}$ to denote the initial residual to system l , where l denotes the index of any of the nonconverged systems given the starting guess $x_0^{j,l}$. We use $r_k^{j,l}$ to denote the residual of system l after k iterations. Since different choices of seed lead to different Krylov subspaces and hence different iterates, the superscript j is used to denote that this is the residual at the k th iteration for system l when system j was used as seed. By the beginning of the k th iterate, QMR has generated biorthogonal bases for two k -dimensional Krylov subspaces, $K_k(A, r_0^{j,j})$ and $K_k(A^T, r_0^{j,j})$. We denote the respective bases by the vectors $v_{j,i}$ and $w_{j,i}$, $i = 1, \dots, k$: the subscript j is used to indicate that this particular set was generated using system j as seed. These vectors are the columns of the $N \times k$ matrices $V_{j,k}$ and $W_{j,k}$, respectively. The corresponding $(k+1) \times k$ tridiagonal matrix is denoted as $\bar{T}_{j,k}$ (compare to (2.3)). By the end of the k th iterate, QMR has also generated the unnormalized versions of the vectors $v_{j,k+1}$ and $w_{j,k+1}$ for use in the $(k+1)$ st iteration.

Let us comment on the values of $x_0^{j,l}$. If we suppose that j was the seed system and converged after m steps and that the index of the next seed system is j^* , then we set $x_0^{j^*,l} = x_m^{j,l}$ for all indices l such that system l has not already converged.

In the previous section, we have seen that the k th iterate corresponding to the seed system is given by

$$x_k^{j,j} = x_0^{j,j} + V_{j,k} z_k^{j,j} \quad \text{with} \quad z_k^{j,j} = \arg \min_{z \in \mathbb{C}^{k \times 1}} \|\beta e_1 - \bar{T}_{j,k} z\|.$$

Now we also want the k th iterate of the (nonconverged) nonseed system, say, system l , to lie in $x_0^{j,l} + K_k(A, r_0^{j,l})$, in other words,

$$(3.1) \quad x_k^{j,l} = x_0^{j,l} + V_{j,k} z_k^{j,l}, \quad l \neq j.$$

Next we must decide how to define $z_k^{j,l}$. From (2.3) (with $V_{j,k}$ in place of V_k) and (3.1),

$$(3.2) \quad \begin{aligned} r_k^{j,l} &= b^{(l)} - A x_k^{j,l} \\ &= r_0^{j,l} - V_{j,k+1} \bar{T}_{j,k} z_k^{j,l}. \end{aligned}$$

Therefore, using biorthogonality,

$$\|D_{j,k+1}^{-1} W_{j,k+1}^T r_k^{j,l}\| = \|D_{j,k+1}^{-1} W_{j,k+1}^T r_0^{j,l} - \bar{T}_{j,k} z_k^{j,l}\|.$$

Finally, we use the above equality to determine $z_k^{j,l}$:

$$(3.3) \quad z_k^{j,l} = \arg \min_{z \in \mathbb{C}^{k \times 1}} \|D_{j,k+1}^{-1} W_{j,k+1}^T r_0^{j,l} - \bar{T}_{j,k} z\|.$$

3.1. Computational issues. Let us sketch how to efficiently compute the iterates and residuals of the nonseed systems. More details can be found in [13]. As above, we will use the index l to denote an arbitrary nonseed system and j to denote the seed system.

Let the QR decomposition of $\bar{T}_{j,k}$ be

$$\bar{T}_{j,k} = Q_{j,k}^* \begin{bmatrix} R_{j,k} \\ 0 \end{bmatrix},$$

where $Q_{j,k}$ is a product of Givens rotations and $R_{j,k}$ is $k \times k$ upper triangular with upper bandwidth 2. It can be shown [13] that using the QR factorization to solve (3.3) results in the easily solved system

$$(3.4) \quad z_k^{j,l} = R_{j,k}^{-1} t_k^{j,l},$$

where $t_k^{j,l}$ differs from $t_{k-1}^{j,l}$ only in its last entry, which we shall denote by $y_k^{j,l}$. The norm of the projected residual in (3.3) is given by the scalar $|\tau_{k+1}^{j,l}|$. Further, $\tau_{k+1}^{j,l}$ and $y_k^{j,l}$ can be updated from $\gamma_{k+1}^{j,l}$ and $\tau_k^{j,l}$ with $\delta_{j,k} \equiv w_{j,k}^T v_{j,k}$:

$$(3.5) \quad \begin{bmatrix} y_k^{j,l} \\ \tau_{k+1}^{j,l} \end{bmatrix} = \begin{bmatrix} c_{j,k} & s_{j,k} \\ -s_{j,k} & c_{j,k} \end{bmatrix} \begin{bmatrix} \tau_k^{j,l} \\ \gamma_{k+1}^{j,l} \end{bmatrix} \quad \text{with} \quad \gamma_{k+1}^{j,l} \equiv \frac{1}{\delta_{j,k+1}} w_{j,k+1}^T r_0^{j,l},$$

where $c_{j,i} \in \mathcal{R}$, $s_{j,i} \in \mathcal{C}$, $c_{j,i}^2 + |s_{j,i}|^2 = 1$.

As in equation (4.8) of [9], we define $P_{j,k} = [p_{j,1}, p_{j,2}, \dots, p_{j,k}] \equiv V_{j,k} R_{j,k}^{-1}$. Since $R_{j,k}$ is upper triangular with bandwidth 2, there is a short-term recurrence relation for the $p_{j,k}$ [9]. Using (3.1) and (3.4), the k th iterate of the l th system is given by

$$(3.6) \quad x_k^{j,l} = x_{k-1}^{j,l} + y_k^{j,l} p_{j,k}.$$

From this, we derive an iterative update for $r_k^{j,l}$ that does not require any additional matrix-vector products per iteration as follows.

LEMMA 3.1. *The residual at the k th iteration corresponding to the l th system is given by*

$$r_k^{j,l} = r_{k-1}^{j,l} - y_k^{j,l} f_{j,k}, \quad \text{where} \quad f_{j,k} \equiv A p_{j,k},$$

and can be computed in $O(N)$ flops.

Proof. The proof follows from (3.6), the bandedness of $R_{j,k}$, and the definition of $p_{j,k}$ above. (A detailed proof can be found in [13].) \square

3.2. Seed selection. Clearly, the success of our QMR-projection approach at reducing the total number of matrix-vector products needed to solve all the systems to the desired tolerance depends on which and in what order systems are selected as seed. We use the approach in [24]; namely, we choose the seed index j such that the norm of the residual of the corresponding system is larger than all the remaining nonconverged systems. Developing more informed selection heuristics remains a subject for future research.

3.3. Theory. Suppose that QMR has been run once and that the initial seed system has converged after m steps. Our algorithm proceeds by choosing another seed and using as its initial guess that solution obtained via projection as the first system was solved. One of the results of this section is that in exact arithmetic, assuming A is diagonalizable, if some of the right eigenvectors are well approximated by Ritz vectors corresponding to the first Krylov subspace generated, the rate of convergence of the second seed system behaves as if the corresponding part of the spectrum of A is cut off. The proof technique follows along the lines of the proof of Lemma 3.2 in [5]. $\kappa_2(\cdot)$ denotes the 2-norm condition number of the argument.

We assume A is diagonalizable with eigendecomposition $A = \hat{Z}\Lambda\hat{S}$, where $\hat{S} = \hat{Z}^{-1}$. Here $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$. We use \hat{z}_j to denote the j th column of \hat{Z} and \hat{s}_j^* to denote the j th row of \hat{S} . Without loss of generality, we may assume that $\|\hat{z}_j\| = 1$. After step m of QMR applied to the seed system, let $T_{1,m}U_{1,m} = U_{1,m}\Sigma_{1,m}$ be the eigendecomposition of $T_{1,m}$, where $T_{1,m}$ denotes the tridiagonal $m \times m$ leading submatrix of $\bar{T}_{1,m}$. Since QMR is built on top of the unsymmetric Lanczos process, in exact arithmetic if m is large enough, we expect some of the Ritz vectors given by the columns of $V_{1,m}U_{1,m}$ to be good approximations to, say, $n \leq m$ of the right eigenvectors² [1, 6, 7, 2] (assuming these eigenvectors are present in the starting vector for the first seed system). Under these definitions and assumptions, we have the following.

THEOREM 3.2. *Consider two systems $Ax^{(1)} = b^{(1)}$ and $Ax^{(2)} = b^{(2)}$ with A an $N \times N$ matrix. Let $x_0^{2,2}$ be the starting vector for the second system obtained via our projection approach after m steps of QMR have been run using the first system as seed; that is, $x_0^{2,2} = x_m^{1,2}$.*

Define \mathcal{I} as the set of indices of the n right eigenvectors that are well approximated by $n \leq N$ of the m Ritz vectors $V_{1,m}U_{1,m}$. Define $\bar{x}_0^{2,2}$ such that $x^{(2)} - \bar{x}_0^{2,2}$ is the projection of $x^{(2)} - x_0^{2,2}$ on $\text{span}\{\hat{z}_j, j \notin \mathcal{I}\}$. Let $\bar{x}_i^{2,2}$ be the i th iterate of GMRES applied to system 2 with initial guess $\bar{x}_0^{2,2}$. Then for any i

$$\|b^{(2)} - Ax_i^{2,2}\| \leq \kappa_2(V_{2,i+1})(\|b^{(2)} - A\bar{x}_i^{2,2}\| + \delta),$$

where

$$\delta = \sum_{k \in \mathcal{I}} |\lambda_k \tilde{p}(\lambda_k)| \|\hat{s}_j^* P_m e_0^{1,2}\|,$$

\tilde{p} is a particular i -degree polynomial with constant term 1, $e_0^{1,2} \equiv x^{(2)} - x_0^{1,2}$, and $P_m \equiv I - V_{1,m}(\bar{T}_{1,m}^* \bar{T}_{1,m})^{-1} \bar{T}_{1,m} D_{1,m+1}^{-1} W_{1,m+1}^T A$ is a projector.

Proof. Since A is diagonalizable, $e_0^{2,2} \equiv x^{(2)} - x_0^{2,2} = \sum_{k=1}^N \phi_k \hat{z}_k$ for some expansion coefficients ϕ_k . Hence, by definition

$$(3.7) \quad b^{(2)} - Ax_0^{2,2} = \sum_{k=1}^N \phi_k \lambda_k \hat{z}_k,$$

$$(3.8) \quad b^{(2)} - A\bar{x}_0^{2,2} = \sum_{k \notin \mathcal{I}} \phi_k \lambda_k \hat{z}_k.$$

²We caution the reader that our notation is somewhat unconventional, as we use n simply to denote an index $n \leq N$ and N to denote the dimension of the matrix.

Now if $\bar{x}_i^{2,2}$ is the i th GMRES iterate with $\bar{x}_0^{2,2}$ as the initial guess, there exists a polynomial \tilde{p}_i of degree less than or equal to i with $\tilde{p}_i(0) = 1$ such that

$$(3.9) \quad b^{(2)} - A\bar{x}_i^{2,2} = \sum_{k \notin \mathcal{I}} \phi_k \lambda_k \tilde{p}_i(\lambda_k) \hat{z}_k,$$

where \tilde{p}_i satisfies

$$(3.10) \quad \tilde{p}_i = \arg \min_{p \in \bar{\Pi}^i} \|p(A)(b^{(2)} - A\bar{x}_0^{2,2})\|.$$

Here, $\bar{\Pi}^i$ denotes the set of all polynomials with degree less than or equal to i with constant term 1. From Theorem 7.1 in [22], we have a bound on the i th QMR residual in terms of the i th GMRES residual:

$$(3.11) \quad \|b^{(2)} - A\bar{x}_i^{2,2}\| \leq \kappa_2(V_{2,i+1}) \min_{p \in \bar{\Pi}^i} \|p(A)(b^{(2)} - A\bar{x}_0^{2,2})\|.$$

Also,

$$(3.12) \quad \min_{p \in \bar{\Pi}^i} \|p(A)(b^{(2)} - A\bar{x}_0^{2,2})\| \leq \|\tilde{p}_i(A)(b^{(2)} - A\bar{x}_0^{2,2})\| = \left\| \sum_{k=1}^N \phi_k \lambda_k \tilde{p}_i(\lambda_k) \hat{z}_k \right\|.$$

Substituting this into (3.11) and using (3.9) and the triangle inequality gives

$$(3.13) \quad \|b^{(2)} - A\bar{x}_i^{2,2}\| \leq \kappa_2(V_{2,i+1}) \left(\|b^{(2)} - A\bar{x}_i^{2,2}\| + \sum_{k \in \mathcal{I}} |\lambda_k \tilde{p}_i(\lambda_k)| |\phi_k| \right).$$

Using the definition of $\bar{x}_0^{2,2}$ as $x_m^{1,2}$, it is straightforward to show

$$e_0^{2,2} = (I - V_{1,m}(\bar{T}_{1,m}^* \bar{T}_{1,m})^{-1} \bar{T}_{1,m}^* D_{1,m+1}^{-1} W_{1,m+1}^T A) e_0^{1,2} = P_m e_0^{1,2}.$$

It is easy to see that P_m is a projector since $(P_m)^2 = P_m$. By the definition of $e_0^{2,2}$ it follows that $P_m e_0^{1,2} = \sum_{k=1}^N \phi_k \hat{z}_k$. Using $\hat{s}_j^* \hat{z}_k = 1$ if $j = k$ and 0 otherwise leads to

$$(3.14) \quad |\hat{s}_j^* P_m e_0^{1,2}| = |\phi_j| \quad j \in \mathcal{I}.$$

Substituting (3.14) into (3.13), we obtain the desired result. \square

Now let us discuss why we expect δ to be small. First, it is clear that in exact arithmetic the Ritz vectors lie in $\mathcal{N}(P_m)$. For any vector $q \in \mathcal{C}^N$, since Z is full rank we may write $q = \hat{Z}c$ for the vector of expansion coefficients $c = \hat{Z}^{-1}q = \hat{S}q$ with components $c_j = \hat{s}_j^* q$. Now suppose $q \in \mathcal{R}(P_m)$. Since the columns of Z with indices in \mathcal{I} are approximated by n of the m Ritz vectors, by assumption, and since the Ritz vectors are in $\mathcal{N}(P_m)$, this implies $c_j \approx 0, j \in \mathcal{I}$, which by definition means $\hat{s}_j^* q \approx 0, j \in \mathcal{I}$. Therefore, with $q \equiv P_m e_0^{1,2}$, $|\phi_j| \approx 0, j \in \mathcal{I}$, so δ should be small in exact arithmetic. It is clear that the quality of the Ritz vector approximation and loss of biorthogonality (e.g., the actual $\mathcal{N}(P_m)$) will affect the size of δ in practice.

Using the definition of \tilde{p}_i in (3.10), we obtain in analogy with the standard GMRES upper bound for diagonalizable matrices [22] the following corollary.

COROLLARY 3.3. *Let \hat{Z}_n denote the $N \times (N - n)$ matrix with columns \hat{z}_j for $j \notin \mathcal{I}$. Then with δ defined as in Theorem 3.2 and P being the $(N - n) \times N$ matrix whose rows are the transposed unit vectors $e_k^T, k \notin \mathcal{I}$,*

$$\|b^{(2)} - A\bar{x}_i^{2,2}\|_2 \leq \kappa_2(V_{2,i+1}) \left(\min_{p \in \bar{\Pi}^i} \max_{\substack{\lambda_k \\ k \notin \mathcal{I}}} |p(\lambda_k)| \|\hat{Z}_n\| \|P \hat{S} \bar{r}_0^{2,2}\| + \delta \right).$$

Proof. The proof follows from the result of the theorem by first writing (3.9) as $\hat{Z}_n \tilde{p}(\hat{\Lambda}) \sum_{k \notin \mathcal{I}} \phi_k \lambda_k e_k$, where $\hat{\Lambda}$ is $\text{diag}(\lambda_k)_{k \notin \mathcal{I}}$, using the identity $e_k = P \hat{S} \hat{z}_k$, and taking norms. \square

Thus, under the assumptions we stated at the beginning of this section so that n right eigenvectors with indices in \mathcal{I} of A have been captured when the first seed system is solved, the second seed system converges in exact arithmetic as if the corresponding part of the spectrum of A has been cut off. The strength of this statement in practice is based on the size of δ , which is affected in finite precision arithmetic as noted above.

In the next theorem, we bound the residual norms of the nonseed systems. A proof and detailed discussion of the size of the upper bound are given in [13]. In short, when the right-hand sides are close and the quasi-residual of the seed system is being efficiently reduced, so are the residuals of the nonseed system.

THEOREM 3.4. *Let j denote the index of the seed system and l denote the index of a (nonconverged) nonseed system. Then*

$$(3.15) \quad \|r_k^{j,l}\| \leq \sqrt{k+1} \left(|\gamma_1^{j,l}| \left| \prod_{i=0}^{k-1} s_{j,k-i} \right| + \sum_{i=0}^{k-1} |\gamma_{k-i+1}^{j,l}| |c_{j,k-i}| \left| \prod_{m=0}^{i-1} s_{j,k-m} \right| \right) + \sqrt{N-k-1} \|h_k^{j,l}\|,$$

where $h_k^{j,l} = [\gamma_{k+2}^{j,l}, \dots, \gamma_N^{j,l}]^T$ and $s_{j,k}, c_{j,k}$ are as defined in section 3.1.

4. BL-QMR background. The BL-QMR algorithm of Freund and Malhotra attempts to solve (1.1) in the following way. First, given K vectors r_i and p vectors l_i , they define

$$R = [r_1, \dots, r_K], \quad L = [l_1, \dots, l_p].$$

The block Krylov sequences generated by R, A and L, A^T are

$$(4.1) \quad \{R, AR, A^2 R, \dots, A^{j-1} R, \dots\} \quad \text{and} \quad \{L, A^T L, \dots, (A^T)^{j-1} L, \dots\}.$$

However, if $A^{j-1} r_i$ (likewise $(A^T)^{j-1} l_i$) is linearly or nearly linearly dependent on the previous vectors, so are all $A^k r_i$ (likewise $(A^T)^k l_i$) for $k \geq j$. Thus, Freund and Malhotra propose scanning the vectors in the two sequences in (4.1) from left to right and deleting those which are linearly or nearly linearly dependent on previous ones. In the process they obtain deflated Krylov sequences whose vectors are linearly independent. We refer to the n -dimensional subspaces generated by these deflated sequences as $K_n^{\text{dl}}(A, R)$ and $K_n^{\text{dl}}(A^T, L)$. Note that in the presence of no deflation, $K_n^{\text{dl}}(A, R)$ and $K_n^{\text{dl}}(A^T, L)$ are spanned by the first n columns of (4.1) with $n \leq jK$ or $n \leq jp$, respectively.

Within BL-QMR is a Lanczos-type algorithm which incorporates the deflation as mentioned above in order to generate biorthogonal bases for $K_n^{\text{dl}}(A, R)$ and $K_n^{\text{dl}}(A^T, L)$: that is, two sequences of right and left Lanczos vectors

$$v_1, \dots, v_n \quad \text{and} \quad w_1, \dots, w_n, \quad n = 1, 2, \dots,$$

such that

$$(4.2) \quad \text{span}\{v_1, \dots, v_n\} = K_n^{\text{dl}}(A, R), \quad \text{span}\{w_1, \dots, w_n\} = K_n^{\text{dl}}(A^T, L),$$

$$w_j^T v_k = \begin{cases} 0 & \text{if } j \neq k, \\ \delta_j \neq 0 & \text{if } j = k. \end{cases}$$

Defining the $N \times n$ matrices $V_n = [v_1, \dots, v_n]$ and $W_n = [w_1, \dots, w_n]$, it follows that

$$W_n^T V_n = D_n \equiv \text{diag}(\delta_1, \dots, \delta_n), \quad n = 1, 2, \dots$$

Also, the matrix equation relating the v 's is

$$AV_\mu = V_n T_\mu + \hat{V}_\mu^{\text{dl}}, \quad \mu \geq 1,$$

where $\mu = n - m_{\text{cr}}$ and m_{cr} is defined by the fact that $K - m_{\text{cr}}$ is the total number of deflations performed in the v sequence up to iteration n in the Lanczos algorithm. Further, T_μ is $n \times \mu$ with lower bandwidth $K + 1$ and upper bandwidth $p + 1$. Also, $\hat{V}_\mu^{\text{dl}} = V_\mu^{\text{dl}} + E_\mu$, where V_μ^{dl} is $N \times \mu$ but has only $K - m_{\text{cr}}$ nonzero columns corresponding to vectors that are deflated and E_μ has nonzero entries in row i column $p + j$, $j = 1, \dots$, only if a deflation of the i th w occurs for $i > K$. We note that if deftol is the deflation tolerance, then $\|V_\mu^{\text{dl}}\| \leq \text{deftol} \sqrt{K - m_{\text{cr}}}$. For further details, the reader is referred to [8].

Now let us assume $R = [r_0^{(1)}, r_0^{(2)}, \dots, r_0^{(K)}]$; that is, the matrix R contains the initial residuals of each of the K systems we would like to solve. Thus, the v 's correspond to the initial residuals. The way the deflation strategy in [8] works is that if a v is deflated, one linear system is also set aside; then upon convergence of the remaining systems, the solution to the deflated system is updated using the solutions of the other systems. Thus, in what follows we consider only the updates to the nondeflated linear systems, and we denote with a subscript “cr” submatrices of the originals with m_{cr} columns that correspond to these systems.

Recall that when QMR is applied to a single linear system, the μ th iterate is an appropriate linear combination of the Lanczos vectors, plus the initial guess. Similarly, the block QMR iterate is defined as

$$X_{\mu, \text{cr}} = X_{0, \text{cr}} + V_\mu Z, \quad Z \in \mathcal{C}^{\mu \times m_{\text{cr}}}.$$

As with QMR, then, we need to find the matrix Z which determines the appropriate linear combination. Following [8], the residual block $R_{\mu, \text{cr}}$ related to $X_{\mu, \text{cr}}$ satisfies

$$\begin{aligned} R_{\mu, \text{cr}} &= B_{\text{cr}} - AX_{\mu, \text{cr}} \\ &= R_{0, \text{cr}} - AV_\mu Z \\ &= R_{0, \text{cr}} - V_n T_\mu Z - \hat{V}_\mu^{\text{dl}} Z \\ &= V_n \left(\begin{bmatrix} \beta_{\text{cr}} \\ 0 \end{bmatrix} - T_\mu Z \right) - \hat{V}_\mu^{\text{dl}} Z, \end{aligned}$$

where β_{cr} is $m_1 \times m_{\text{cr}}$ defined by taking the appropriate columns of β , with

$$V_{m_1} \beta + V_0^{\text{dl}} = R,$$

and m_1 is the number of columns of R (recall R has K columns) that are not deflated as the first K Lanczos vectors are created ($m_1 \leq K$). Here B contains the $b^{(j)}$'s as its columns, and B_{cr} is the submatrix of B with the appropriate m_{cr} columns.

Because the columns of V_n are not unitary and \hat{V}_μ^{dl} has nonzero columns, one cannot find Z such that $\|R_{\mu, \text{cr}}\|$ is minimal. Rather, we seek $Z = Z_\mu$ such that

$$Z_\mu = \arg \min_{Z \in \mathcal{C}^{\mu \times m_{\text{cr}}}} \left\| \begin{bmatrix} \beta_{\text{cr}} \\ 0 \end{bmatrix} - T_\mu Z \right\|.$$

Since T_μ is banded, the standard approach based on the QR factorization of T_μ is used to implicitly determine Z_μ and ultimately determine short-term recurrences for $X_{\mu,\text{cr}}$. Following [8] we have

$$T_\mu = (Q^{(\mu)})^* \begin{bmatrix} R^{(\mu)} \\ 0 \end{bmatrix}$$

for a unitary $n \times n$ matrix $Q^{(\mu)}$ and a nonsingular, $\mu \times \mu$, upper triangular matrix $R^{(\mu)}$. Thus,

$$Z_{(\mu)} = (R^{(\mu)})^{-1} t_\mu, \quad \text{where} \quad \begin{bmatrix} t_\mu \\ \tau_\mu \end{bmatrix} = Q^{(\mu)} \begin{bmatrix} \beta_{\text{cr}} \\ 0 \end{bmatrix}.$$

Finally,

$$\begin{aligned} X_{\mu,\text{cr}} &= X_{0,\text{cr}} + V_{(\mu)} (R^{(\mu)})^{-1} t_\mu \\ (4.3) \quad &= X_{\mu-1,\text{cr}} + p_\mu y_\mu^T, \end{aligned}$$

where p_μ and y_μ^T are given by (refer to [8, equations (5.10), (5.8)])

$$(4.4) \quad p_\mu = \left(v_\mu - \sum_{i=j^*}^{\mu-1} p_i \theta_i \right) / \theta_\mu, \quad \begin{bmatrix} y_\mu^T \\ \tau_\mu \end{bmatrix} = Q_\mu \begin{bmatrix} \tau_{\mu-1} \\ 0 \end{bmatrix}.$$

The θ_i are scalars corresponding to the last column of $R^{(\mu)}$, Q_μ (not to be confused with $Q^{(\mu)}$) is a particular matrix of Givens rotations described in (5.2) of [8], and j^* is an index described in [8] satisfying $(\mu - j^*) \leq 2m$.

5. The block QMR-projection method. In a manner similar to section 3, we describe a block QMR-projection approach to solving (1.1) that combines the advantageous properties of the block QMR algorithm and our single-seed projection algorithm.

Suppose that we select a subset of size $m < K$ linear systems to serve as “seed” from among the original K . Let \mathcal{I}_{m_1} be the index set i_1, \dots, i_m of the chosen systems. We use $\mathcal{I}_{m_1}^c$ to denote the indices from 1 to K which are not in \mathcal{I}_{m_1} . Let $b^{(j)}$ with $j \in \mathcal{I}_{m_1}$ be the m columns of the matrix $B^{(1)}$ and let the remaining $J = K - m$ right-hand sides (corresponding to nonseed systems indexed by $\mathcal{I}_{m_1}^c$) be the columns of the matrix $B^{(2)}$. We define $X_0^{(1)}$ as the matrix $[x_0^{(i_1)}, \dots, x_0^{(i_m)}]$ of initial guesses for the m seed systems, and $X_0^{(2)}$ as the matrix of initial guesses for the nonseed systems. The corresponding initial block residuals are $R_0^{(1)} = B^{(1)} - AX_0^{(1)}$ and $R_0^{(2)} = B^{(2)} - AX_0^{(2)}$.

The idea is to set R (and L) defined in the previous section to $R_0^{(1)}$ and run BL-QMR to solve the seed systems while using a projection idea to update the nonseed systems. Once BL-QMR converges on the seed system, the process is repeated by choosing a new subset, indexed by $\mathcal{I}_{m_2} \subset \mathcal{I}_{m_1}^c$, of the nonconverged, nonseed systems. The systems indexed by \mathcal{I}_{m_2} now serve as seed, where the columns of $X_0^{(1)}$ are understood to be the estimated solutions, generated in the first round of projected BL-QMR, to the systems with indices in \mathcal{I}_{m_2} . The remaining systems, indexed by $\mathcal{I}_{m_2}^c = \mathcal{I}_{m_1}^c \setminus \mathcal{I}_{m_2}$, are updated by projection. In the following, $X_\mu^{(1)}$ ($R_\mu^{(1)}$) denotes the μ th block iterate (residual) of the current block seed while $X_\mu^{(2)}$ ($R_\mu^{(2)}$) denotes the μ th block iterate (residual) of the current nonseed block. We shall further assume that m

is the number of *current* seed systems and J is the number of *current* nonconverged, nonseed systems. The numbers m and J can change at each round.

At iteration μ , we want our nonseed systems to also lie in the current Krylov subspace. That is, we desire

$$X_\mu^{(2)} \in X_0^{(2)} + K_\mu^{\text{dl}}(A, R_0^{(1)}).$$

Since the columns of V_μ span this subspace, this means

$$X_\mu^{(2)} = X_0^{(2)} + V_\mu Z_\mu^{(2)}$$

for some $\mu \times J$ matrix $Z_\mu^{(2)}$. Now we must decide how to define $Z_\mu^{(2)}$. We observe

$$\begin{aligned} R_\mu^{(2)} &= B^{(2)} - A(X_0^{(2)} + V_\mu Z_\mu^{(2)}) \\ &= R_0^{(2)} - V_n T_\mu Z_\mu^{(2)} - \hat{V}_\mu^{\text{dl}} Z_\mu^{(2)}. \end{aligned}$$

Using biorthogonality

$$D_n^{-1} W_n^T R_\mu^{(2)} = D_n^{-1} W_n^T R_0^{(2)} - T_\mu Z_\mu^{(2)} - D_n^{-1} W_n^T \hat{V}_\mu^{\text{dl}} Z_\mu^{(2)}.$$

Then

$$\|D_n^{-1} W_n^T R_\mu^{(2)}\| \leq \|D_n^{-1} W_n^T R_0^{(2)} - T_\mu Z_\mu^{(2)}\| + \|D_n^{-1} W_n^T \hat{V}_\mu^{\text{dl}} Z_\mu^{(2)}\|.$$

Note that if no deflations have occurred, \hat{V}_μ^{dl} is zero, so we have equality rather than inequality. Therefore, in analogy with the single-seed algorithm of section 3, we define

$$Z_\mu^{(2)} \equiv \arg \min_{Z \in \mathbb{C}^{\mu \times J}} \|D_n^{-1} W_n^T R_0^{(2)} - T_\mu Z\|.$$

Using the QR factorization of T_μ described in the previous section, we obtain

$$Z_\mu^{(2)} = \arg \min_{Z \in \mathbb{C}^{\mu \times J}} \left\| Q^{(\mu)} G_n - \begin{bmatrix} R^{(\mu)} \\ 0 \end{bmatrix} Z \right\|,$$

where G_n is the $n \times J$ matrix $G_n = D_n^{-1} W_n^T R_0^{(2)}$. If

$$(5.1) \quad \begin{bmatrix} t_\mu^{(2)} \\ \tau_\mu^{(2)} \end{bmatrix} = Q^{(\mu)} G_n,$$

we obtain

$$(5.2) \quad Z_\mu^{(2)} = (R^{(\mu)})^{-1} t_\mu^{(2)},$$

so that

$$(5.3) \quad \|D_n^{-1} W_n^T R_0^{(2)} - T_\mu Z_\mu^{(2)}\| = \|\tau_\mu^{(2)}\|.$$

Using $G_n^T = [G_{n-1}^T, g_n]$, together with (5.1) and the definition of $Q^{(\mu)}$ in (5.1) of [8], it is easy to show that

$$(5.4) \quad \begin{bmatrix} t_\mu^{(2)} \\ \tau_\mu^{(2)} \end{bmatrix} = \begin{bmatrix} I_{\mu-1} & 0 \\ 0 & Q_\mu \end{bmatrix} \begin{bmatrix} t_{\mu-1}^{(2)} \\ \tau_{\mu-1}^{(2)} \\ g_n^T \end{bmatrix} = \begin{bmatrix} t_{\mu-1}^{(2)} \\ Q_\mu \begin{bmatrix} \tau_{\mu-1}^{(2)} \\ g_n^T \end{bmatrix} \end{bmatrix}.$$

Thus, $t_\mu^{(2)}$ differs from $t_{\mu-1}^{(2)}$ only in its last row, which we call $(y_\mu^{(2)})^T$:

$$t_\mu^{(2)} = \begin{bmatrix} t_{\mu-1}^{(2)} \\ (y_\mu^{(2)})^T \end{bmatrix}, \text{ where } (y_\mu^{(2)})^T \in \mathcal{C}^{1 \times J}.$$

From the above relation and (5.4) it follows that to obtain $(y_\mu^{(2)})^T$ one need only perform a product with Q_μ :

$$\begin{bmatrix} (y_\mu^{(2)})^T \\ \tau_\mu^{(2)} \end{bmatrix} = Q_\mu \begin{bmatrix} \tau_{\mu-1}^{(2)} \\ g_n^T \end{bmatrix},$$

which, since Q_μ by definition is a product of m_{cr} Givens rotations, is an easy task.

With p_i defined as in (4.4), it is now easy to show that the μ th nonseed block iterate is

$$(5.5) \quad X_\mu^{(2)} = X_{\mu-1}^{(2)} + p_\mu (y_\mu^{(2)})^T.$$

Thus, we may readily show

$$(5.6) \quad R_\mu^{(2)} = R_{\mu-1}^{(2)} - A p_\mu (y_\mu^{(2)})^T.$$

However, using the definition of p_μ , we find an update formula for the block residual which does not actually require any additional matrix-vector products.

LEMMA 5.1. *$R_\mu^{(2)}$ can be updated from $R_{\mu-1}^{(2)}$ in at most $O(N(J+2m))$ additional floating point operations.*

Proof. By substituting (4.4) into (5.6), we obtain a formula for updating $R_\mu^{(2)}$:

$$(5.7) \quad R_\mu^{(2)} = R_{\mu-1}^{(2)} - f_\mu (y_\mu^{(2)})^T \quad \text{with} \quad f_i \equiv A p_i = \frac{1}{\theta_i} \left(A v_i - \sum_{k=j^*}^{i-1} \theta_k f_k \right).$$

Consider forming f_μ . Now the matrix-vector product $A v_\mu$ is computed in the course of the Lanczos process at iteration μ and need not be recomputed. Therefore, it is clear that to compute the length N vector f_μ requires at most $O(2mN)$ flops since $(\mu - j^*) \leq 2m$ by definition (see section 5 of [8]). We note that the computation of the outer product $f_\mu (y_\mu^{(2)})^T$ requires $O(JN)$ operations, and the proof is complete. \square

We note that a similar update is valid for $R_\mu^{(1)}$:

$$(5.8) \quad R_\mu^{(1)} = R_{\mu-1}^{(1)} - f_\mu (y_\mu^{(1)})^T.$$

6. Issues in practical computation for the block algorithm.

6.1. Block-seed selection heuristic. Clearly, the performance of our multiple-seed algorithm in terms of savings of matrix-vector products depends on which, and how many, systems are chosen to be seed. Deflation in BL-QMR solves the problem of removing redundancy if systems with starting residuals which are nearly linearly dependent are chosen as seed. Ideally, however, we would like to choose as seed systems some subset of the nonconverged systems which are in some sense optimally independent in order to increase the chance that the solutions to the nonseed systems will lie nearly in the Krylov subspaces generated by the seed systems, thereby ensuring the effectiveness of the projection process.

In our examples, we used the following heuristics to determine which and how many seed systems to use. First, we let $B = [b^{(1)}, \dots, b^{(K)}]$. Since $K \ll N$ was not too large in these examples, we computed a compact pivoted QR factorization of B ,

$$\tilde{B} \equiv B\Pi = QR, \quad Q \in \mathcal{C}^{N \times K}, \quad R \in \mathcal{C}^{K \times K},$$

to determine which of the remaining were most independent. Here, Π is just a permutation matrix which serves to permute the columns of B such that the first few columns of $B\Pi$ are the most independent. In particular, if ρ denotes the diagonal entries of R and if $|\rho(1)|/|\rho(i)| > \alpha$ for any $1 \leq i \leq K$, then we discard the corresponding column of \tilde{B} . The remaining m columns of \tilde{B} serve as the seed systems. We set the maximum value of α to 10^5 to ensure the columns were not too linearly dependent, but adjusted it lower if necessary so that the size of the seed block was no bigger than about $K/2$. On the next round of projected BL-QMR, however, we simply decided on a new number of seeds to use ($m \leftarrow \lceil m/2 \rceil$) and took those with the largest m relative residuals to serve as seed. More efficient means of selecting m for each round and for determining the m seeds need to be examined in the future.³

6.2. Loss of biorthogonality. One additional problem that we encountered in practice in using either our single-seed or our multiple-seed algorithm was that loss of biorthogonality could affect the accuracy of the $\gamma_n^{jl} = (1/\delta_{j,n})w_{j,n}^T r_0^{jl}$, or $g_n^T = (1/\delta_n)w_n^T R_0^{(2)}$. This loss of accuracy would thereby adversely affect the convergence of the computed solution. To avoid this difficulty for the block projection algorithm, we used the following fact. If no deflations were performed up to the μ th iteration when solving the single-seed system,

$$R_{\mu-1}^{(2)} = R_0^{(2)} - V_{n-1}T_{\mu-1}Z_{\mu-1}^{(2)} \Rightarrow g_n^T = \frac{1}{\delta_n}w_n^T R_{\mu-1}^{(2)} = \frac{1}{\delta_n}w_n^T R_0^{(2)},$$

where it is understood that $R_j^{(2)} = R_0^{(2)}$, $j < 0$. Thus, at the beginning of iteration $\mu \geq 1$, we computed g_n^T based on the current residual estimate, then updated the residual estimate using Lemma 5.1. If deflations do occur, observe

$$g_n^T = \frac{1}{\delta_n}w_n^T R_{\mu-1}^{(2)} - \frac{1}{\delta_n}w_n^T \hat{V}_{\mu-1}^{\text{dl}} Z_{\mu-1}^{(2)}.$$

In our examples, the second term was on the order of the deflation tolerance. This was because $\hat{V}_{\mu}^{\text{dl}} = V_{\mu}^{\text{dl}}$ since no w deflations occurred for indices larger than J . Hence nonzero columns of V_{dl} were nearly linear combinations of the first m_1 v 's for which $\frac{1}{\delta_n}w_n^T v = 0$. In this work we choose to ignore the second term rather than go to the extra expense of computing inner products with the nonzero columns of \hat{V}_{dl} .

Likewise, for the single-seed algorithm we use

$$\gamma_n^{jl} = \frac{1}{\delta_{j,n}}w_{j,n}^T r_0^{jl} = \frac{1}{\delta_{j,n}}w_{j,n}^T r_{n-2}^{jl}, \quad n \geq 2.$$

An investigation into the reason behind the success of these approaches in finite precision arithmetic will be the subject of future work. We note that a similar phenomenon was observed in [20] with respect to practical implementation of GMRES variants, and an explanation for such behavior in finite precision arithmetic was given.

³In the worst case, if the seed block is too small, then it could require many rounds and much computation for all the systems to converge, and performance could be worse than BL-QMR without projection. If the seed block is too large, gains in execution time over BL-QMR would probably also be reduced, and our algorithm's behavior would become more dependent on the deflation tolerance.

6.3. Computational aspects of the block-seed algorithm. It may be relatively expensive in terms of execution time to move data between the smaller, faster cache and the larger, slower main memory. When data is available in cache, it is desirable to use it as much as possible. The level-2 and level-3 BLAS are better for achieving this than level-1 BLAS operations. Thus, one advantage that the block Krylov subspace algorithms enjoy over standard Krylov subspace algorithms (and our single-seed algorithm) is that the former can be implemented to be rich in higher level BLAS operations, whereas the latter class of algorithms requires a large number of level-1 BLAS operations [14, 2].

Further, Krylov subspace algorithms require a large number of vector inner products relative to the remaining number of computations. These inner products, when implemented on a distributed memory parallel machine, correspond to synchronization points (that is, computation cannot proceed until all processors receive the result of the inner product) and require numerous smaller messages among processors [17]. Our single-seed method inherits these problems, although updates to the seed and nonseed systems can be done independently. Block Krylov subspace algorithms, however, can be implemented to provide more computation between communications and larger but fewer messages among processors [14, 17]. Below, we provide one implementation of the block-seed projection algorithm from the proceeding section. We do not claim that this implementation is optimal in terms of cache utilization or parallelism; our goal is to illustrate the potential efficiencies of the block-seed algorithm and show that it retains the same advantages that the block QMR algorithm enjoys.

Suppose that m is the number of right-hand sides in a given seed block and that J is the number of systems in the nonseed block. Let $m_1 \leq m$ be the number of linearly independent right Lanczos vectors that are formed, using deflation, from the initial residuals to the seed block. In the following, $V_{(k)} = [v_{m_1 k+1}, \dots, v_{m_1(k+1)}]$ and $W_{(k)} = [w_{m_1 k+1}, \dots, w_{m_1(k+1)}]$, $0 \leq k$.

ALGORITHM 1. $\mu = 1 = \phi$. Given $X_0^{(1)}, R_0^{(1)} \in C^{N \times m_1}$ and $X_0^{(2)}, R_0^{(2)} \in C^{N \times J}$; Given the m_1 columns of $V_{(1)}$ and $W_{(1)}$ and $D_{(1)} = \text{diag}(\delta_1, \dots, \delta_{m_1})$.

For $k = 2$ until seed block converges do

1. $V_{(k)} = AV_{(k-1)}$; $W_{(k)} = A^T W_{(k-1)}$.
2. If deflations occurred in the W (or V) sequence, update $V_{(k)}$ (or $W_{(k)}$).
3. Biorthogonalize the columns of $V_{(k)}$ against the columns of $V_{(k-2)}, V_{(k-1)}$; biorthogonalize the columns of $W_{(k)}$ against the columns of $W_{(k-2)}, W_{(k-1)}$.
4. Set $i = 0, j = 0, s = 1, \hat{s} = 1$. Set $\mu = \mu + 1$; $\phi = \phi + 1$.
5. For $n = (k-1)m_1 + 1, \dots, km_1$, set $i = i + 1$ and do
 - (a) If $V_{(k)}(:, i)$ does not exist, put $V_{(k)}(:, i) = AV_{(k)}(:, s)$ and biorthogonalize against ϕ_μ previous Lanczos pairs; $s = s + 1$.
 - (b) If $\|V_{(k)}(:, i)\| < \text{deftol}$, then $j = j + 1$ and deflate:
 - i. Compute $(y_\mu^{(1)})^T, (y_\mu^{(2)})^T, \tau_\mu^{(1)}, \tau_\mu^{(1)}$; compute f_μ, p_μ (via gaxpy's).
 - ii. "Delete" i th column of $V_{(k)}$, "shift" remaining columns left 1.
 - iii. Compute which system gets dropped from the seed block (that row of $y_\mu^{(1)}$ will have a zero entry).
 - iv. Let $y_\mu^{(i_v)}, f_\mu, p_\mu$ be the j th columns of $Y^{(i_v)}, i_v = 1, 2; F; P$.
 - v. Update deflation index sets. Set $\mu = \mu + 1$, goto step 5(a).
 - (c) Normalize $V_{(k)}(:, i)$ and set $\mu_n = \mu$; $j = j + 1$.
 - (d) If $W_{(k)}(:, i)$ does not exist, put $W_{(k)}(:, i) = A^T W_{(k)}(:, \hat{s})$ and biorthogonalize against μ_ϕ previous Lanczos pairs; $\hat{s} = \hat{s} + 1$.
 - (e) If $\|W_{(k)}(:, i)\| < \text{deftol}$, then deflate:

- i. “Delete” i th column of $W_{(k)}$, “shift” remaining columns left 1.
- ii. Update deflation index sets. Set $\phi = \phi + 1$, goto step 5(d).
- (f) Normalize $W_{(k)}(:, i)$. Set $\phi_n = \phi$ and $\delta_n = W_{(k)}(:, i)^T V_{(k)}(:, i)$.
- (g) Continue the MGS⁴ process on columns $\geq i$ of $V_{(k)}, W_{(k)}$.
- (h) Set $g_n^T = (1/\delta_n)W_{(k)}(:, i)^T R_{\mu-j}^{(2)}$.
- (i) Compute $(y_\mu^{(1)})^T, (y_\mu^{(2)})^T, \tau_\mu^{(1)}, \tau_\mu^{(2)}$; compute f_μ, p_μ (via gaxpy’s).
- (j) Let $y_\mu^{(i_v)}, f_\mu p_\mu$ be the j th columns of $Y^{(i_v)}, i_v = 1, 2; F; P$.
- 6. $X_\mu^{(1)} = X_{\mu-j}^{(1)} + P(Y^{(1)})^T$; $R_\mu^{(1)} = R_{\mu-j}^{(1)} - F(Y^{(1)})^T$.
- 7. $X_\mu^{(2)} = X_{\mu-j}^{(2)} + P(Y^{(2)})^T$; $R_\mu^{(2)} = R_{\mu-j}^{(2)} - F(Y^{(2)})^T$.
- 8. “Remove” deflated systems from $X_\mu^{(1)}$.

One benefit of this implementation is that A, A^T are accessed only once each block iteration, if no deflation occurs, and computing products of A and A^T with dense, rectangular matrices of Lanczos vectors makes better use of cache than products of A with a single vector. In step 2 (also 5(a), 5(d)) some columns of either or both of the current blocks may have to be biorthogonalized against some previous Lanczos vectors if certain previous deflations occurred in the V and/or W sequence. In step 3, the current blocks of Lanczos vectors are biorthogonalized against the appropriate columns⁵ of the previous two blocks. Considering the matrix $V_{(k)}$ (or $W_{(k)}$) rather than its columns separately, we can do this using level-2 BLAS with a two-sided modified Gram–Schmidt approach or, at the expense of some stability, we could accomplish this with level-3 BLAS via a block modified Gram–Schmidt approach [14]. For each deflation step, however, we incur the price of one matrix-vector product and several vector-wise inner products. It is possible to reorganize the algorithm so that a Lanczos block effectively decreases in size after deflation (possibly leaving left and right blocks of different sizes) and thereby put off this extra work until it can be done with higher-level BLAS, but as the notation is more tedious, we use the current implementation to illustrate our points.

Notice we are using modified Gram–Schmidt to biorthogonalize within the current block, but that computation toward updating the solution and residual blocks is done between each step of the process. Notice also that the block iterates and residuals are only updated after a new block of Lanczos vectors has been generated; this was done in order to minimize the number of accesses to the block iterates and residuals and to incorporate level-3 BLAS operations. The updating could be done (via level-2 BLAS operations) inside the innermost loop according to (4.3), (5.5), (5.7), (5.8), or one might opt to track the size of $\tau_\mu^{(1)}$ and update only when necessary.

One way to implement the algorithm on a distributed memory parallel machine is to row partition [17, 14] the matrices $F, P, V_{(k)}, W_{(k)}, X_\mu^{(i_v)}, R_\mu^{(i_v)}, i_v = 1, 2$. Thus, the matrix-multiplications with A, A^T , the biorthogonalization steps, and steps 5(b), (c), (e), (f), (h) require communication among processors; most of the other steps require only local updates of portions of the rectangular matrices. As in [17, section 3.1], it may be possible in a parallel implementation to exploit any computations that are mostly decoupled: for example, the updates to solution and residual blocks are somewhat independent of the generation of the Lanczos vectors and of each other.

⁴Modified Gram–Schmidt.

⁵If deflations have occurred in the W sequence, then one need only biorthogonalize $V_{(k)}$ against some of the columns of $W_{(k-2)}$ (or $W_{(k-1)}$), rather than against the whole block, and similarly for computing $W_{(k)}$ if deflations occurred in the V sequence.

Techniques for performing matrix-vector products with A, A^T in parallel should also be employed.

Before executing Algorithm 1, using our heuristic in section 6.1, we do a compact pivoted QR factorization of B to determine the seed block and therefore should not incur much overhead beyond what the BL-QMR algorithm in [8] would need to generate its initial block using deflation. Every other time that a seed must be chosen, we select those $\lceil m/2 \rceil$ systems with largest residual norm, so a small amount of additional computations/communication are needed at the seed selection steps.

Clearly, if the first time Algorithm 1 is called $J = 0$, it is just the block QMR algorithm with deflation. If $J \neq 0$, then the first time Algorithm 1 is executed, for a given k , the same number of solution vectors and residual vectors must be updated (at most) as if J had been 0: the difference is that updates are separated into updates on two different solution and residual blocks, and these updates are independent of one another. Thus, if Algorithm 1 is implemented in parallel, the independent updates may help compensate for the execution time that is due to processor communication. The computation of g_n^T and $(y_\mu^{(2)})^T$ are the most notable differences between the unprojected and projected algorithms. Overall, the block-seed projection approach may yield the following advantages over block QMR:

- fewer accesses to A and less communication among processors;
- smaller seed block sizes with the block-projection approach mean fewer flops are needed to compute products with A, A^T and fewer flops are needed to construct Lanczos pairs and to factor T_μ ;
- less storage per processor during a given run of Algorithm 1 (as fewer vectors are needed in the recurrences for p_μ, f_μ and the number of columns of $V_{(k)}, W_{(k)}$, etc., are reduced);
- updating the seed and nonseed blocks can be done independently.

7. Numerical results. In this section we give numerical results which indicate the potential effectiveness of our approach on electromagnetic scattering problems. All experiments were conducted in Matlab using IEEE double precision floating point arithmetic on a 600 MHz Pentium II processor. We compare our results with results from the Matlab implementation of block QMR with deflation, algorithm BL-QMR in [8]. For comparison purposes, the implementation of our block-seed algorithm mirrors the implementation of BL-QMR in [8] with modifications where necessary, rather than the one in the preceding section.

Mathematically, we would like to solve a two-dimensional Helmholtz-type equation for the scattered electric field $E(x, y)$:

$$(7.1) \quad (\Delta + k^2(x, y))E(x, y) = \chi_m(x, y)E_0(x, y) \quad \text{in } \Omega$$

with perfectly matched layer (PML) boundary conditions [3, 18]: the specific mathematical formulation we use is described in [12]. Here $k^2(x, y) = \omega^2 \mu_0 \epsilon(x, y)$ is the square of the wave number, with ω representing angular frequency and μ_0 a constant denoting the magnetic permeability. The function $\epsilon(x, y)$, called the electrical permittivity, is defined as $\epsilon = \epsilon_0 \epsilon_{rel} + i \frac{\sigma}{\omega}$ for some real $\sigma \geq 0$, $\epsilon_{rel} \geq 1$ with $i = \sqrt{-1}$ and ϵ_0 a constant (the permittivity of free-space). The value σ is the conductivity of the material. The function χ_m describes the properties of the buried object and has support only over the object location. $E_0(x, y)$ is the known incident electric field.

We discretize using finite differences, which leads to a matrix equation involving the matrix A which is $N \times N$, sparse, complex, and structured but neither symmetric nor Hermitian due to the boundary conditions. Because the matrix is highly indefinite,

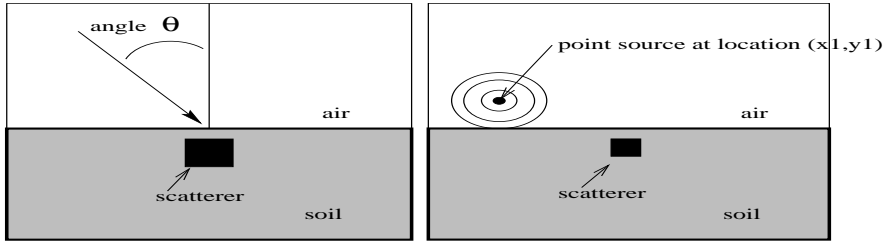


FIG. 7.1. Physical configurations for Example 1 (left) and Example 2 (right).

we need to use a preconditioner to speed convergence. The preconditioner we use is the one described in [12], and we perform all preconditioning from the right.

For all algorithms, we take the initial starting guesses $x_0^{(j)}$ to be zero. We stop running our algorithms when the relative residual norms of all of the systems are less than $tol = 10^{-7}$. For the two block-based algorithms, ours and BL-QMR, we update (seed) block residuals via (5.8) ((5.7) is used for the nonseed block⁶). We monitor convergence of the current seed block by checking norms of the columns of $R_\mu^{(1)}$. However, the true norms of the residuals in the seed block were computed and checked to satisfy the convergence tolerance before the block was deemed to have converged. Since for these examples the major computational expense per iteration is the two matrix-vector products with applications of the preconditioner, we consider the total number of matrix-vector products required for all the systems to converge as our primary measure of success and discuss some timing results.

7.1. Example 1. In this experiment we would like to find the scattered electric fields caused when plane waves at various angles impinge on a horizontal air-soil interface and scatter from a $7\text{cm} \times 6\text{cm}$ object buried 5cm below the surface. Each angle corresponds to a different E_0 in (7.1), which in turn corresponds to a different right-hand side $b^{(j)}$ in (1.1). Figure 7.1 gives a physical illustration of the problem.

In this example, we use a soil type (referred to as “Seabee” in the literature [19]) and conduct experiments at two different frequencies, $\omega/(2\pi) = 45\text{ MHz}$ and 475 MHz . At 45 MHz , Seabee has $\epsilon_{rel} = 35.65$ and $\sigma = .13$, while at 475 MHz $\epsilon_{rel} = 21.31$ and $\sigma = .23$. We assumed that the buried object has $\epsilon_{rel} = 2.9$ and $\sigma = .001$ at both frequencies. For air, $\epsilon_{rel} = 1$, $\sigma = 0$. We have discretized at a rate of 50 points per wavelength of soil at 45 MHz and 20 points per wavelength at 475 MHz . In both cases, the total number of unknowns (N) is $(2^7 + 15)^2$.

We centered the buried object (refer to Figure 7.1) and considered the scattered field due to plane waves impinging on the surface at evenly spaced angles from -60 to 60 degrees from the normal. The second columns in Tables 7.1 and 7.2 give the total number of matrix-vector products needed if preconditioned QMR is applied to each system separately.⁷ The third column gives the total number of matrix-vector products needed if our preconditioned QMR with projection algorithm is used. The next several columns give the total number of matrix-vector products computed when solving the problem using BL-QMR with various deflation tolerances. The final

⁶Note that the latter needs to be computed each iteration to determine the update for the nonseed block, whereas one could use the bound in [8] to monitor convergence of the seed block at the possible expense of computing many extra matrix-vector products.

⁷Note that we define the matrix-vector product count as the number of multiplies by AM^{-1} or its transpose where M is the preconditioner.

TABLE 7.1

Example 1. 45 MHz: Number of matrix-vector products required for convergence by each of the methods (each system independently, single-seed QMR-projection, BL-QMR with deflation tolerances 10^{-8} , 10^{-9} , and 10^{-10} , and block QMR with projection) for experiments involving different numbers of right-hand sides (RHS). Dashes indicate no convergence of the method in under 300 iterations.

No. matrix-vector products, 45 MHz							
No. RHS	Seq.	Prj.	BQ $1E-8$	BQ $1E-9$	BQ $1E-10$	BQ $1E-11$	BQ + Prj.
7	856	292	–	201	219	219	173
13	1590	254	–	211	227	239	180
25	3054	280	–	233	243	267	207

TABLE 7.2

Example 1. 475 MHz: Number of matrix-vector products required for convergence by each of the methods for experiments involving different numbers of right-hand sides.

No. matrix-vector products, 475 MHz							
No. RHS	Seq.	Prj.	BQ $1E-7$	BQ $1E-8$	BQ $1E-9$	BQ $1E-10$	BQ + Prj.
7	664	420	223	223	223	223	221
13	1236	468	–	303	321	319	237
25	2378	492	–	321	319	–	245

column shows results when our block-seed approach is used ($deftol = 10^{-9}$), where the seed blocks are chosen using the heuristic outlined in section 6.1 with $\alpha = 10^5$. Dashes indicate that the convergence tolerance was not met within $maxit = 300$ iterations.

As Table 7.1 shows, for the 45 MHz case, BL-QMR failed to converge after 300 iterations in all cases when the deflation tolerance was set to 10^{-8} , but it outperformed our single-seed projection method if the deflation tolerance was small enough. Our block-seed projection approach performs better than all the other methods in terms of the number of matrix-vector products; however, we note that for these nonoptimized implementations, the execution times for the best BL-QMR runs and our block algorithm are about the same. Solving sequentially took 4.5 times longer than our single-seed method and over 6 times longer than block-based algorithms.

At 475 MHz, we expected our $x^{(j)}$'s not to be as close as in the previous case due to the underlying physics of the problem, and therefore we did not expect as much savings with our single-seed projection approach. Indeed, Table 7.2 shows that the difference between the second and third columns is not as dramatic as in Table 7.1. Table 7.2 also shows that BL-QMR, with the deflation tolerance set at either 10^{-8} or 10^{-9} , outperforms our single-seed projection approach. However, comparing the last column with the others, we find that our block-seed projection approach can provide substantial savings over the other methods. There is also a difference in execution times: for example, for 25 systems our block-seed projection method takes about 6.2 minutes while BL-QMR with deflation tolerance of 10^{-9} takes about 7.1 minutes.

7.2. Example 2. For our second example, each of our K systems corresponds to solving for the scattered electric field from a buried object when the source of the incident field is a point source, located at position x_i, y_i above the earth (see Figure 7.1). We consider the case when the frequency is 480 MHz, and the soil has $\epsilon_{rel} = 6.5$ and $\sigma = .019$. As before, the buried object has $\epsilon_{rel} = 2.9$ and $\sigma = .001$. The buried object is $7\text{cm} \times 4\text{cm}$ buried 5cm deep and centered left to right. The width of each cell in the discrete grid is 1cm, and the total number of unknowns (N) is $(2^6 + 15)^2$. Our point sources are each located 3cm above the earth and either

TABLE 7.3

Example 2. Number of matrix-vector products required for convergence by each of the methods for experiments involving different numbers of right-hand sides. Dashes indicate no convergence of the method in under 300 iterations.

No. RHS	No. matrix-vector products					
	Seq.	Prj.	BQ 1E-9	BQ 1E-11	BQ 1E-12	BL-QMR + Prj.
25	1894	1084	–	–	345	209
35	2658	1468	–	–	411	289

TABLE 7.4

Example 2. Approximate execution time in minutes. Dashes indicate no convergence of the method in under 300 iterations.

No. RHS	Minutes					
	Seq.	Prj.	BQ 1E-9	BQ 1E-11	BQ 1E-12	BQ + Prj.
25	4.2	4.5	–	–	10.6	2.0
35	5.9	7.4	–	–	33.1	3.7

vary in the horizontal direction, with 0 being in the middle, from -12cm to 12cm in 1cm increments or -17cm to 17cm in 1cm increments. The numbers of matrix-vector products required by each of the different methods to solve these systems are given in Table 7.3. However, as illustrated by the timing results in Table 7.4, both the single- and block-seed projection give dramatic improvements over BL-QMR without projection.⁸

8. Conclusions and future work. We have introduced two new projection approaches, based on QMR and block QMR, respectively, for solving multiple linear systems with the same coefficient matrix but different right-hand sides. Compared to solving each system separately by QMR, both approaches can significantly reduce the work and execution time needed to solve all the systems to within a specified tolerance provided there is sufficient shared information among the right-hand sides; the block-seed algorithm requires less shared information to perform well. We provided theory for the single-seed approach which suggests that under certain conditions in exact arithmetic, QMR on subsequent seed systems converges as if part of the spectrum has been cut off; we also gave an upper bound for the rate of convergence of the nonseed systems. More work needs to be done to determine convergence behavior of the block-seed algorithm in both exact and finite precision arithmetic.

As our numerical results showed, with appropriate deflation tolerance, the BL-QMR algorithm [8] could outperform our single-seed QMR-projection method in terms of matrix-vector product savings (although not always reflected in the execution times) particularly when the right-hand sides are not as close; however, our block-seed projection method consistently exhibited the greatest savings in such cases. The performance of our block-seed approach depends on our choices of successive seed blocks, and overall execution time depends on the actual implementation. We gave one block-seed selection heuristic and discussed possible efficiencies of block-seed algorithm. Determining good seed selection strategies, efficient serial and parallel implementations, and formal time comparisons with other methods remain subjects for future research.

⁸Part of the improvement can be attributed to the difference in size of the BL-QMR block iterate with the size of the seed for our methods.

REFERENCES

- [1] Z. BAI, *Error analysis of the Lanczos algorithm for the nonsymmetric eigenvalue problem*, Math. Comp., 62 (1994), pp. 209–226.
- [2] Z. BAI, D. DAY, AND Q. YE, *ABLE: An adaptive block Lanczos method for non-Hermitian eigenvalue problems*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 1060–1082.
- [3] J. BERENGER, *A perfectly matched layer for the absorption of electromagnetic waves*, J. Math. Phys., 114 (1994), pp. 185–200.
- [4] W. E. BOYSE AND A. A. SEIDL, *A block QMR method for computing multiple simultaneous solutions to complex symmetric systems*, SIAM J. Sci. Comput., 17 (1996), pp. 263–274.
- [5] T. F. CHAN AND W. L. WAN, *Analysis of projection methods for solving linear systems with multiple right-hand sides*, SIAM J. Sci. Comput., 18 (1997), pp. 1698–1721.
- [6] J. CULLUM, *Arnoldi versus nonsymmetric Lanczos algorithms for solving nonsymmetric eigenvalue problems*, BIT, 36 (1996), pp. 470–493.
- [7] D. DAY, *An efficient implementation of the nonsymmetric Lanczos algorithm*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 566–589.
- [8] R. FREUND AND M. MALHOTRA, *A block QMR algorithm for non-Hermitian linear systems with multiple right-hand sides*, Linear Algebra Appl., 254 (1997), pp. 197–257.
- [9] R. FREUND AND N. NACHTIGAL, *QMR: A quasi-minimal residual method for non-Hermitian linear systems*, Numer. Math., 60 (1991), pp. 315–339.
- [10] R. W. FREUND AND N. M. NACHTIGAL, *An implementation of the QMR method based on coupled two-term recurrences*, SIAM J. Sci. Comput., 15 (1994), p. 313–337.
- [11] P. JOLY, *Résolution de Systèmes Linéaires avec Plusieurs Seconds Membres par la Méthode du Gradient Conjugué*, Tech. report R-91012, Publications du Laboratoire d'Analyse Numérique, Université Pierre et Marie Curie, Paris, 1991.
- [12] M. KILMER, E. MILLER, AND C. RAPPAPORT, *Preconditioners for Structured Matrix Problems Arising in Subsurface Object Detection*, preprint, <http://www.tufts.edu/~mkilme01>.
- [13] M. KILMER, E. MILLER, AND C. RAPPAPORT, *QMR-Based Projection Techniques for the Solution of Non-Hermitian Systems with Multiple Right Hand Sides*, Tech. report TR-CDSP-00-51, Center for Communications and Digital Signal Processing, Northeastern University, Boston, MA, 2000.
- [14] G. LI, *A block variant of the GMRES method on massively parallel processors*, Parallel Comput., 23 (1997), pp. 1005–1019.
- [15] A. A. NIKISHIN AND A. YU. YEREMIN, *Variable block CG algorithms for solving large sparse symmetric positive definite linear systems on parallel computers. I. General iterative scheme*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1135–1153.
- [16] D. P. O'LEARY, *The block conjugate gradient algorithm and related methods*, Linear Algebra Appl., 29 (1980), pp. 293–322.
- [17] D. P. O'LEARY, *Parallel implementation of the block conjugate gradient algorithm*, Parallel Comput., 5 (1987), pp. 127–139.
- [18] C. RAPPAPORT, *Interpreting and improving the PML absorbing boundary condition using anisotropic lossy mapping of space*, IEEE Trans. Magn., 32 (1996), pp. 968–974.
- [19] E. M. ROSEN AND T. W. ALTSHULER, *Analysis of UXO and clutter signatures from the DARPA background clutter experiment*, in Proceedings of the UXO Forum, 1998.
- [20] M. ROZLOŽNÍK AND Z. STRAKOŠ, *Variants of the residual minimizing Krylov space methods*, in Proceedings of the Eleventh Summer School Software and Algorithms of Numerical Mathematics, I. Marek, ed., University of West Bohemia, Plzen, Czech Republic, 1995, pp. 208–225.
- [21] Y. SAAD, *On the Lanczos method for solving symmetric linear systems with several right hand sides*, Math. Comp., 48 (1987), pp. 651–662.
- [22] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, PWS Publishing, New York, 1996.
- [23] V. SIMONCINI, *A stabilized QMR version of block BICG*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 419–434.
- [24] V. SIMONCINI AND E. GALLOPOULOS, *An iterative method for nonsymmetric systems with multiple right-hand sides*, SIAM J. Sci. Comput., 16 (1995), pp. 917–933.
- [25] C. F. SMITH, *The Performance of Preconditioned Iterative Methods in Computational Electromagnetics*, Ph.D. thesis, University of Illinois at Urbana-Champaign, Illinois, 1987.
- [26] C. F. SMITH, A. F. PETERSON, AND R. MITTRA, *A conjugate gradient algorithm for the treatment of multiple incident electromagnetic fields*, IEEE Trans. Antennas and Propagation, 37 (1989), pp. 1490–1493.

ITERATIVE REGULARIZATION AND MINRES*

MISHA KILMER[†] AND G. W. STEWART[‡]

Abstract. In this paper we present three theorems which give insight into the regularizing properties of MINRES. While our theory does not completely characterize the regularizing behavior of the algorithm, it provides a partial explanation of the observed behavior of the method. Unlike traditional attempts to explain the regularizing properties of Krylov subspace methods, our approach focuses on convergence properties of the residual rather than on convergence analysis of the harmonic Ritz values. The import of our analysis is illustrated by two examples. In particular, our theoretical and numerical results support the following important observation: in some circumstances the dimension of the optimal Krylov subspace can be much smaller than the number of the components of the truncated spectral solution that must be computed to attain comparable accuracy.

Key words. regularization, ill-posed problem, MINRES, iterative regularization

AMS subject classification. 65F10

PII. S0895479898348623

1. Introduction. The discretization of a compact self-adjoint operator equation with errors in the right-hand side results in a linear system whose exact solution bears no relation to the solution of the original error-free system. Specifically, in the spectral domain—that is, in the coordinate system of the eigenvectors of the matrix—the exact linear system can be written in the form

$$\Lambda x = b,$$

where

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n), \quad 1 = \lambda_1 > \dots > \lambda_n > 0.$$

Note that we assume that the system is positive definite and has been scaled so that $\lambda_1 = 1$. To the extent that the quantities in this equation track the corresponding quantities in the original operator, they have the following properties:

- (1) If the operator is smooth, the eigenvalues λ_i of the discrete operator will eventually decay rapidly to zero.
- (2) If the solution of the original problem is square integrable, then the components ξ_i of x decay to zero.
- (3) Since the components β_i of the right-hand side satisfy $\beta_i = \xi_i \lambda_i$, they decay faster than either the components of the solution or the eigenvalues of the operator.

The system with error is

$$\Lambda \tilde{x} = b + e \equiv \tilde{b}.$$

*Received by the editors December 2, 1998; accepted for publication (in revised form) by M. Hanke April 26, 1999; published electronically December 28, 1999.

<http://www.siam.org/journals/simax/21-2/34862.html>

[†]Mathematics Department, Tufts University, Medford, MA 02155 (mkilme01@tufts.edu). This research was supported in part by the Army Research Office under grant DAAG55-97-1-0013.

[‡]Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 (stewart@cs.umd.edu). This research was supported by the National Science Foundation under grant CCR-95-03126.

If we model the error as white noise, then we can append the following condition to the above list.

- (4) The components ϵ_i of e are random variables with mean zero and standard deviation ϵ .

The components of the solution for the perturbed right-hand side are

$$(1.1) \quad \tilde{\xi}_i = \frac{\beta_i + \epsilon_i}{\lambda_i}.$$

Since the β_i decay rapidly, they soon fall below the error level. The subsequent components of the solution are effectively random variables with standard deviation ϵ/λ_i . The division by λ_i magnifies the originally small error so that it dominates the solution.

As long as the β_i are larger than the errors ϵ_i , the $\tilde{\xi}_i$ contain useful information. The process of extracting this information is called *regularization*. There are many regularization schemes. A natural one is to stop computing components of x when the β_i get near the error level. We will call this procedure *truncated spectral regularization*.¹

In this paper we will be concerned with regularization based on the iterative method MINRES [9]. The method can be described briefly as follows. Let

$$K_k = (\tilde{b} \ \Lambda \tilde{b} \ \cdots \ \Lambda^{k-1} \tilde{b})$$

be the k th *Krylov matrix*. Then we seek an approximation to x in the form

$$x_k = K_k a_k,$$

where a_k is determined so that

$$\rho_k^2 \equiv \|\tilde{b} - \Lambda x_k\|^2 = \min.$$

Here $\|\cdot\|$ denotes the usual Euclidean norm.

The residual $r_k = \tilde{b} - \Lambda x_k$ has an alternate expression that is used in the analysis of the algorithm. Given a vector $a_k = (\alpha_1, \dots, \alpha_k)^T$, let a polynomial p be defined by

$$p(t) = 1 - \alpha_1 t - \alpha_2 t^2 - \cdots - \alpha_k t^k.$$

Then the residual $r_k = \tilde{b} - \Lambda K_k a_k$ can be written in the form $r_k = p_k(\Lambda) \tilde{b}$, where p_k is the polynomial associated with a_k . Since MINRES minimizes $\|r_k\|$, we have

$$(1.2) \quad \rho_k \equiv \|r_k\| = \min_{\substack{\deg(p_k)=k \\ p_k(0)=1}} \|p(\Lambda) \tilde{b}\|.$$

Because the i th component of the residual is the value of a polynomial at λ_i , we will often call it the value of the residual at λ_i .

It has been widely observed that up to a certain index k_0 the MINRES solutions x_k are increasingly accurate approximates to the exact solution x , after which their accuracy rapidly deteriorates. Thus if we can determine k_0 (a difficult problem), MINRES can be used as a regularization method. What seems less well known is that

¹In the nonsymmetric case this process is also known as truncated SVD regularization (see [7] and the references therein).

in some circumstances the critical index k_0 can be quite small — much smaller than the number of components of the truncated spectral solution that must be computed to attain the same accuracy.²

In this paper we give three theorems that help explain the regularizing properties of MINRES. The phrase “help explain” is a deliberate warning to the reader not to expect a complete analysis. The theorems do not say that MINRES has to behave in a particular way. Nonetheless, they give considerable insight into the method.

There is a vast body of literature on regularization, much of which is devoted to determining the asymptotic behavior of the error in the optimally regularized solution as the error in the right-hand side approaches zero. (This literature has been admirably surveyed by Engl, Hanke, and Neubauer [4].) There is also a body of literature for well-posed problems in which the asymptotic convergence of Krylov methods like MINRES are analyzed (see, for example, the recent book by Greenbaum [5]). Our approach, which consists of leaving the error in b fixed and determining how the solution behaves as the regularization parameter varies, is essentially nonasymptotic and fits in neither of these categories. On the other hand, our work is closely related to analyses that attempt to find the optimal value of the regularization parameter. (For a survey of this literature see Hansen [8].)

In the next section we give a toy example that will be used to illustrate the subsequent results. We will also describe in a general way what is happening. The next three sections are devoted to establishing our three results. In section 6 we apply our results to another example. The paper concludes with a discussion of the results and suggestions for future work.

A little informal terminology will help in our discussions. We will call the part of the spectrum for which the β_i are little affected by the error the *initial part* of the spectrum. We will call the part where they are fully contaminated the *terminal part* of the spectrum. The part of the spectrum lying between will be called the *intermediate part* or the *transitional part*.

2. An example. In this section we will introduce a simple-minded example from image processing. Let $\underline{x}(t)$ represent an “image” on $[0, 1]$, and consider the Gaussian blurring operator

$$\underline{b}(t) = \int_0^1 \underline{x}(s) e^{-\left(\frac{s-t}{\sigma}\right)^2} ds.$$

We will discretize this operator by choosing an integer n and setting

$$t_i = \frac{i-1}{n}, \quad i = 1, \dots, n+1,$$

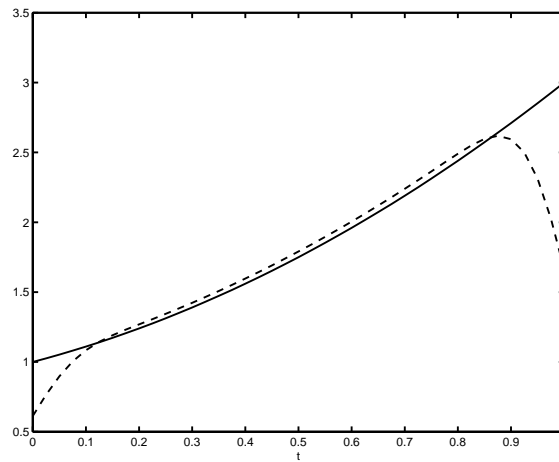
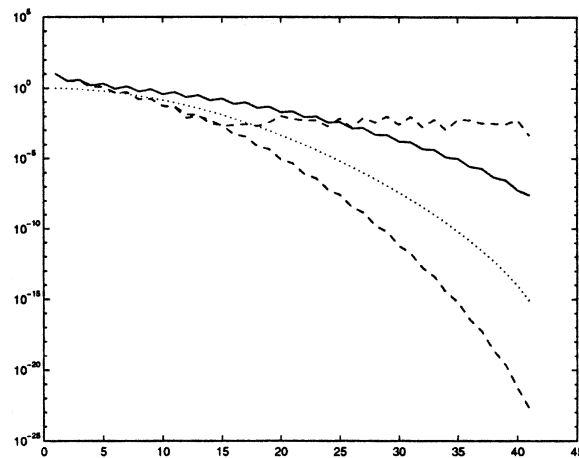
and generating a matrix K whose elements are

$$\kappa_{ij} = e^{-\left(\frac{t_j - t_i}{\sigma}\right)^2}.$$

The resulting matrix is then scaled so that its dominant eigenvalue is one. Although this problem is highly simplified, its regularization properties are typical of more realistic problems.

In the following experiments, we took $n = 40$ and $\epsilon = 0.1$. Our function was $\underline{x}(t) = 1 + t + t^2$, and it was discretized by evaluating it at the points t_i to give the components x_i of x . The vector b was generated in the form Ax . Figure 2.1 is a plot

²Martin Hanke has also observed this phenomenon [6].

FIG. 2.1. $x(t)$ (solid) and $b(t)$ (dashed).FIG. 2.2. ξ_i (solid), λ_i (dotted), and $\beta_i, \tilde{\beta}_i$ (dashed).

of $\underline{x}(t)$ and $\underline{b}(t)$. The divergence at the ends of the intervals is due to the fact that at those points the part of the blurring distribution that lies outside $[0, 1]$ becomes significant.

Figure 2.2 shows the vectors b , \tilde{b} , and x after they have been transformed to the spectral coordinates. Note that for $i > 15$ values $\tilde{\beta}_i$ are significantly contaminated with error. The plot also shows the eigenvalues of A , which decrease slowly at first and then rapidly plunge to about 10^{-15} . This rapid decrease is typical of smooth operators.

Truncated spectral regularization and MINRES regularization behave quite differently for this problem. Figure 2.3 plots the norms of the errors at the k th step of each method. The error for truncated spectral regularization decreases slowly until $k = 15$, after which it increases rapidly. The error in the MINRES iterates decreases rapidly until $k = 5$ and then rises. Truncated spectral regularization produces a bet-

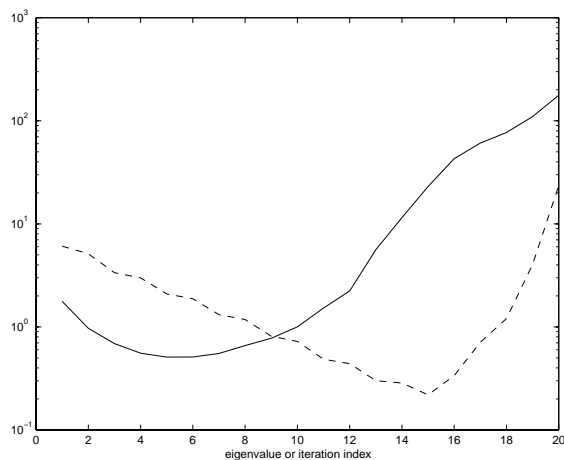


FIG. 2.3. Errors from spectral regularization (dashed) and MINRES regularization (solid).

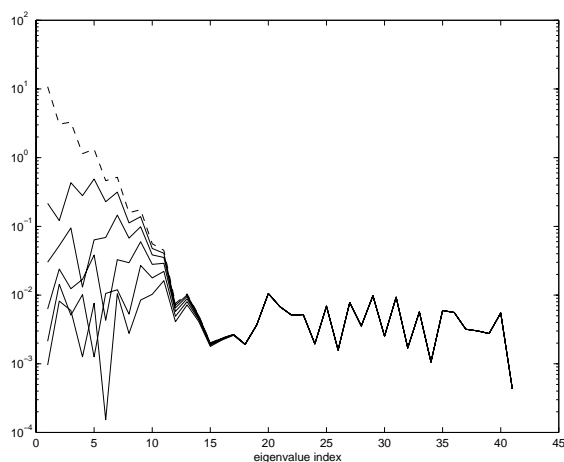


FIG. 2.4. First five MINRES residuals and \tilde{b} (dashed line).

ter solution with about half the error but arrives at it more slowly.³ In fact, not until $k = 11$ does it produce a solution as accurate as the MINRES solution for $k = 5$.

It is informative to examine the behavior of the MINRES residuals. Figure 2.4 displays the magnitudes of the residuals at the eigenvalues λ_i as a function of i for the first five MINRES approximations along with the magnitudes of the components of \tilde{b} . It is seen that they decrease in the initial part of the spectrum. At other points in the spectrum, they change very little. The plot in Figure 2.5 shows the same thing in terms of the residual polynomials that satisfy (1.2).⁴ The polynomials are small in

³The relative quality of the solutions depends on the throw of the dice in constructing the random vector e . For ten other simulations the ratio of the minimum error for truncated spectral regularization to that for MINRES was the following: 0.80, 0.57, 0.76, 0.41, 1.14, 0.65, 0.67, 0.67, 0.82, 0.68. These numbers show the solutions to be usually comparable, with the MINRES solution actually better in a single case.

⁴The plots are not strictly the polynomials. The values between i and $i + 1$ were obtained by evaluating the polynomial at ten equally spaced points between λ_i and λ_{i+1} .

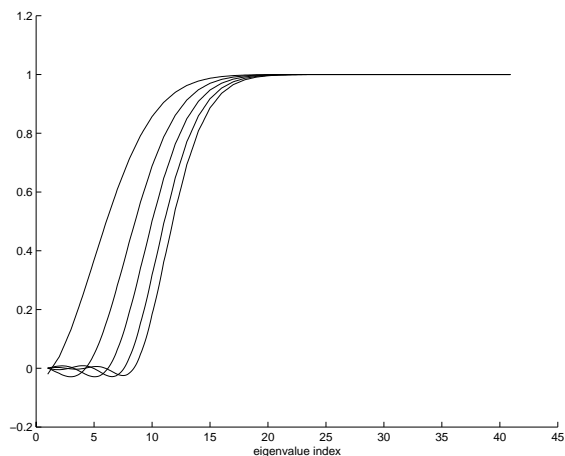


FIG. 2.5. *First five MINRES residual polynomials.*

the initial part of the spectrum and then rise swiftly to a value of one.

Stepping back from this example, we see that three factors are operating to give the fast regularization.

- (1) From (1.1) it follows that the residual cannot be reduced indefinitely without harming the solution.
- (2) The plot in Figure 2.4 shows that the residual decreases rapidly to a point where further reduction is harmful.
- (3) Notice in Figure 2.4 that the values of the first five residuals corresponding to terminal eigenvalues λ_i coincide with $\tilde{\beta}_i$. Equation (1.1) implies that if we reduce the residual at λ_i from $\tilde{\beta}_i$ to $\tilde{\beta}_i(1 - \alpha)$, where $0 < \alpha < 1$, then the corresponding component of the approximate solution is $\alpha\tilde{\beta}_i/\lambda_i$. Now in the terminal part of the spectrum $\tilde{\beta}_i \cong \epsilon_i$, and hence the component of the approximate solution is essentially $\alpha\epsilon_i/\lambda_i$. Because the terminal eigenvalues are extremely small compared to the error, even a very small reduction of the residual at a terminal eigenvalue must result in a large component in the approximate solution. Thus if MINRES is to produce an acceptably regularized solution, it must not reduce the residuals at the terminal part of the spectrum while it is still reducing the residuals at the initial part. As it turns out (see Figure 2.5), the first 5 residual polynomials for our running example are very near one on the terminal spectrum, so there is virtually no reduction of the residual there.

In the next three sections we will establish results that give quantitative substance to these observations.

3. The residual and the error. The first of our results concerns the relation between the size of the error in a purported solution and the size of the corresponding residual. We have already noted that if the residual is required to be too small, the error must be large. However, we can be more precise.

Specifically, let y be a purported solution with residual norm δ . There are many such solutions, and among them one must give minimum error; i.e., there must be a

vector y that solves the following problem. Given $\delta > 0$,

$$(3.1) \quad \begin{aligned} & \text{minimize } \|x - y\|^2 \\ & \text{subject to } \|\tilde{\beta} - \Lambda y\|^2 = \delta^2. \end{aligned}$$

The following theorem shows when the solution to this problem has a large error. Below we have assumed $\lambda_1 = 1$.

THEOREM 3.1. *If $\tau > -1$ is the solution of the equation*

$$(3.2) \quad \sum_i \frac{\epsilon_i^2}{(1 + \tau \lambda_i^2)^2} = \delta^2,$$

then

$$(3.3) \quad \eta_i = \xi_i + \frac{\tau \lambda_i \epsilon_i}{1 + \tau \lambda_i^2}$$

is a solution of (3.1). In this case

$$(3.4) \quad \|x - y\|^2 = \sum_i \left(\frac{\tau \lambda_i}{1 + \tau \lambda_i^2} \right)^2 \epsilon_i^2.$$

Proof. Consider the Lagrangian

$$\sum_i (\xi_i - \eta_i)^2 + \tau \left[\sum_i (\tilde{\beta}_i - \lambda_i \eta_i)^2 - \delta^2 \right],$$

where τ is the Lagrange multiplier. Differentiating with respect to η_i and setting the results to zero, we get

$$\xi_i - \eta_i + \tau \lambda_i (\tilde{\beta}_i - \lambda_i \eta_i) = 0.$$

Equation (3.3) follows on solving this equation for η_i and using the fact that $\tilde{\beta}_i = \lambda_i \xi_i + \epsilon_i$. Equation (3.2) defines the value of τ for which the solution satisfies the constraint. Finally, (3.4) follows by direct computation. \square

The best way to understand this theorem is to consider the solution y_τ generated as τ varies from -1 to ∞ . From (3.2) we see that the residual norm δ decreases monotonically from ∞ to 0. The error norm $\|x - y_\tau\|$ decreases from ∞ to 0 at $\tau = 0$, where $\delta = \|e\|$. Thereafter, small reductions in the residual cause large increases in the error. To see this, consider the term

$$\frac{\epsilon_i^2}{(1 + \tau \lambda_i^2)^2}$$

in (3.2). For this term to be reduced by a factor of four from its value ϵ_i^2 at $\tau = 0$, a very modest reduction in the total, we must have $\tau \lambda_i^2 = 1$. The corresponding term in the square of the error norm (3.4) is

$$\left(\frac{\tau \lambda_i \epsilon_i}{1 + \tau \lambda_i^2} \right)^2 = \frac{\epsilon_i^2}{4 \lambda_i^2},$$

which for small λ_i is large. It is worth noting that these observations depend only on Λ and e and not on x or b .

The above comments suggest that an attempt to reduce the residual norm below $\|e\|$ will increase the error in the solution. This is certainly true of our example, for which $\|e\| = 0.033$. The error turns around at $k = 5$. The residual norms are $\rho_4 = 0.055$, $\rho_5 = 0.039$, and $\rho_6 = 0.031$. Thereafter the error increases sharply, as it must (see Figure 2.3).

It is worth noting that Theorem 3.1 supports the discrepancy principle of Morozov, which says that a regularization parameter should be chosen to make the norm of the residual approximately the size of the norm of the error (see [4, section 4.3]). However, the rapid increase of the error when the residual becomes too small suggests that methods based on the discrepancy principle will be extremely sensitive to variations in the estimate of the error norm (see the experiments in section 7.7.1 of [8]).

4. Reduction of the residual. We now turn to the reduction of the residual. The usual analysis of MINRES is based on (1.2). Specifically, given any polynomial p_k of degree k satisfying $p_k(0) = 1$, we have

$$(4.1) \quad \|r_k\|^2 \leq \sum_{i=1}^n p_k(\lambda_i)^2 \tilde{\beta}_i^2.$$

One convergence result amounts to choosing a sequence of polynomials p_k such that

$$\lambda \in [\lambda_n, \lambda_1] \implies p_k(\lambda) \rightarrow 0.$$

The rate at which the p_k approach zero on $[\lambda_n, \lambda_1]$ is an upper bound on the convergence rate of MINRES.

The key to analyzing the regularizing properties of MINRES is to note that we are not interested in the convergence of the method on the terminal part of the spectrum. In fact (1.1) shows that MINRES *must not* converge there until after a reasonable regularized solution has been formed. We can therefore restrict our attention to an interval, say, $[\lambda_m, \lambda_1]$ in which the $\tilde{\beta}_i$ are relatively error free. Since the interval is farther removed from the origin, the convergence will be faster.⁵

The polynomial $p_{k,m}$ for this analysis is constructed in the usual way from the Chebychev polynomials $c_k(\tau)$. Specifically, let

$$(4.2) \quad s_m(t) = -\frac{2t - \lambda_1 - \lambda_m}{\lambda_1 - \lambda_m}$$

and set

$$(4.3) \quad p_{k,m}(t) = \frac{c_k[s_m(t)]}{c_k[s_m(0)]}.$$

By bounding the values of these polynomials we obtain the following theorem.

THEOREM 4.1. *Let $m > 1$ and set*

$$\kappa = \frac{\lambda_1}{\lambda_m}$$

⁵For the standard analysis see [5, Chapter 3]. In fact, one reviewer has pointed out that Björck and Eldén [2] used polynomials of this kind to construct regularizing iterative methods. Here, on the other hand, we use them to analyze the regularizing properties of a particular iterative method.

and

$$\tilde{b}_m = (\tilde{\beta}_1, \dots, \tilde{\beta}_m)^T.$$

Then

$$(4.4) \quad \begin{aligned} \|r_k\|^2 &\leq \|p_{k,m}(\Lambda)\tilde{b}\|^2 \\ &\leq 4 \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \right)^{2k} \|\tilde{b}_m\|^2 \\ &\quad + \sum_{i=m+1}^n \left(1 - 2 \frac{\lambda_i}{(\sqrt{\lambda_1} + \sqrt{\lambda_i})^2} \right)^{2k} \tilde{\beta}_i^2. \end{aligned}$$

The proof of this theorem is highly technical and will be found in the appendix. The first term in (4.4) is of the same form as the bound from the usual analysis of MINRES mentioned above. However, for that bound κ is λ_1/λ_n , which is larger than λ_1/λ_m and implies slower convergence. The first factor in the remaining terms is essentially one for small eigenvalues. However, for eigenvalues near λ_m it has a reducing effect on the corresponding part of the residual.

The bound is too crude to use in applications, but it provides insight into the way MINRES converges. In our running example, when m is small, κ is near one, and the polynomials $p_{k,m}$ rapidly reduce the residual on $\lambda_1, \dots, \lambda_m$. But this is precisely the place where there is a large residual to reduce. Since the reduction by $p_{k,m}$ bounds the MINRES reduction, we conclude that initially MINRES will rapidly reduce the residual.

But the bound implies more. As k increases, for fixed m , $p_{k,m}$ will reach a point of diminishing returns when there is not enough residual to reduce. This can be seen by looking at the upper bound in (4.4). Consider m fixed at 4. Since κ is effectively 1, the first term becomes insignificant for $k \geq 2$, and we have a tight bound on the residual for $k = 2$. Yet as k increases and the residual continues to be reduced, the bound stagnates because the second term decreases slowly. However, if we increase m a little, the first term remains small while the second term becomes smaller (since there are fewer indices over which to sum). Thus with increasing k one should increase m to get the smallest possible upper bound. The following table shows that this is precisely what happens in our example.

k	1	2	3	4	5
m	2	4	5	6	7
$\ p_{k,m}(\Lambda)\tilde{b}\ $	8.8e-01	2.8e-01	1.3e-01	7.1e-02	4.7e-02
$\max_{[\lambda_m, \lambda_1]} p_{k,m}$	8.0e-02	2.9e-02	1.0e-02	6.1e-03	5.2e-03
ρ_k	8.7e-01	2.5e-01	1.0e-01	5.5e-02	3.9e-02

For each value of k , the value of m for which $p_{k,m}$ minimizes the residual $\|p_{k,m}(\Lambda)\tilde{b}\|$ was determined. As the table shows, the value of m increases with k . This increase is accompanied by a modest decrease in the size of $p_{k,m}$ on $[\lambda_m, \lambda_1]$. Note that by holding m fixed one can obtain a dramatic decrease in the size of $p_{k,m}$ over $[\lambda_m, \lambda_1]$, but this strategy is not as effective at reducing the bound on the residual norm as the strategy of increasing m and putting up with a modest decrease in $p_{k,m}$. For example, if m is fixed at two, then $\max_{[\lambda_2, \lambda_1]} p_{5,2} = 1.5e-09$, but the residual norm is only $8.9e-02$.

The last line in the table shows the MINRES residual norms. They are well approximated by the values in the third line, which suggests that MINRES is following a similar strategy. Figure 2.5 supports our contention that MINRES reduces the residual by increasing the interval in which the polynomial is small.

5. Nonreduction of the residual. Let $\alpha \in [0, 1]$, and suppose we wish to reduce the residual at λ_i from $\tilde{\beta}_i$ to $\tilde{\beta}_i(1 - \alpha)$. We have already noted that the corresponding component of \tilde{x} is then

$$\tilde{\xi}_i = \frac{\alpha \tilde{\beta}_i}{\lambda_i}.$$

If λ_i is small, then $\tilde{\xi}_i$ is large and the solution will be unsatisfactory. Thus while MINRES is producing a regularized solution, it must not reduce the residuals on the terminal part of the spectrum. From Figure 2.5 we see that this is precisely what is happening in our example. The residual polynomials are effectively one at small λ_i .

The reason MINRES behaves in this manner is that if any polynomial is small at a small eigenvalue, it must be very large at the initial eigenvalues, say $\lambda_1, \dots, \lambda_m$. Specifically, we have the following theorem.

THEOREM 5.1. *Let $p_{k,m}$ be defined by (4.3). If p is any polynomial of degree k satisfying $p(0) = 1$, $p(\lambda_i) = 1 - \alpha$ for some $i > m$, and*

$$(5.1) \quad \frac{p_{k-1,m}(\lambda_i)\lambda_i}{p_{k-1,m}(\lambda_i) - 1 + \alpha} < \lambda_m,$$

then there is a point $t_{\text{big}} \in [\lambda_m, \lambda_1]$ such that

$$(5.2) \quad |p(t_{\text{big}})| \geq \left| 1 - \frac{p_{k-1,m}(\lambda_i) - 1 + \alpha}{p_{k-1,m}(\lambda_i)} \frac{\lambda_m}{\lambda_i} \right| |p_{k-1,m}(\lambda_m)|.$$

Proof. The proof is a variant of the standard proof of the theorem of de laVallée Poussin [1, p. 191]. Consider the polynomial

$$q(t) = \left[1 - \frac{p_{k-1,m}(\lambda_i) - 1 + \alpha}{p_{k-1,m}(\lambda_i)} \frac{t}{\lambda_i} \right] p_{k-1,m}(t).$$

This polynomial satisfies the conditions $p(0) = 1$ and $p(\lambda_i) = 1 - \alpha$. If, in addition, (5.1) is satisfied, the root of the linear term in brackets is less than λ_m , and the polynomial $q(t)$ alternates in sign at k points

$$\lambda_m = t_1, t_2, \dots, t_k = 1$$

on the interval $[\lambda_m, \lambda_1]$.

Now let p also satisfy the conditions of the theorem. We claim that $|p(t_i)| \geq |q(t_i)|$ for at least one t_i . For if not, the polynomial $p - q$ alternates in sign at the t_i and hence has $k - 1$ zeros in the interval $[\lambda_m, \lambda_1]$. But by the conditions of the theorem $p - q$ is zero at zero and λ_i . Hence $p - q$ is a polynomial of degree not greater than k with $k + 1$ zeros and must be identically zero — a contradiction.

Because the values of $|p_{k-1,m}(t_i)|$ are equal, the absolute value of $q(t_i)$ is smallest at λ_m , and evaluating $|q(\lambda_m)|$ gives the lower bound (5.2). \square

Note that when i is large the condition (5.1) is not very restrictive. For in that case, the right-hand side is near λ_i/α , which cannot approach λ_m unless α is very small.

The import of this theorem is that the cost of reducing the residual at the small eigenvalues is to make the polynomial large on an interval containing the initial eigenvalues. This is because the factor λ_m/λ_i in (5.2) becomes enormous for i large. For

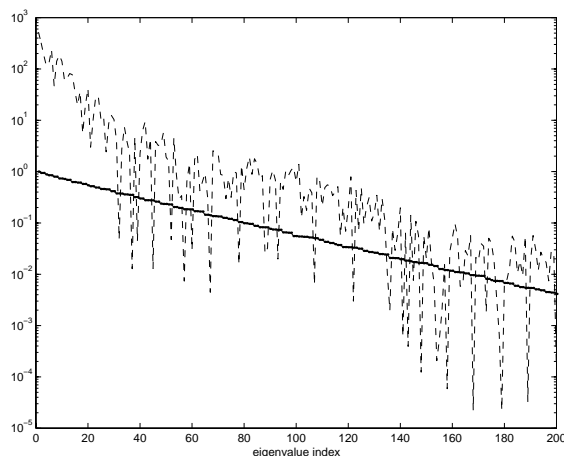


FIG. 6.1. *Eigenvalues (solid line) and components of b (dashed line).*

example, Theorem 5.1 implies that if $m = 8$ and $k = 5$ in our running example, then to get a reduction of .9 ($\alpha = .1$) in the residual at $\lambda_{25} = 6.8\text{e-}6$, the absolute value of the residual polynomial will have to be greater than 71 at some point in $[\lambda_8, \lambda_1]$. Such a large polynomial is more likely to cause a residual magnification than a residual reduction in $[\lambda_8, \lambda_1]$. This likelihood is increased by the fact that for our example the roots of p_k are spread out over $[\lambda_m, \lambda_1]$, where $m > k$, and the zeros of p_k cannot be near all m of these eigenvalues.

On the other hand, in our example, suppose we wish to reduce the residual at λ_{15} from β_{15} to $\beta_{15}(1 - \alpha)$ for $\alpha = 0.11$. The lower bound (5.2) for the residual polynomial is 0.0086. The maximum value of the residual polynomial in $[\lambda_8, \lambda_1]$ is, in fact, 0.037, which is insufficiently greater than 0.0086 to render the reduction at λ_{15} harmless. Once again, the dispersal of the roots of $p_{k,m}$ has a beneficial effect — this time allowing a modest decrease in the residual at the intermediate eigenvalues.

In the last section we observed that MINRES eschews a quick reduction of the residual polynomials over a limited range in favor of expanding the range in which they are moderately small. The theory of this section shows that the deferred reduction has another benefit: it gives the algorithm an opportunity to reduce the residual by a modest amount on the intermediate eigenvalues.

6. Another example. To confirm our analysis of MINRES, we consider a two dimensional blurring operator. Specifically, the original matrix A is the Kronecker product $T \otimes T$, where T is a symmetric Toeplitz matrix whose first row is given by

$$t_{1k} = e^{-0.1(1-k)^2}, \quad k = 1, \dots, 32.$$

Figure 6.1 exhibits the first 200 eigenvalues of the matrix and the absolute values components of b in the spectral coordinate system. The error e consists of white noise scaled so that $\|e\|/\|b\| = 0.01$.

Figure 6.2 plots the errors from spectral regularization and MINRES. The contrast is dramatic. The MINRES iteration reduces the error to a minimum of 195 in four iterations. Spectral regularization requires the first 100 components of the solution be computed to reduce the error to a minimum of 192. (However, the plot shows that acceptable accuracy is attained for, say, 30 components.)

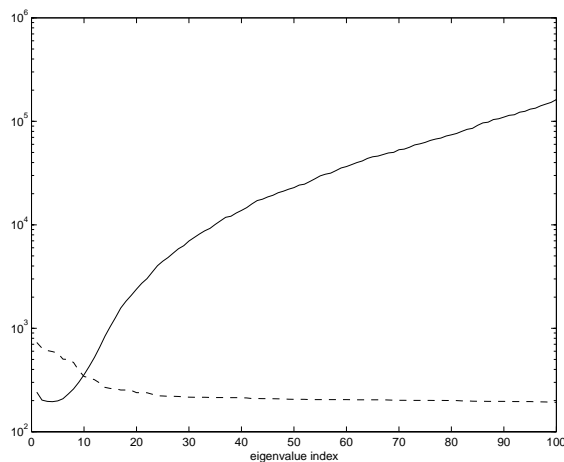


FIG. 6.2. Errors from spectral regularization (dashed line) and MINRES regularization (solid line).

The analysis of section 3 suggests that the solution will deteriorate when the residual norm becomes less than the error norm. In fact, the fourth, fifth, and sixth residual norms are 8.66, 7.86, and 7.52. This should be compared with the error norm 7.78.

From the comments in section 4 we would expect that residual polynomials would reduce the error by spreading their roots over a number of eigenvalues rather than by approximating specific eigenvalues. The roots for the first five residual polynomials are given below.

(1)	9.3e-01				
(2)	9.8e-01	7.3e-01			
(3)	9.9e-01	8.1e-01	5.7e-01		
(4)	9.9e-01	8.7e-01	6.9e-01	3.6e-01	
(5)	1.0e+00	9.0e-01	7.4e-01	5.4e-01	2.2e-01

The largest root is converging to one, which in our normalization is an eigenvalue of A . But otherwise, the roots are dispersed. For example, the smallest root of p_4 is 0.36, which corresponds to λ_{34} .

7. Conclusions. In the introduction to this paper we emphasized that this work would not provide a complete analysis of MINRES as a regularization technique. We can analyze MINRES applied to well-posed problems because asymptotic bounds become applicable long before a solution of the desired accuracy is attained. On the other hand, for ill-posed problems in our examples, MINRES does not have a chance to settle down before the optimal solution is obtained. In effect, we are analyzing the early transient behavior of the algorithm — always a difficult problem.

In spite of not concluding with a general theorem, our analysis sheds considerable light on the behavior of MINRES as a regularizer. Theorem 3.1 is a very general result about ill-posed problems. It suggests that attempting to reduce the residual norm below the error norm will result in a worsened solution. As we have already noted, the theorem can be taken as a justification of the Morozov discrepancy principle.

We have mentioned that the bound (4.4) on the residual norm is too crude to be used in practical applications. Nonetheless, it is good enough to suggest the strategy that MINRES uses to obtain rapid reduction of the residual norm. The free parameter m in the bound is essentially a measure of the spread of the zeros of the residual polynomial. As we have seen, the bound is optimized by keeping m greater than the iteration number k . This suggests that MINRES will attempt to minimize the residual by spreading the roots of its residual polynomials over a larger number of eigenvalues. Our numerical examples support this conjecture.

Finally, Theorem 5.1 shows us why MINRES does not tend to reduce the residual at a small eigenvalue while homing in on the regularized solution. Essentially, a small value of the residual polynomial at a small eigenvalue causes the polynomial to be large on the larger eigenvalues—just where it is most desirable to obtain a decrease in the residuals.

For definiteness, we have restricted ourselves to MINRES applied to a positive definite system. Obviously, some of our results apply to other systems and methods—e.g., least squares regularization of overdetermined systems. Such generalizations, however, will be subject for future work.

8. Appendix. *Proof of Theorem 4.1.* Let $c_k(\tau)$ denote the Chebychev polynomial of degree k and define $s_m(t)$ and $p_{k,m}(t)$ as in (4.2) and (4.3), respectively. Now $p_{k,m}$ is a polynomial of degree k satisfying $p_{k,m}(0) = 1$, and therefore from (4.1), we have

$$(8.1) \quad \|r^{(k)}\|_2^2 \leq \sum_{i=1}^n p_{k,m}^2(\lambda_i) \tilde{\beta}_i = \|p_{k,m}(\Lambda) \tilde{b}\|_2^2.$$

When $i \leq m$, $c_k^2(s_m(\lambda_i)) \leq 1$. Moreover, one can show that the denominator of $p_{k,m}$, which is independent of i , is bounded above [10]:

$$(8.2) \quad \left| \frac{1}{c_k\left(\frac{\lambda_1 + \lambda_m}{\lambda_1 - \lambda_m}\right)} \right| \leq 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k.$$

Thus, it remains to determine an upper bound for the values $c_k(s_m(\lambda_i))$, $i > m$.

For notational convenience, we fix i , let $d = \frac{\lambda_i}{\lambda_1}$, and note that since $\kappa = \frac{\lambda_1}{\lambda_m}$, $\frac{\lambda_i}{\lambda_m} = d\kappa$. Using this notation, we write $s_{m,i} \equiv s_m(\lambda_i) = \left(\frac{\kappa+1-2d\kappa}{\kappa-1}\right)$.

Now $c_k(s) = \cosh(k \cosh^{-1} s)$. We use the formulas

$$(8.3) \quad \cosh x = \frac{e^x + e^{-x}}{2}$$

and

$$(8.4) \quad \ln z = \cosh^{-1} \left(\frac{z + z^{-1}}{2} \right) \quad \text{when } x = \ln z$$

to determine our upper bound as outlined below. We note that this type of proof technique was used at least as early as the 1960s (see [3]).

- (1) Find a z such that $\left(\frac{z+z^{-1}}{2}\right) = s_{m,i}, i > m$.
- (2) Use z and (8.3) to determine $c_k(s_{m,i}), i > m$.
- (3) Find an upper bound for $c_k^2(s_{m,i}), i > m$.
- (4) Use this bound and (8.2) to deduce an upper bound for $p_{k,m}^2(s_{m,i}) i > m$.
- (5) Use the above information to obtain the upper bound on the right-hand side of (8.1).

Step 1. We first set

$$\frac{z + z^{-1}}{2} = s_{m,i}$$

to obtain

$$z^2 - 2s_{m,i}z + 1 = 0.$$

Since this is a quadratic equation, z is given by

$$\begin{aligned} z &= \frac{2s_{m,i} \pm \sqrt{4s_{m,i}^2 - 4}}{2} \\ &= \left(\frac{\kappa + 1 - 2d\kappa}{\kappa - 1} \right) \pm \sqrt{\left(\frac{\kappa + 1 - 2d\kappa}{\kappa - 1} \right)^2 - 1} \\ &= \frac{1}{\kappa - 1} \left(\kappa + 1 - 2d\kappa \pm 2\sqrt{(\kappa - \kappa d)(1 - \kappa d)} \right) \\ &= \frac{(\sqrt{\kappa - \kappa d} \pm \sqrt{1 - \kappa d})^2}{\kappa - 1}. \end{aligned}$$

(8.5)

For convenience, we take z to be

$$z = \frac{(\sqrt{\kappa - \kappa d} + \sqrt{1 - \kappa d})^2}{\kappa - 1}. \quad (8.6)$$

Step 2. Now $c_k(s_{m,i}) = \cosh(k \ln z) = \cosh(\ln z^k)$. Using (8.3), we obtain

$$\begin{aligned} c_k(s_{m,i}) &= \frac{z^k + z^{-k}}{2} \\ &= \frac{z^{2k} + 1}{2z^k}. \end{aligned}$$

(8.7)

Step 3. Therefore,

$$c_k^2(s_{m,i}) = \frac{1}{4} \left(z^{2k} + \frac{1}{z^{2k}} + 2 \right).$$

(8.8)

To get an upper bound on $c_k^2(s_{m,i})$, let us first try to get an upper bound on $\frac{1}{z^{2k}}$ by getting a lower bound on z . Observe that the numerator of z in (8.6) is bounded below by $\kappa - d\kappa$, so our lower bound for z is $\frac{\kappa - d\kappa}{\kappa - 1}$. Since $d\kappa \leq 1$, we therefore obtain

$$\frac{1}{z} \leq \frac{\kappa - 1}{\kappa - d\kappa} \leq 1,$$

and so $\frac{1}{z^{2k}} \leq 1$.

Next, we have the following expression for z^{2k} :

$$z^{2k} = \left(\frac{\kappa + 1 + 2((\kappa - d\kappa)(1 - d\kappa))^{1/2} - 2d\kappa}{\kappa - 1} \right)^{2k}.$$

Therefore, we have the following bound on $c_k^2(s_{m,i})$:

$$(8.9) \quad c_k^2(s_{m,i}) \leq \frac{1}{4} \left(3 + \left(\frac{\kappa + 1 + 2((\kappa - d\kappa)(1 - d\kappa))^{1/2} - 2d\kappa}{\kappa - 1} \right)^{2k} \right).$$

Step 4. Now we may use (8.2) and (8.9) to determine an upper bound on $p_{k,m}^2(s_{m,i})$ for $i > m$:

$$p_{k,m}^2(s_{m,i}) \leq \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^{2k} \left(3 + \left(\frac{\kappa + 1 + 2((\kappa - d\kappa)(1 - d\kappa))^{1/2} - 2d\kappa}{\kappa - 1} \right)^{2k} \right).$$

Using $\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} = \frac{\kappa-1}{\kappa+2\sqrt{\kappa}+1}$, we cancel like terms on the right-hand side above and factor out $\sqrt{\kappa}$ from the radical to obtain

$$p_{k,m}^2(s_{m,i}) \leq 3 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^{2k} + \left(\frac{\kappa + 1 + 2\sqrt{\kappa} \left(\left(1 - \frac{\lambda_i}{\lambda_1}\right) \left(1 - \frac{\lambda_i}{\lambda_m}\right) \right)^{1/2} - 2\frac{\lambda_i}{\lambda_m}}{\kappa + 1 + 2\sqrt{\kappa}} \right)^{2k}.$$

The second term of the right-hand side above can be bounded above by

$$\left(1 - \frac{2\frac{\lambda_i}{\lambda_m}}{\kappa + 1 + 2\sqrt{\kappa}} \right)^{2k} = \left(1 - 2\frac{\lambda_i}{\lambda_1 + \lambda_m + 2\sqrt{\lambda_1\lambda_m}} \right)^{2k}.$$

Thus,

$$(8.10) \quad p_{k,m}^2(s_{m,i}) \leq 3 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^{2k} + \left(1 - 2\frac{\lambda_i}{\lambda_1 + \lambda_m + 2\sqrt{\lambda_1\lambda_m}} \right)^{2k}.$$

Step 5. From (8.2) when $i \leq m$,

$$p_{k,m}^2(s_{m,i}) \leq \left(\frac{l}{c_k(s_{m,i})} \right)^2 \leq 4 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^{2k}.$$

Breaking up the sum over the first m and last $n - m$ terms from (8.1) and (8.10), we observe

$$\|r_k\|_2^2 \leq 4 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^{2k} \|\tilde{b}_m\|_2^2 + \sum_{i=m+1}^n \left(1 - 2\frac{\lambda_i}{(\sqrt{\lambda_1} + \sqrt{\lambda_m})^2} \right)^{2k} \tilde{\beta}_i^2.$$

9. Acknowledgments. We wish to thank Per Christian Hansen and Dianne O’Leary for their useful comments on the first draft of this paper.

REFERENCES

- [1] K. E. ATKINSON, *An Introduction to Numerical Analysis*, John Wiley, New York, 1978.
- [2] A. BJÖRCK AND L. ELDÉN, *Methods in Numerical Algebra for Ill-Posed Problems*, Technical Report LiTH-MAT-R-1979-33, Department of Mathematics, Linköping University, Linköping, Sweden, 1979.
- [3] J. W. DANIEL, *The conjugate gradient method for linear and nonlinear operator equations*, SIAM J. Numer. Anal., 4 (1967), pp. 10–26.
- [4] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.
- [5] A. GREENBAUM, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.
- [6] M. HANKE, private communication.
- [7] P. C. HANSEN, *The truncated SVD as a method for regularization*, BIT, 27 (1987), pp. 534–553.
- [8] P. C. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
- [9] C. C. PAIGE AND M. A. SAUNDERS, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.
- [10] J. STOER AND R. BULIRSCH, *Introduction to Numerical Analysis*, 2nd ed., Springer-Verlag, New York, 1993.

CHOOSING REGULARIZATION PARAMETERS IN ITERATIVE METHODS FOR ILL-POSED PROBLEMS*

MISHA E. KILMER[†] AND DIANNE P. O'LEARY[‡]

Abstract. Numerical solution of ill-posed problems is often accomplished by discretization (projection onto a finite dimensional subspace) followed by regularization. If the discrete problem has high dimension, though, typically we compute an approximate solution by projecting the discrete problem onto an even smaller dimensional space, via iterative methods based on Krylov subspaces. In this work we present a common framework for efficient algorithms that regularize after this second projection rather than before it. We show that determining regularization parameters based on the final projected problem rather than on the original discretization has firmer justification and often involves less computational expense. We prove some results on the approximate equivalence of this approach to other forms of regularization, and we present numerical examples.

Key words. ill-posed problems, regularization, discrepancy principle, iterative methods, L-curve, Tikhonov, truncated singular value decomposition, projection, Krylov subspace

AMS subject classifications. 65F10, 65F22

PII. S0895479899345960

1. Introduction. Linear, discrete ill-posed problems of the form

$$(1.1) \quad \min_x \|Ax - b\|_2$$

arise, for example, from the discretization of first-kind Fredholm integral equations and occur in a variety of applications. We shall assume that the full-rank matrix A is $m \times n$ with $m \geq n$. In discrete ill-posed problems, A is ill-conditioned and there is often no gap in the singular value spectrum. Typically, the right-hand side b contains noise due to measurement and/or approximation error. This noise, in combination with the ill-conditioning of A , means that the exact solution of (1.1) has little relationship to the noise-free solution and is worthless.

Instead, we use a *regularization* method to determine a solution that approximates the noise-free solution. We replace the original operator by a better conditioned but related one in order to diminish the effects of noise in the data. Sometimes this regularized problem is too large to solve exactly. In that case, we typically project the problem onto an even smaller dimensional space, perhaps via iterative methods based on Krylov subspaces. Sometimes this projection provides enough regularization to produce a good approximate solution, but often (see, for example, [28, 15]) additional regularization is needed.

A fundamental decision to be made in such cases is whether to regularize before or after the projection. *One subtle issue is that the regularization parameter that is optimal for the discretized problem may not be optimal for the lower-dimensional problem actually solved by the iteration*, and this leads to the research discussed in this paper.

*Received by the editors December 3, 1999; accepted for publication (in revised form) by P. C. Hansen November 13, 2000; published electronically April 6, 2001. This work was supported by the National Science Foundation under grants CCR 95-03126 and CCR 97-32022 and by the Army Research Office, MURI grant DAAG55-97-1-0013.

<http://www.siam.org/journals/simax/22-4/34596.html>

[†]Department of Mathematics, Tufts University, Medford, MA 02155 (mkilme01@tufts.edu).

[‡]Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 (oleary@cs.umd.edu).

At first glance, there can appear to be a lot of work associated with the selection of a good regularization parameter, and many algorithms proposed in the literature are needlessly complicated, repeating a Krylov iteration multiple times. By regularizing after projection by the iterative method, so that we are regularizing the lower dimensional problem that is actually being solved, this difficulty vanishes.

The purpose of this paper is to present a common framework for parameter selection techniques applied to the problem resulting from iterative methods such as Krylov subspace techniques. We show that by determining regularization parameters based on the final projected problem rather than on the original discretization, we can better approximate the optimal parameter and reduce the cost of solution.

Our paper is organized as follows. In section 2 we survey some methods for choosing the corresponding regularization parameters. In section 3, we show how any standard parameter selection technique for the original problem can be applied instead to a projected problem obtained from an iterative method, greatly reducing the cost without much degradation in the solution. We give experimental results in section 4 and conclusions in section 5.

In the following we shall assume that $b = b_{true} + e$, where b_{true} denotes the unperturbed data vector and e denotes zero-mean white noise. We will also assume that b_{true} satisfies the discrete Picard condition; that is, the spectral coefficients of b_{true} decay faster, on average, than the singular values.

Let $\hat{U}\Sigma\hat{V}^*$ denote the singular value decomposition (SVD) of A , where the columns of \hat{U} and \hat{V} are the singular vectors, and the singular values are ordered as $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$. Then the solution (1.1) is given by

$$(1.2) \quad x = \sum_{i=1}^n \frac{\hat{u}_i^* b}{\sigma_i} \hat{v}_i = \sum_{i=1}^n \left(\frac{\hat{u}_i^* b_{true}}{\sigma_i} + \frac{\hat{u}_i^* e}{\sigma_i} \right) \hat{v}_i.$$

As a consequence of the white noise assumption, $|\hat{u}_i^* e|$ is roughly constant for all i , while the discrete Picard condition guarantees that $|\hat{u}_i^* b_{true}|$ decreases with i faster than σ_i does. The matrix A is ill-conditioned, so small singular values magnify the corresponding coefficients $\hat{u}_i^* e$ in the second sum, and it is this large contribution of noise that renders the exact solution x defined in (1.2) worthless. The following four classes of regularization methods try in different ways to lessen the contribution of noise. For further information on these methods, see, for example, [19, 15].

In **Tikhonov regularization**, (1.1) is replaced by

$$(1.3) \quad \min_x \|Ax - b\|_2^2 + \lambda^2 \|Lx\|_2^2,$$

where λ is a positive scalar regularization parameter, and we choose L to be the identity matrix I . Solving (1.3) is equivalent to solving

$$(1.4) \quad (A^*A + \lambda^2 I)x_\lambda = A^*b.$$

In analogy with (1.2) we have

$$(1.5) \quad x_\lambda = \sum_{i=1}^n \left(\frac{\sigma_i \hat{u}_i^* b_{true}}{\sigma_i^2 + \lambda^2} + \frac{\sigma_i \hat{u}_i^* e}{\sigma_i^2 + \lambda^2} \right) \hat{v}_i.$$

In **truncated SVD** we compute the regularized solution by truncating the expansion in (1.2) as

$$(1.6) \quad x_\ell = \sum_{i=1}^{\ell} \frac{\hat{u}_i^* b}{\sigma_i} \hat{v}_i.$$

Here the regularization parameter is ℓ , the number of terms retained in the sum. Rust [33] introduced a related truncation strategy, including in the sum (1.2) only those terms corresponding to a spectral coefficient $\hat{u}_i^* b$ whose magnitude is greater than or equal to some tolerance ρ , which can be regarded as the regularization parameter.

Solving (1.4) or (1.6) can be impractical if n is large, but fortunately, regularization can be achieved through **projection** onto a k -dimensional subspace; see, for example, [9]. The truncated SVD (TSVD) is one example, but projection is often achieved through the use of iterative methods such as conjugate gradients, GMRES, QMR, and other Krylov subspace methods [28, 1]. Krylov subspace algorithms tend to produce, at early iterations, solutions that resemble x_{true} more than later iterates. Therefore, the choice of the regularization parameter k , the stopping point for the iteration and the dimension of the subspace, is very important.

Another important family of regularization methods, termed **hybrid methods** [19, 15], was introduced by O'Leary and Simmons [28]. These methods combine a projection method with a direct regularization method such as TSVD or Tikhonov regularization. Since the dimension k is usually small relative to n , regularization of the restricted problem is much less expensive, but the end results can be very similar to those achieved by applying the same direct regularization technique to the original problem; see section 3.5.

2. Existing parameter selection methods. In this section, we discuss three parameter selection techniques that have been proposed in the literature. They differ in the amount of a priori information required as well as in the decision criteria.

The discrepancy principle [26] says that if δ is the expected value of $\|e\|_2$, then the regularization parameter should be chosen so that the norm of the residual corresponding to the regularized solution x_{reg} is $\tau\delta$; that is,

$$(2.1) \quad \|Ax_{reg} - b\|_2 = \tau\delta,$$

where $\tau > 1$ is some predetermined real number. Note that as $\delta \rightarrow 0$, $x_{reg} \rightarrow x_{true}$. Other methods based on knowledge of the variance are given, for example, in [3, 13, 7].

Generalized cross-validation (GCV) [11] does not depend on a priori knowledge about the noise variance. We find the parameter λ that minimizes the GCV functional

$$(2.2) \quad G(\lambda) = \frac{\|(I - AA_\lambda^\sharp)b\|_2^2}{(\text{trace}(I - AA_\lambda^\sharp))^2},$$

where A_λ^\sharp denotes the matrix that maps the right-hand side b onto the regularized solution x_λ . In Tikhonov regularization, for example, A_λ^\sharp is $(A^*A + \lambda^2 I)^{-1}A^*$.

The L-curve, the plot of the norm of the regularized solution versus the corresponding residual norm for each of a set of regularization parameter values, was introduced by Lawson and popularized by Hansen [17, 25]. Intuitively, the best regularization parameter should lie on the corner of the L-curve, since for values higher than this, the residual increases without reducing the norm of the solution much, while for values smaller than this, the norm of the solution increases rapidly without much decrease in residual. In practice, only a few points on the L-curve are computed and the corner is located by estimating the point of maximum curvature [20].

The appropriate choice of regularization parameter—especially for projection algorithms—is a difficult problem, and each method has severe flaws. The discrepancy principle is convergent as the noise goes to zero, but it relies on information that

TABLE 2.1

Summary of additional flops needed to compute the regularization parameter for each of four regularization methods with various parameter selection techniques. Notation: q is the cost of multiplication of a vector by A ; p is the number of discrete parameters that must be tried; k is the dimension of the projection; m and n are problem dimensions.

	Basic cost	Added cost		
		Disc.	GCV	L-curve
Tikhonov	$O(mn^2)$	$O(p(m+n))$	$O(p(n+m))$	$O(p(m+n))$
TSVD	$O(mn^2)$	$O(m)$	$O(m)$	$O(m+n)$
Rust's TSVD	$O(mn^2)$	$O(m \log m)$	$O(m \log m)$	$O(m \log m)$
Projection	$O(qk)$	0	$O(q)$	$O(q)$

is often unavailable or erroneous. Even with a correct estimate of the variance, the solutions tend to be oversmoothed [21, p. 96]. (See also the discussion in section 6.1 of [17].) One noted difficulty with GCV is that G can have a very flat minimum, making it difficult to determine the optimal λ numerically [37]. The L-curve is usually more tractable numerically, but its limiting properties are nonideal. The solution estimates fail to converge to the true solution as $n \rightarrow \infty$ [38] or as the error norm goes to zero [8]. All methods that assume no knowledge of the error norm—including GCV—have this latter property [8].

For further discussion and references about parameter choice methods, see [7, 19]. The cost of these methods is tabulated in Table 2.1.

2.1. Previous work on parameter choice for hybrid methods. At first glance, it appears that for Tikhonov regularization, multiple systems of the form (1.4) must be solved in order to evaluate candidate values of λ for the discrepancy principle or the L-curve.

Chan and Ng [5] note that the systems involve matrices $C(\lambda) = A^*A + \lambda I$, which they solve using a Galerkin projection method on a sequence of “seed” systems. Although economical in storage, this is unnecessarily expensive in time because they do not exploit the fact that for each fixed k , the Krylov subspace $\mathcal{K}_k(A^*b, C(\lambda))$ is the same for all values of λ .

Frommer and Maass [10] propose two algorithms for approximating the λ that satisfies the discrepancy principle (2.1). The first is a “truncated conjugate gradient (CG)” approach, solving k systems of the form (1.4), truncating the iterative process early for large λ , and using previous solutions as starting guesses for later problems. Like Chan and Ng, this algorithm does not exploit the redundant Krylov subspaces. In the second method, however, they update the CG iterates for all k systems simultaneously, stopping their “shifted CG” algorithm when $\|Ax_\lambda - b\|_2 \leq \tau\delta$ for one of their λ values. The methods we propose in section 3 will usually require less work than the shifted CG algorithm because of less overhead.

Calvetti, Golub, and Reichel [4] use upper and lower bounds on the L-curve, generated by the matrices $C(\lambda)$ using a Lanczos bidiagonalization process, to approximate the best parameter for Tikhonov regularization before projection.

Kaufman and Neumaier [22] suggest an envelope guided conjugate gradient approach for the Tikhonov L-curve problem. Their method is necessarily somewhat more expensive than ours because they maintain nonnegativity constraints on the variables.

Substantial work has also been done on TSVD regularization of the projected problems. Björck, Grimme, and van Dooren [2] use GCV to determine the truncation point for the projected SVD. Their emphasis is on maintaining an accurate factor-

ization when many iterations are needed, using full reorthogonalization and implicit restart strategies. O'Leary and Simmons [28] take the viewpoint that the problem should be preconditioned appropriately so that a massive number of iterations is unnecessary. That viewpoint is echoed in this current work, so we implicitly assume that the problem has been preconditioned [28] so that $A = M^{-1}\hat{A}$ and $b = M^{-1}\hat{b}$, where \hat{A} and \hat{b} are the original data and M is a preconditioning matrix. See [16, 27, 24, 23] for preconditioners appropriate for certain types of ill-posed problems.

3. Regularizing the projected problem. In this section we categorize a dozen approaches to regularization of the projected problem that arise from using Krylov methods, giving enough detail to make the costs apparent and to show that the ideas are easy to program. Many Krylov methods have been proposed; for ease of exposition we focus on just one of these: the LSQR algorithm of Paige and Saunders [30].

LSQR iteratively computes a bidiagonalization related to that introduced by Golub and Kahan [12]. After k iterations, it has effectively computed three matrices: an upper-bidiagonal matrix B_k and two matrices $U_k \equiv [u_1, \dots, u_k]$ and $V_k \equiv [v_1, \dots, v_k]$, with orthonormal columns, related by

$$(3.1) \quad b = \beta_1 u_1 = \beta_1 U_{k+1} e_1,$$

$$(3.2) \quad AV_k = U_{k+1} B_k,$$

$$(3.3) \quad A^T U_{k+1} = V_k B_k^T + \alpha_{k+1} v_{k+1} e_{k+1}^T,$$

where e_i denotes the i th unit vector.

In numeric computations, the columns of U_k and V_k can fail to be orthonormal. This has never given us convergence difficulties, but if it becomes troublesome, there are well-known techniques to handle it [31, 32, 36, 6].

Now suppose we want to solve

$$(3.4) \quad \min_{x \in \mathcal{S}} \|b - Ax\|_2,$$

where \mathcal{S} denotes the k -dimensional subspace spanned by the first k vectors v_i . The solution we seek is of the form $x^{(k)} = V_k y^{(k)}$ for some vector $y^{(k)}$ of length k . Define $r^{(k)} = b - Ax^{(k)}$ to be the corresponding residual and observe that

$$\begin{aligned} r^{(k)} &= \beta_1 u_1 - AV_k y^{(k)} \\ &= U_{k+1} (\beta_1 e_1 - B_k y^{(k)}). \end{aligned}$$

Since U_{k+1} has, in exact arithmetic, orthonormal columns, the projected problem we wish to solve is

$$(3.5) \quad \min_{y^{(k)}} \|\beta_1 e_1 - B_k y^{(k)}\|_2.$$

Solving this minimization problem is mathematically equivalent to solving the normal equations involving the bidiagonal matrix

$$(3.6) \quad B_k^* B_k y^{(k)} = \beta_1 B_k^* e_1,$$

although more stable means are used in practice. Typically k is small, so reorthogonalization to combat round-off error might or might not be necessary. The matrix B_k may be ill-conditioned because some of its singular values approximate some of the small singular values of A . Therefore, solving the projected problem might not yield

TABLE 3.1

Summary of flops for projection plus inner regularization with various parameter selection techniques, in addition to the $O(qk)$ flops required for projection itself. Here k is the number of iterations (i.e., the size of the projection) taken and p is the number of discrete parameters that must be tried.

Projection plus –	Disc.	GCV	L-curve
Tikhonov	$O(pk)$	$O(k^3)$	$O(pk)$
TSVD	$O(k^3)$	$O(k^3)$	$O(k^3)$
Rust's	$O(k^3)$	$O(k^3)$	$O(k^3)$

TABLE 3.2

Summary of additional storage for each of four regularization methods under each of three parameter selection techniques. The original matrix is $m \times n$ with q nonzeros, p is the number of discrete parameters that must be tried, k iterations are used in projection, and the factorizations are assumed to take \hat{q} storage.

	Basic cost	Added cost		
		Disc.	GCV	L-curve
Tikhonov	$O(\hat{q})$	$O(1)$	$O(p)$	$O(p)$
TSVD	$O(\hat{q})$	$O(1)$	$O(m)$	$O(m)$
Rust's TSVD	$O(\hat{q})$	$O(m)$	$O(m)$	$O(m)$
Projection	$O(kn)$	$O(1)$	$O(k)$	$O(k)$

TABLE 3.3

Summary of storage, not including storage for the matrix, for projection plus inner regularization approach and various parameter selection techniques. Here p denotes the number of discrete parameters tried. Each of these regularization methods also requires us to save the basis V or else regenerate it in order to reconstruct x .

Projection plus –	Disc.	GCV	L-curve
Tikhonov	$O(1)$	$O(p)$	$O(p)$
TSVD	$O(1)$	$O(k)$	$O(k)$
Rust's TSVD	$O(k)$	$O(k + p)$	$O(k + p)$

a good solution $y^{(k)}$, but we can use any of the methods of section 2 to regularize this projected problem; we discuss options in detail below.

If we used the algorithm GMRES [35] instead of LSQR, we would derive similar relations. Here, though, the U and V matrices are identical and the B matrix is upper Hessenberg rather than bidiagonal. Conjugate gradients would yield similar relationships.

For cost comparisons for these methods, see Tables 2.1 and 3.1. Storage comparisons are given in Tables 3.2 and 3.3.

3.1. Regularization by projection. As mentioned earlier, if we terminate the iteration after k steps, we have projected the solution onto a k -dimensional subspace and this has a regularizing effect that is sometimes sufficient. Determining the best value of k can be accomplished, for instance, by one of our three methods of parameter choice. Efficient implementation relies on LSQR recurrences for determining $\|r^{(k)}\|$ and $\|x^{(k)}\|$ cheaply, without computing either $r^{(k)}$ or $x^{(k)}$ [30, 34].

For the **discrepancy principle**, we stop the iteration for the smallest value of k for which $\|r_k\| \leq \tau\delta$.

To apply **GCV**, we note that in LSQR (see section 3.1), the operator AA^\sharp is given by $U_{k+1}B_kB_k^\dagger U_{k+1}^*$, where B_k^\dagger is the pseudoinverse of B_k . Thus from (2.2), the

GCV functional is [19]

$$G(k) = \frac{\|r^{(k)}\|_2^2}{(m-k)^2}.$$

We note that there are in fact two distinct definitions for A^\sharp and hence two definitions for the denominator in $G(k)$; for small enough k , the two are comparable, and the definition we use here is less expensive to calculate [19, section 7.4].

To determine the **L-curve** associated with LSQR, values of $\|r^{(k)}\|_2$ and $\|x^{(k)}\|_2$ are needed for several values of k . In using this method or GCV, one must go a few iterations beyond the optimal k in order to verify the optimum [20].

3.2. Regularization by projection plus TSVD. If projection alone does not regularize, then we can compute the TSVD regularized solution to the projected problem (3.6). We need the SVD of the $(k+1) \times k$ matrix B_k . This requires $O(k^3)$ operations but can also be computed from the SVD of B_{k-1} in $O(k^2)$ operations [14].

Clearly, we still need to use some type of parameter selection technique to find a good value of $\ell(k)$. First, notice that it is easy to compute the norms of the residual and the solution resulting from retaining only the ℓ largest singular values. If ξ_{jk} is the component of e_1 in the direction of the j th left singular vector of B_k , and if γ_j is the j th singular value (ordered largest to smallest), then the residual and solution 2-norms are

$$(3.7) \quad \|r_\ell^{(k)}\| = \beta_1 \left(\sum_{j=\ell(k)+1}^{k+1} \xi_{jk}^2 \right)^{1/2} \quad \text{and} \quad \|x_\ell^{(k)}\| = \beta_1 \left(\sum_{j=1}^{\ell(k)} \left(\frac{\xi_{jk}}{\gamma_j} \right)^2 \right)^{1/2}.$$

Using this fact, we can use any of our three sample methods.

For the **discrepancy principle** we choose $\ell(k)$ to be the smallest value for which $\|r_\ell^{(k)}\| \leq \tau\delta$, if such a value exists. As k increases, the number of neglected singular values will be monotonically nondecreasing (exact arithmetic).

The **GCV** functional for the k th projected problem is obtained by substituting B_k for A and B_k^\sharp for A^\sharp , and substituting the expression of the residual in (3.7) for the numerator in (2.2):

$$G_k(\ell) = \frac{\beta_1^2 \sum_{j=\ell+1}^{k+1} \xi_{jk}^2}{(k-\ell+1)^2}.$$

We now have many **L-curves**, one for each value of k . The coordinate values in (3.7) form the discrete L-curve for a given k , from which the desired value of $\ell(k)$ can be chosen without forming the approximate solutions or residuals.

3.3. Regularization by projection plus Rust's TSVD. As in standard TSVD, to use Rust's version of TSVD for regularization of the projected problem requires computing the SVD of the $(k+1) \times k$ matrix B_k . Using the previous notation, Rust's strategy is to set

$$y_\rho^{(k)} = \sum_{j \in \mathcal{I}_\rho^{(k)}} \frac{\xi_{jk}}{\gamma_j} q_j^{(k)},$$

where $q_j^{(k)}$ are the right singular vectors of B_k and $\mathcal{I}_\rho^{(k)} = \{i < k+1 : |\xi_{ik}| > \rho\}$. We focus on three ways to determine ρ .

For the **discrepancy principle**, the norm of the residual of the regularized solution is given by $\|r_\rho^{(k)}\|_2 = \beta_1 (\sum_{j \notin \mathcal{I}_\rho^{(k)}} \xi_{jk}^2)^{1/2}$. According to the discrepancy principle, we must choose ρ so that the residual is less than $\tau\delta$. In practice, this would require that the residual be evaluated by sorting the values $|\xi_{ik}|$ and adding terms in that order until the residual norm is less than $\tau\delta$.

For **GCV**, let $\text{card}(\mathcal{I}_\rho^{(k)})$ denote the cardinality of the set $\mathcal{I}_\rho^{(k)}$. From (2.2), it is easy to show that the GCV functional corresponding to the projected problem for this regularization technique is given by

$$G_k(\rho) = \frac{\beta_1^2 \sum_{j \in \mathcal{I}_\rho^{(k)}} \xi_{jk}^2}{(k+1 - \text{card}(\mathcal{I}_\rho^{(k)}))^2}.$$

In practice, for each k we first sort the values $|\xi_{ik}|, i = 1, \dots, k$, from smallest to largest. Then we define k discrete values ρ_j to be equal to these values with ρ_1 being the smallest. We set $\rho_0 = 0$. Note that because the values of $\rho_j, j = 1, \dots, k$, are the sorted magnitudes of the SVD expansion coefficients, we have

$$G_k(\rho_j) = \frac{\beta_1^2 (|\xi_{(k+1),k}|^2 + \sum_{i=1}^j \rho_i^2)}{(j+1)^2}, \quad j = 0, \dots, k.$$

Finally, we take the regularization parameter to be the ρ_j for which $G_k(\rho_j)$ is a minimum.

As with standard TSVD, we now have one **L-curve** for each value of k . For fixed k , if we define the $\rho_j, j = 0, \dots, k$, as we did for GCV above and we reorder the γ_i in the same way that the $|\xi_{ik}|$ were reordered when sorted, then we have

$$\|x_{\rho_j}^{(k)}\|_2^2 = \beta_1^2 \sum_{i=j+1}^k \left(\frac{\rho_i}{\gamma_i} \right)^2; \quad \|r_{\rho_j}^{(k)}\|_2^2 = \beta_1^2 \left(|\xi_{(k+1),k}|^2 + \sum_{i=1}^j \rho_i^2 \right), \quad j = 0, \dots, k.$$

When these solution and residual norms are plotted against each other as functions of ρ , the value of ρ_j corresponding to the corner is selected as the regularization parameter.

3.4. Regularization by projection plus Tikhonov. Finally, let us consider using Tikhonov regularization to regularize the projected problem (3.5) for some integer k . Thus, for a given regularization parameter λ , we would like to solve

$$(3.8) \quad \min_y \|\beta_1 e_1 - B_k y\|_2^2 + \lambda^2 \|y\|_2^2.$$

The solution $y_\lambda^{(k)}$ satisfies

$$(3.9) \quad (V_k^* A^* A V_k + \lambda^2 I) y_\lambda^{(k)} = V_k^* A^* b.$$

We need to address how to choose a suitable value of λ .

For the **discrepancy principle**, note that in exact arithmetic, we have

$$(3.10) \quad r_\lambda^{(k)} = b - A x_\lambda^{(k)} = U_{k+1}^* (\beta_1 e_1 - B_k y_\lambda^{(k)}).$$

Hence $\|B_k y_\lambda^{(k)} - \beta_1 e_1\|_2 = \|r_\lambda^{(k)}\|_2$. Therefore, to use the discrepancy principle requires that we choose λ so that $\|r_\lambda^{(k)}\|_2 \leq \tau\delta$ with p discrete trial values λ_j . For a given k ,

we take λ to be the largest value λ_j for which $\|r_\lambda^{(k)}\|_2 < \tau\delta$, if it exists; if not, we increase k and test again.

For **GCV**, let us define $(B_k)_\lambda^\dagger$ to be the operator mapping the right-hand side of the projected problem onto the regularized solution of the projected problem:

$$(B_k)_\lambda^\dagger = (B_k^* B_k + \lambda^2 I)^{-1} B_k^*.$$

Given the SVD of B_k as above, the denominator in the GCV functional defined for the projected problem (refer to (2.2)) is

$$\left(k + 1 - \sum_{j=1}^k \frac{\gamma_j^2}{\gamma_j^2 + \lambda^2} \right)^2.$$

The numerator is simply $\|r_\lambda^{(k)}\|_2^2$. For values of $k \ll n$, it is feasible to compute the singular values of B_k .

The **L-curve** is comprised of the points $(\|B_k y_\lambda^{(k)} - \beta_1 e_1\|_2, \|y_\lambda^{(k)}\|_2)$. But using (3.10) and the orthonormality of the columns of V_k , we see these points are precisely $(\|r_\lambda^{(k)}\|_2, \|x_\lambda^{(k)}\|_2)$. For p discrete values of λ , $\lambda_i, 1 \leq i \leq p$, the quantities $\|r_{\lambda_i}^{(k)}\|_2$ and $\|x_{\lambda_i}^{(k)}\|_2$ can be obtained by updating their respective estimates at the $(k-1)$ st iteration.¹

3.5. Correspondence between direct regularization and projection plus regularization. In this section, we demonstrate why the projection plus regularization approaches can be expected to yield regularized solutions nearly equivalent to the direct regularization counterpart. The following theorem, a simple corollary of the invariance of Krylov sequences under shifts, establishes the desired result for the case of Tikhonov vs. projection plus Tikhonov.

THEOREM 3.1. *Fix $\lambda > 0$ and define $x_\lambda^{(k)}$ to be the k th iterate of conjugate gradients applied to the Tikhonov problem*

$$(A^* A + \lambda^2 I)x = A^* b.$$

Let $y_\lambda^{(k)}$ be the exact solution to the regularized projected problem

$$(B_k^* B_k + \lambda^2 I)y = B_k^*(\beta_1 e_1),$$

where B_k, V_k are derived from the original problem $A^ A = A^* b$, and set $z_\lambda^{(k)} = V_k y_\lambda^{(k)}$. Then $z_\lambda^{(k)} = x_\lambda^{(k)}$.*

Proof. See [15, p. 301]. \square

Let us compare TSVD regularization applied to the original problem to the projection plus TSVD approach. Direct computation convinces us that the two methods compute the same regularized solution if $k = n$ and arithmetic is exact. An approximate result holds in exact arithmetic when we take k iterations, with $\ell \leq k \leq n$. Let the SVD of B_k be denoted by $B_k = Z_k \Gamma_k Q_k^T$, and define the $s \times \ell$ matrix $W_{s,\ell}$ as

$$W_{s,\ell} = \begin{bmatrix} I \\ 0 \end{bmatrix}.$$

¹The technical details of the approach are found in [29, pp. 197–198], from which we obtain $\|r_\lambda^{(k)}\| = \sqrt{\|\bar{r}_\lambda^{(k)}\|^2 - \lambda^2 \|x_\lambda^{(k)}\|^2}$. The implementation details for estimating $\|x_\lambda^{(k)}\|$ and $\|\bar{r}_\lambda^{(k)}\|$ were taken from the Paige and Saunders algorithm at <http://www.netlib.org/linalg/lsqqr>.

Then the regularized solution obtained from the TSVD regularization of the projected problem is

$$x_{reg}^{(k)} = V_k(Q_k W_{k,\ell} \Gamma_{k,1}^{-1} W_{k+1,\ell}^T Z_k^T U_k^T b),$$

where $\Gamma_{k,1}$ denotes the leading $\ell \times \ell$ principal submatrix of Γ_k . If k is taken to be sufficiently larger than ℓ so that $V_k Q_k W_{k,\ell} \approx \hat{V} W_{n,\ell}$, $W_{k+1,\ell}^T Z_k^T U_{k+1}^T \approx W_{n,\ell}^T \hat{U}^T$, and $\Gamma_{k,1} \approx \Sigma_1$ with Σ_1 the leading principal submatrix of Σ , then we expect $x_{reg}^{(k)}$ to be a good approximation to x_ℓ . This is made more precise in the following theorem.

THEOREM 3.2. *Let $k \geq \ell$ such that*

$$(V_k Q_k W_{k,\ell}) = \hat{V}_1 + E_1 \quad \text{with } \|E_1\| \leq \delta_1 \ll 1,$$

$$(U_{k+1} Z_k W_{k+1,\ell}) = \hat{U}_1 + E_2 \quad \text{with } \|E_2\| \leq \delta_2 \ll 1,$$

where \hat{V}_1 and \hat{U}_1 contain the first ℓ columns of \hat{V} and \hat{U} , respectively. Let $D = \text{diag}(d_1, \dots, d_\ell)$ satisfy

$$\Gamma_{k,1} = \Sigma_1 + D \quad \text{with } |d_i| \leq \delta_3 \ll 1.$$

Then

$$\|x_{reg}^{(k)} - x_\ell\| \leq \max_{1 \leq i \leq \ell} \frac{1}{\sigma_i + d_i} \left(\frac{\delta_3}{\sigma_\ell} + 3 \max(\delta_1, \delta_2) \right) \|b\|.$$

Proof. Using the representations $x_\ell = \hat{V}_1 \Sigma_1^{-1} \hat{U}_1^T b$ and $x_{reg}^{(k)} = (\hat{V}_1 + E_1) \Gamma_{k,1}^{-1} (\hat{U}_1^T + E_2^T) b$, we obtain

$$\|x_{reg}^{(k)} - x_\ell\| \leq (\|\Gamma_{k,1}^{-1} - \Sigma_1^{-1}\| + \|\Gamma_{k,1}^{-1}\| \|E_2\| + \|E_1\| \|\Gamma_{k,1}^{-1}\| + \|E_1\| \|\Gamma_{k,1}^{-1}\| \|E_2\|) \|b\|,$$

and the conclusion follows from bounding each term. \square

Note that typically $\sigma_\ell \gg \sigma_n$ so that $1/\sigma_\ell$ is not too large. The bound says that the better LSQR captures the first ℓ singular values and vectors, the more we are assured the solution obtained by projection plus TSVD is close to the TSVD regularized solution to the original problem. For some results relating to the value of k necessary for the hypothesis of the theorem to hold, refer to the theory of Kaniel-Paige and Saad [31, section 12.4]. There is no universal recipe, but if k is large enough that the projected problem satisfies the discrete Picard condition, then this is some indication that the approximability property holds.

4. Numerical results. In this section, we present three numerical examples. All experiments were carried out using Matlab with IEEE double precision floating point arithmetic. Where noted, we made use of certain routines in Hansen's Regularization Tools [18]. Since the exact, noise-free solutions were known in these examples, we evaluated the methods using the relative, 2-norm difference between the regularized solutions and the exact solutions. When we applied Rust's method to the original problem, the ρ_i were taken to be the magnitudes of the spectral coefficients of b sorted in increasing order.

TABLE 4.1

Example 1: comparison of $\|x_{true} - x_{reg}\|_2 / \|x_{true}\|_2$ for each of four regularization methods on the original problem, where the regularization method was chosen using methods indicated. The parameter values selected for each method are indicated in parentheses.

	Disc.		GCV		L-curve		Optimal	
Tikhonov	(1.6E-1)	2.2E-2	(8.0E-2)	2.2E-2	(4.0E-2)	4.3E-2	(1.3E-1)	2.1E-2
TSVD	(6)	1.1E-1	(9)	1.6E-2	(10)	1.6E-2	(9)	1.6E-2
Rust's TSVD	(1.6E-2)	2.5E-2	(5.3E-5)	2.2E+4	(1.6E-2)	2.5E-2	(1.6E-2)	2.5E-2
Projection	(5)	2.5E-2	(5)	2.5E-2	(10)	2.2E-2	(9)	2.2E-2

4.1. Example 1. The 200×200 matrix A and true solution x_{true} for this example were generated using the function `phillips` in Hansen's Regularization Tools. We generated $b_{true} = Ax_{true}$ and then computed the noisy vector b as $b + e$, where e was generated using the Matlab `randn` function and was scaled so that the noise level, $\frac{\|e\|}{\|b_{true}\|}$, was 5×10^{-3} . The condition number of A was on the order of 4×10^7 .

Table 4.1 displays the values of the regularization parameters chosen when the original problem was solved using one of the three parameter selection techniques together with one of the four regularization methods. We set $\tau\delta$ for the discrepancy principle to be $8E-2$, close to the value $\|e\|_2 = 7.65E-2$.

The last column in the table gives the value of the parameter that yielded a regularized solution with minimum relative error. Several values of λ were tested: $\log_{10} \lambda = -4, -3.9, \dots, 0$. The relative error values for regularized solutions corresponding to the parameters are also presented in this table. The GCV and L-curve parameters for projection were determined after 15 iterations. Note that using GCV to determine a regularization parameter for Rust's TSVD resulted in an extremely noisy solution with huge error.

The corners of the L-curves for the Tikhonov, projection, and TSVD methods were determined using Hansen's `lcorner` function, with the modification that sometimes points not strictly on the portion of the curve that was L-shaped (that is, points with very large residual or very small residual) were not considered (otherwise, a false corner resulted); this was most often a concern with the TSVD method. Since the corner was so clearly defined for Rust's method but the function had trouble automatically finding the corner, the corner was picked manually.

Next, we projected using LSQR and then regularized the projected problem with one of the other three regularization methods together with one of the three parameter selection techniques. Results at iterations 10 and 25 are given in Tables 4.2 and 4.3, respectively. As before, the `lcorner` routine was used to determine the corners of the respective L-curves, with the modifications as mentioned above.

Comparing Tables 4.1 and 4.2, we observe that using either the discrepancy principle or the L-curve, 10 steps of projection plus Tikhonov gives results as good as or much better than if those techniques had been used with Tikhonov on the original problem. A similar statement can be made for projection plus Rust's TSVD when any of the 3 selection methods are used and for projection plus TSVD when the discrepancy principle is used. After 25 iterations, the errors for projection plus Tikhonov or Rust's TSVD closely resemble the errors in Table 4.1 with one exception. We note that at 25 iterations, the parameters chosen for projection plus Tikhonov by the discrepancy principle or the L-curve method and their corresponding errors are identical to those chosen for the original problem.

In fact, the L-curve, GCV, and discrepancy methods applied to the projected

TABLE 4.2

Example 1, iteration 10: comparison of $\|x_{true} - x_{reg}\|_2 / \|x_{true}\|_2$ for projection plus Tikhonov, TSVD, and Rust's TSVD. The parameter values for each method are indicated in parentheses.

	Disc.		GCV		L-curve		Optimal	
Tikhonov	(1.6E-1)	2.1E-2	(2.5E-2)	2.5E-2	(2.0E-4)	2.2E-2	(2.0E-2)	2.0E-2
TSVD	(7)	2.5E-2	(7)	2.5E-2	(10)	2.2E-2	(10)	2.2E-2
Rust's TSVD	(9.7E-3)	2.5E-2	(9.7E-3)	2.5E-2	(5.5E-4)	2.2E-2	(9.1E-3)	2.1E-2

TABLE 4.3

Example 1, iteration 25: comparison of $\|x_{true} - x_{reg}\|_2 / \|x_{true}\|_2$ for projection plus Tikhonov, TSVD, and Rust's TSVD. The parameter values are given in parentheses.

	Disc.		GCV		L-curve		Optimal	
Tikhonov	(1.6E-1)	2.2E-2	(2.0E-1)	2.3E-2	(4.0E-2)	4.3E-2	(1.3E-1)	2.1E-2
TSVD	(17)	2.5E-2	(17)	2.5E-2	(21)	2.4E-2	(19)	1.6E-2
Rust's TSVD	(2.0E-2)	2.5E-2	(2.0E-2)	2.5E-2	(1.5E-2)	2.5E-2	(1.5E-2)	2.5E-2

problem with Tikhonov regularization consistently chose the same parameter for future iterations (see Figure 4.1, for instance), and correspondingly the errors remain constant; however, the results at earlier iterations are actually better than after the parameter on the projected problem has converged to the L-curve parameter on the original. For the projection plus TSVD, both the discrepancy principle and GCV method yielded parameters for which the solutions had similar errors from one iteration to the next for at least the first 80 iterations (see the top of Figure 4.2); the L-curve behaved slightly less consistently for iterations beyond about 50. Discrepancy and GCV when applied to projection plus Rust's TSVD also gave consistent solutions for about 40 iterations, after which the GCV solutions began to grow very large in error, much like GCV applied to the original problem (refer to the bottom of Figure 4.2).

4.2. Example 2. The 3969×3969 matrix A for this example was a symmetric, block Toeplitz matrix with Toeplitz blocks formed according to $A = T \otimes T$. Here T is a symmetric, banded Toeplitz matrix with entries $T_{i,j} = t_{i-j}$; the nonzero entries in the first row were $t_k = (\sin(k/B)/(k/B))^2$, $0 \leq k \leq 4$, $B = .8$. The singular values of this matrix range from 5.7 to 8.6×10^{-8} but do not decay very quickly, and the matrix has a condition number of about 7×10^7 . x was obtained by stacking by columns the 63×63 image that was zero except for a rectangle with value 1 from rows 20 to 49, columns 4 to 24, and another rectangle with value .8 at rows 23 to 53, columns 29 to 52. We generated $b_{true} = Ax_{true}$ and then computed the noisy vector b as $b + e$, where e was generated using the Matlab `randn` function and was scaled so that the noise level, $\frac{\|e\|}{\|b_{true}\|}$, was 2×10^{-3} .

We generated our discrete λ_i using $\log_{10} \lambda = -4, -4.9, \dots, 0$. The norm of the noise vector was $3.66E-1$, so we took $\tau\delta = 4.00E-1$ for the discrepancy principle.

In this example, when no preconditioning was used, it took 90 iterations for LSQR to reach a minimum relative error of $7.93E-2$. Likewise, the dimension k of the projected problem had to be at least 90 to obtain good results with the projection-plus-regularization approaches and even larger for the parameter selection techniques to work well on the projected problem. Therefore, for the projection based techniques, we chose to work with a left preconditioned system (refer to the discussion at the end of section 2.1). Our preconditioner was chosen as in [23] where the parameter defining the preconditioner was taken to be $m = 2080$. Results for right preconditioning were

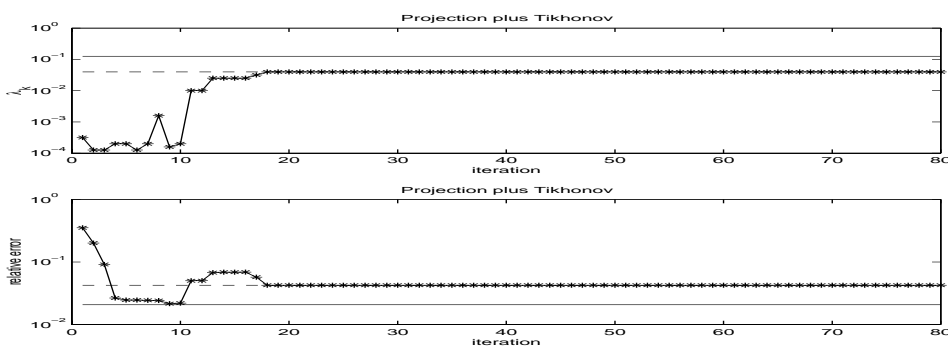


FIG. 4.1. Example 1. Top: λ_k as selected by L-curve method; bottom: relative error for corresponding solution. The solid line indicates the optimal value on the original problem, and the dashed line indicates value selected by L-curve on the original problem.

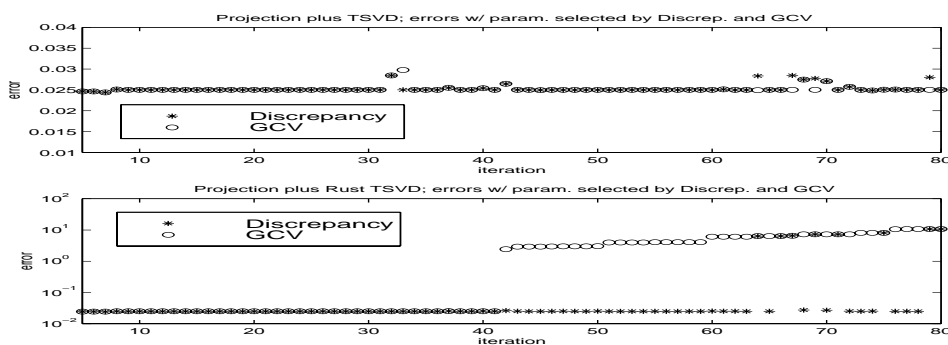


FIG. 4.2. Example 1. Relative error between computed and exact solutions for projection plus TSVD (top) and projection plus Rust's TSVD (bottom) when the parameters for the projected problem are selected by either the discrepancy principle (*) or GCV method (o).

similar, although the errors were not quite as small. On other examples, though, we found that right preconditioning by this type of preconditioner was only effective in certain instances, even when left preconditioning was effective.²

The results of the resulting regularization for the original problem parameters are given in Table 4.4. We note that GCV with Rust's TSVD was ineffective. Also, after 50 iterations on the left preconditioned system, the GCV functional for projection was still decreasing, so the value in Table 4.4 corresponds to the value after 50 iterations. The L-curve parameter in the table was determined after 20 iterations.

Although we projected using LSQR, we note that since the matrix and preconditioner were symmetric, we could have used MINRES as in [23]. The results in each case at iterations 10, 20, and 40 are given in Tables 4.5, 4.6, and 4.7, respectively, and we summarize results up to 60 iterations in the discussion below.

Again, we used the `lcorner` routine to determine the corners of the respective L-curves, with the modification that for 20 iterations and beyond for TSVD, we first removed points on the curve with residual norm greater than 10 to avoid detecting a false corner.

²In the language of [23], right preconditioning worked well only when K was a very good approximation to C so that right preconditioning did not mix noise into early iterates; left preconditioning was not nearly as sensitive to the approximation on the transition and noise subspaces.

TABLE 4.4

Example 2: comparison of $\|x_{true} - x_{reg}\|_2 / \|x_{true}\|_2$ for each of four regularization methods on the original problem. The parameter values are given in parentheses. The projection was performed on a left preconditioned system.

	Disc.		GCV		L-curve		Optimal
Tikhonov	(1.6E-1)	8.5E-2	(5.0E-2)	8.0E-2	(3.2E-3)	5.3E-1	(6.3E-2) 7.8E-2
TSVD	(2073)	9.9E-2	(2534)	8.1E-2	(1509)	1.2E-1	(2521) 8.0E-2
Rust's TSVD	(2.1E-2)	7.6E-2	(9.2E-2)	4.0E+3	(1.6E-2)	2.3E-1	(2.0E-2) 7.6E-2
Projection	(2)	9.7E-2	(50+)	2.7E-1	(13)	8.3E-2	(8) 7.9E-2

TABLE 4.5

Example 2, iteration 10: comparison of $\|x_{true} - x_{reg}\|_2 / \|x_{true}\|_2$ for projection plus Tikhonov, TSVD, and Rust's TSVD. Parameter values are given in parentheses.

	Disc.		GCV		L-curve		Optimal
Tikhonov	(7.9E-2)	7.9E-2	(6.3E-2)	7.9E-2	(2.0E-4)	7.9E-2	(5.0E-2) 7.9E-2
TSVD	(6)	9.9E-2	(6)	7.9E-2	(8)	9.8E-2	(10) 7.9E-2
Rust's TSVD	(2.2E-1)	8.5E-2	(2.6E-1)	9.9E-2	(2.3E-1)	9.9E-2	(3.9E-4) 7.9E-2

TABLE 4.6

Example 2, iteration 20: comparison of $\|x_{true} - x_{reg}\|_2 / \|x_{true}\|_2$ for projection plus Tikhonov, TSVD, and Rust's TSVD. Parameter values are given in parentheses.

	Disc.		GCV		L-curve		Optimal
Tikhonov	(7.9E-2)	7.9E-2	(6.3E-2)	7.8E-2	(2.0E-4)	1.1E-1	(6.3E-2) 7.8E-2
TSVD	(12)	9.9E-2	(12)	9.9E-2	(19)	8.3E-2	(19) 8.3E-2
Rust's TSVD	(1.6E-1)	9.5E-1	(7.9E-2)	1.1E-1	(4.6E-2)	1.1E-1	(1.3E-1) 8.3E-2

TABLE 4.7

Example 2, iteration 40: comparison of $\|x_{true} - x_{reg}\|_2 / \|x_{true}\|_2$ for projection plus Tikhonov, TSVD, and Rust's TSVD. Parameter values are given in parentheses.

	Disc.		GCV		L-curve		Optimal
Tikhonov	(7.9E-2)	7.9E-2	(6.3E-2)	7.8E-2	(2.0E-1)	2.3E-1	(6.3E-2) 7.9E-2
TSVD	(24)	9.9E-2	(24)	9.9E-2	(28)	9.9E-2	(38) 8.3E-2
Rust's TSVD	(1.5E-1)	9.2E-2	(5.8E-2)	2.3E-1	(1.6E-1)	9.2E-2	(1.5E-1) 9.2E-2

Discrepancy and GCV consistently chose the same regularization parameter and hence gave the same error for projection plus Tikhonov for 10 to 60 iterations. From the tables, we see that these are not the same parameters as those chosen when applied to the original problem and that, in fact, the solutions for projection plus Tikhonov have smaller error. The errors for the solutions obtained using any of the 3 parameter selection methods applied to find ℓ for projection plus TSVD were also consistent for 10 to 60 iterations, as alluded to in the tables. Figure 4.3 shows the errors from iterations 5 to 60 for projection plus Tikhonov and projection plus TSVD when GCV is used. For Rust's TSVD, the L-curve and discrepancy rules are fairly consistent at picking parameters that give solutions with similar error from iteration to iteration. We note that GCV for Rust's TSVD picked parameters giving solutions with reasonably small errors, even though GCV for Rust's TSVD on the original problem failed, giving a solution with huge error. A similar statement can be made for the L-curve with projection plus Tikhonov.

Summarizing, we observe two phenomena. First, the parameters selected to regularize the projected problem can be different from those chosen on the original problem but still yield solutions of better or comparable error. Second, as this and the previous

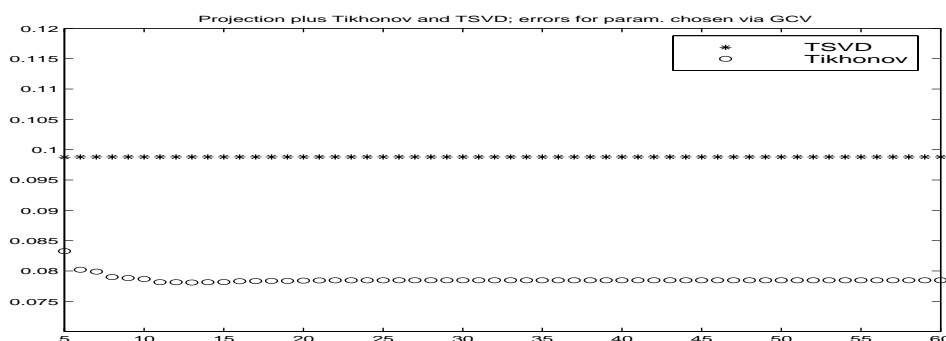


FIG. 4.3. Example 2: Errors for projection plus Tikhonov (*) and projection plus TSVD (o) when the regularization parameter for the projected problem was given by GCV.

TABLE 4.8

Example 3: comparison of $\|x_{true} - x_{reg}\|_2 / \|x_{true}\|_2$ for each of the 4 regularization methods on the original problem. Parameter values are given in parentheses. Those for GCV and the L-curve are those selected after 30 iterations.

	Disc.	GCV	L-curve	Optimal
Tikhonov	(1.0) 3.9E-1	(1.3) 4.0E-1	(5.0E-1) 4.1E-1	(7.9E-1) 3.9E-1
TSVD	(232) 4.2E-1	(400) 1.3E+4	(261) 4.0E-1	(241) 4.0E-1
Rust's TSVD	(3.0E-1) 7.4E+2	(1.8E-1) 1.2E+4	(3.7) 4.6E-1	(3.9E-1) 4.6E-1
Projection	(9) 4.0E-1	(23) 4.3E-1	(16) 4.0E-1	(12) 3.9E-1

example show, loss of orthogonality does not seem to hamper the parameter selection process, at least not for a reasonable number of iterations. This may be due to the fact that the parameter selection methods are applied directly to the projected problem: for example, the denominator of our GCV function for projection plus TSVD is different from the denominator of the GCV function given in [2, (3.8)].

4.3. Example 3. Our final example is from the field of computed tomography. In this example, the true vector x corresponded to the 20×20 image created with the `phantom.m` function. The matrix A was the corresponding 561×400 Radon transform matrix where it is understood that the data was taken at angles from 0 to 179 degrees in increments of 11 degrees. The matrix itself was computed (albeit naively) in Matlab column by column using successive applications of `radon.m` on images of point sources. The singular values fall off very slowly at first (the first 260 of the 400 singular values range between 18 and about 1) after which they fall off rapidly, resulting in a condition number for A of about 10^7 .

Since the norm of the noise vector was about 3.44, we took the tolerance for the discrepancy principle to be 3. The discrete values λ_i used for Tikhonov regularization were 51 evenly log-spaced points between 10^{-4} and 10^1 . The results computed using discrepancy, GCV, and L-curve methods for Tikhonov, TSVD, Rust's TSVD, and projection on the original problem are given in Table 4.8.

Table 4.9 gives the results after 10 iterations of LSQR. Notice that the errors for the projection plus Tikhonov solutions via GCV and L-curve are slightly better than the corresponding error for Tikhonov without projection at only 10 iterations. Also interesting is the fact that at 10 iterations the discrepancy and GCV methods for projection plus Rust's TSVD give solutions with reasonable errors, whereas these techniques give solutions with very large errors when applied to the original problem.

TABLE 4.9

Example 3, iteration 10: comparison of $\|x_{true} - x_{reg}\|_2 / \|x_{true}\|_2$ for projection plus Tikhonov, TSVD, and Rust's TSVD. Parameter values are given in parentheses.

	Disc.		GCV		L-curve		Optimal	
Tikhonov	(1.0)	4.0E-1	(2.2)	4.0E-1	(1.6E-4)	3.9E-1	(4.0E-1)	4.0E-1
TSVD	(10)	3.9E-1	(1)	8.6E-1	(5)	8.3E-1	(10)	3.9E-1
Rust's TSVD	(1.0)	3.9E-1	(1.5)	4.0E-1	(2.2)	4.0E-1	(0.0)	3.9E-1

TABLE 4.10

Example 3, iteration 40: comparison of $\|x_{true} - x_{reg}\|_2 / \|x_{true}\|_2$ for projection plus Tikhonov, TSVD, and Rust's TSVD. Parameter values are given in parentheses.

	Disc.		GCV		L-curve		Optimal	
Tikhonov	(1.0)	3.9E-1	(1.2)	4.1E-1	(5.0E-1)	4.1E-1	(7.9E-1)	3.9E-1
TSVD	(37)	4.0E-1	(15)	7.8E-1	(39)	4.1E-1	(38)	4.0E-1
Rust's TSVD	(6.0E-1)	4.2E-1	(1.2)	4.1E-1	(2.7E-1)	4.1E-1	(6.6E-1)	4.0E-1

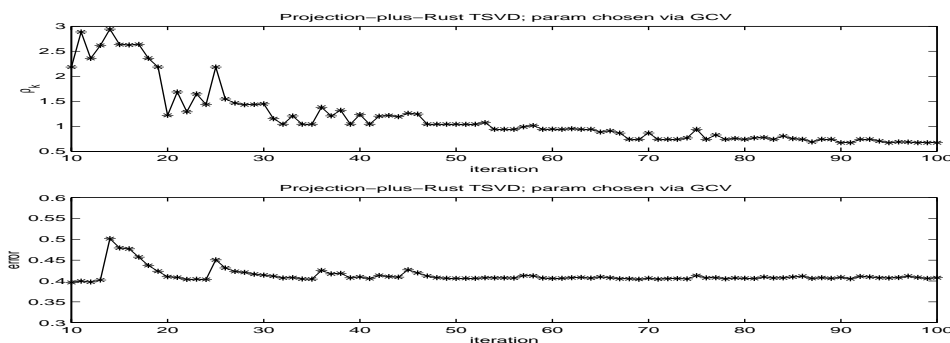


FIG. 4.4. Example 3. Top: Value of ρ_k selected by GCV for projection plus Rust's TSVD; Bottom: Relative error of the corresponding solutions.

Table 4.10 shows the parameters and the errors after 40 iterations. From these results, we see that the L-curve for projection plus Tikhonov eventually gives the same regularization parameter and same solution error as when applied to the larger problem, and we observed this to be true for several iterations beyond 40. Again, we see that discrepancy and GCV used with projection plus Rust's TSVD is effective, whereas they are ineffective when used on the original problem; we observed this behavior well beyond 40 iterations (see Figure 4.4).

5. Conclusions. In this work we have given a common framework for methods based on regularizing a projected problem. We have shown that determining regularization parameters based on the final projected problem rather than on the original discretization has firmer mathematical justification and often involves less computational expense. We presented results that in fact the regularized solution obtained by backprojecting the TSVD or Tikhonov solution to the projected problem is almost equivalent to applying TSVD or Tikhonov to the original problem, where “almost” depends on the size of k . The examples indicate the practicality of the method and illustrate that our regularized solutions are usually as good as those computed using the original system, and they can be computed in a fraction of the time, using a fraction of the storage. We note that similar approaches are valid using other Krylov subspace methods for computing the projected problem.

In this work, we did not address potential problems from loss of orthogonality as the iterations progress. In this discussion, we did, however, assume that either k was naturally very small compared to n or that preconditioning had been applied to enforce this condition. Possibly for this reason, we found that for modest k , round-off did not appear to degrade either the LSQR estimates of the residual and solution norms or the computed regularized solution in the following sense: the regularization parameters chosen via the projection-regularization and the corresponding regularized solutions were comparable to those chosen and generated for the original discretized problem. Another possible reason for the success of our approach is that we chose parameters for the projected problem directly, rather than for the backprojected, larger problem. In our experiments, we found that the parameters selected usually leveled out after a few iterations. The stagnation of the parameters themselves may suggest when k is large enough.

For the Tikhonov approach in this paper, we have assumed that the regularization operator L was the identity or was related to the preconditioning operator; this allowed us to efficiently compute $\|r_\lambda^{(k)}\|$ and $\|x_\lambda^{(k)}\|$ for multiple values of λ efficiently for each k . If L is not the identity but is invertible, we can first implicitly transform the problem to “standard form” [19]. With $\bar{A} = AL^{-1}$, $\bar{x} = Lx$, we can solve the equivalent system

$$\min_{\bar{x}} = \|\bar{A}\bar{x} - b\|_2^2 + \lambda^2 \|\bar{x}\|_2^2.$$

Then the projection plus regularization schemes may be applied to this transformed problem. Clearly the projection based schemes will be useful as long as solving systems involving L can be done efficiently.

REFERENCES

- [1] A. BJÖRCK, *A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations*, BIT, 28 (1988), pp. 659–670.
- [2] A. BJÖRCK, E. GRIMME, AND P. V. DOOREN, *An implicit shift bidiagonalization algorithm for ill-posed systems*, BIT, 34 (1994), pp. 510–534.
- [3] P. BLOMGREN AND T. F. CHAN, *Modular Solvers for Constrained Image Restoration Problems*, Tech. Report, Mathematics Department, UCLA, Los Angeles, 1999.
- [4] D. CALVETTI, G. GOLUB, AND L. REICHEL, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT, 39 (1999), pp. 603–619.
- [5] T. CHAN AND M. NG, *Galerkin projection method for solving multiple linear systems*, SIAM J. Sci. Comput., 21 (1999), pp. 836–850.
- [6] J. CULLUM AND R. A. WILLOUGHBY, *Lanczos and the computation in specified intervals of the spectrum of large, sparse real symmetric matrices*, in Sparse Matrix Proceedings 1978, I. S. Duff and G. W. Stewart, eds., SIAM, Philadelphia, 1979, pp. 220–255.
- [7] L. DESBAT AND D. GIRARD, *The “minimum reconstruction error” choice of regularization parameters: Some more efficient methods and their application to deconvolution problems*, SIAM J. Sci. Comput., 16 (1995), pp. 1387–1403.
- [8] H. W. ENGL AND W. GREVER, *Using the L-curve for determining optimal regularization parameters*, Numer. Math., 69 (1994), pp. 25–31.
- [9] H. E. FLEMING, *Equivalence of regularization and truncated iteration in the solution of ill-posed image reconstruction problems*, Linear Algebra Appl., 130 (1990), pp. 133–150.
- [10] A. FROMMER AND P. MAASS, *Fast CG-based methods for Tikhonov-Phillips regularization*, SIAM J. Sci. Comput., 20 (1999), pp. 1831–1850.
- [11] G. GOLUB, M. HEATH, AND G. WAHBA, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223.
- [12] G. GOLUB AND W. KAHAN, *Calculating the singular values and pseudo-inverse of a matrix*, J. Soc. Indust. Appl. Math. Ser. B Numer. Anal., 2 (1965), pp. 205–224.
- [13] W. GROETSCH, *Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, Boston, 1984.

- [14] M. GU AND S. EISENSTAT, *A Stable and Fast Algorithm for Updating the Singular Value Decomposition*, Tech. Report RR-939, Department of Computer Science, Yale University, New Haven, 1993.
- [15] M. HANKE AND P. C. HANSEN, *Regularization methods for large-scale problems*, *Surveys Math. Indust.*, 3 (1993), pp. 253–315.
- [16] M. HANKE, J. NAGY, AND R. PLEMMONS, *Preconditioned iterative regularization for ill-posed problems*, in *Numerical Linear Algebra and Scientific Computing*, L. Reichel, A. Ruttan, and R. S. Varga, eds. 1993, pp. 141–163.
- [17] P. C. HANSEN, *Analysis of discrete ill-posed problems by means of the L-curve*, *SIAM Rev.*, 34 (1992), pp. 561–580.
- [18] P. C. HANSEN, *Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems*, *Numer. Algorithms*, 6 (1994), pp. 1–35.
- [19] P. C. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problems. Numerical Aspects of Linear Inversion*, SIAM Monogr. Math. Model Comput., SIAM, Philadelphia, 1998.
- [20] P. C. HANSEN AND D. P. O'LEARY, *The use of the L-curve in the regularization of discrete ill-posed problems*, *SIAM J. Sci. Comput.*, 14 (1993), pp. 1487–1503.
- [21] B. HOFMANN, *Regularization for Applied Inverse and Ill-Posed Problems*, Teubner-Texte Mathe. 85, Teubner, Leipzig, 1986.
- [22] L. KAUFMAN AND A. NEUMAIER, *Regularization of ill-posed problems by envelope guided conjugate gradients*, *J. Comput. Graph. Statist.*, 6 (1997), pp. 451–463.
- [23] M. KILMER, *Regularization of ill-posed problems using (symmetric) Cauchy-like preconditioners*, in *Proceedings of the SPIE Annual Meeting, Advanced Signal Processing Algorithms, Architectures, and Implementations VIII*, 1998, SPIE, San Diego, CA, pp. 381–392.
- [24] M. KILMER AND D. P. O'LEARY, *Pivoted Cauchy-like preconditioners for regularized solution of ill-posed problems*, *SIAM J. Sci. Stat. Comput.*, 21 (1999), pp. 88–110.
- [25] C. L. LAWSON AND R. J. HANSON, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.
- [26] V. A. MOROZOV, *On the solution of functional equations by the method of regularization*, *Soviet Math. Dokl.*, 7 (1966), pp. 414–417.
- [27] J. NAGY, R. PLEMMONS, AND T. TORGERSEN, *Iterative image restoration using approximate inverse preconditioning*, *IEEE Trans. Image Process.*, 5 (96), pp. 1151–1163.
- [28] D. P. O'LEARY AND J. A. SIMMONS, *A bidiagonalization-regularization procedure for large scale discretization of ill-posed problems*, *SIAM J. Sci. Statist. Comput.*, 2 (1981), pp. 474–489.
- [29] C. C. PAIGE AND M. A. SAUNDERS, *Algorithm 583, LSQR: Sparse linear equations and least squares problems*, *ACM Trans. Math. Software*, 8 (1982), pp. 195–209.
- [30] C. C. PAIGE AND M. A. SAUNDERS, *LSQR: An algorithm for sparse linear equations and sparse least squares*, *ACM Trans. Math. Software*, 8 (1982), pp. 43–71.
- [31] B. N. PARLETT, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
- [32] B. N. PARLETT AND D. S. SCOTT, *The Lanczos algorithm with selective orthogonalization*, *Math. Comp.*, 33 (1979), pp. 217–238.
- [33] B. W. RUST, *Truncating the Singular Value Decomposition for Ill-Posed Problems*, Tech. Report NISTIR 6131, Mathematical and Computational Sciences Division, National Institute of Standards and Technology, Gaithersburg, MD, 1998.
- [34] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, Boston, 1996.
- [35] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, *SIAM J. Sci. Statist. Comput.*, 7 (1986), pp. 856–869.
- [36] H. D. SIMON, *Analysis of the symmetric Lanczos algorithm with reorthogonalization methods*, *Linear Algebra Appl.*, 61 (1984), pp. 101–131.
- [37] J. M. VARAH, *Pitfalls in the numerical solution of linear ill-posed problems*, *SIAM J. Sci. Statist. Comput.*, 4 (1983), pp. 164–176.
- [38] 1993, pp. 141–163. C. R. VOGEL, *Non-convergence of the L-curve regularization parameter selection method*, *Inverse Problems*, 12 (1996), pp. 535–547.

A Unified Discipline of Subsurface Sensing and Imaging Systems

M. B. Silevitch, S. W. McKnight, and C. Rappaport
Center for Subsurface Sensing and Imaging Systems (CenSSIS)
Northeastern University, Boston, MA 02115

Abstract

Subsurface sensing and imaging seeks to locate and identify objects or conditions underneath an obscuring media by monitoring a probe or wave outside the surface. Many of the mathematical and physical models used in this process are common to underground and underwater environmental exploration, medical imaging, and three-dimensional microscopies, allowing a common framework of physic-based signal processing (PBSP) to be applied. The basis for a unified discipline of subsurface sensing and imaging can be identified from a few general subsurface *information extraction strategies*. These strategies and their related families of PBSP algorithms can be used to guide a systems-oriented approach to subsurface solutions.

1. The Need for a Unified Discipline

The problem of imaging under a surface arises in a wide variety of contexts, and these problems are among the most difficult and intractable system challenges known. Place one hundred plastic landmines on top of a cleared field and they can be safely removed in hours by trained workers. Bury them under one centimeter of soil, and you have a problem that has been the subject of intensive research for over half a century and remains far from solved. State-of-the-art inductive sensors in the hands of an experienced operator can detect non-metallic mines from the signal received from the firing pin and other small metal parts. In typical operation, however, over 300

false alarms are recorded for every mine identified, each requiring lengthy and delicate examination. In the end, operational mine detection systems have little, if any, advantage over probing each square centimeter of the ground with a titanium rod, a process that can clear a field at a rate of 1 meter by 25 meters of ground per person per day. No one has any idea how the three million landmines buried in Bosnia or the 10 million in Cambodia can be removed at any reasonable cost.¹

De-mining, in common with nearly all subsurface sensing and imaging problems, is an *information* problem. If we could see precisely where the mines were buried, world-wide humanitarian de-mining could be achieved with available physical resources. Yet in an Information Age, when the cost of computation and communications is reduced by a sizable fraction each year, the full potential of applying our exponentially expanding information technology sector to subsurface problems has not been realized because of lack of equivalent progress in subsurface detection and identification.

In addition to the technical problems of probes and processing that we will discuss below, we identify two major systems obstacles to progress in subsurface sensing and imaging:

- 1) the problems of sensor design, modeling, image processing, and recognition have been compartmentalized, viewed as separate disciplines rather than as integrated parts of a *system* optimization problem.

- 2) the subsurface problems in different media and different length scales are commonly viewed as unrelated problems and addressed with *ad hoc* solutions. Lessons learned in one subsurface technology are rarely applied to other problems, and no overarching theory exists to identify fundamental limitations, predict what can be detected and the optimal way to do it.

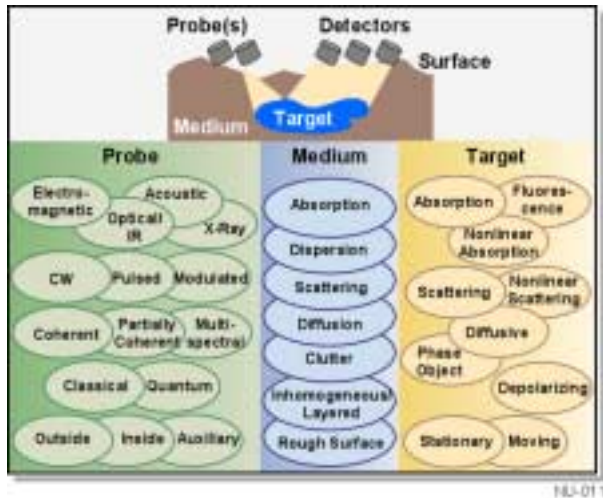


Figure 1. Subsurface problems can be classified on the basis of the probe-medium-target interactions.

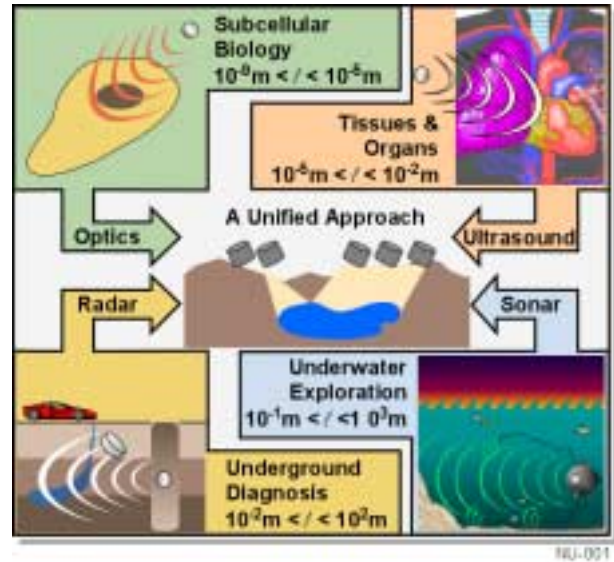


Figure 2. A unified physics-based approach can unite subsurface problems from many domains at different length scales.

The subsurface sensing and imaging problem is to extract information about a subsurface target from scattered and distorted waves received above the surface. Imaging techniques, whether ultrasound sensors in tissue or electromagnetic probes in soil, can be described by the properties of probe wave, the wave propagation characteristics of the medium and surface, and the nature of target/probe interaction as shown in Figure 1. The framework of Figure 1 describes not only underground imaging, but also underwater imaging, medical imaging inside the body, and 3D biological microscopies inside a cell or collection of cells. A unified theory of subsurface sensing and imaging, as illustrated in Figure 2, should encompass all of these applications and permit progress in one domain to be transparently applied in other domains with similar elements in the taxonomy of Figure 1.

For example, diffusive wave optical imaging for medical diagnosis and crosswell electromagnetic induction tomography for geophysical exploration both involve extracting an image of, or

information about, anomalous regions (e.g., diabetic lesions under the skin or oil-bearing rock formations under the ground; see Figure 3). Although the problems occur on vastly different length scales, both require solution of the frequency-domain diffusion equation in the presence of an inhomogeneous, layered medium, and a need to filter large data sets from multiple transmitters and receivers that are, nevertheless, sparse compared to the information set sought. Attacking these two problems within the same framework allows the synergy of the two solutions to be exploited. Thus, even the critical differences between the two problems (lossy vs. lossless propagation, Poisson vs. Gaussian noise statistics, the diffusion equation as a limit of the radiative transfer equation vs. the diffusion equation derived by neglecting the displacement current in Maxwell's Equations) become a basis for more complete understanding of the unified problem, rather than just an obstacle to applying the same specialized algorithm to each problem.

2. Physics-Based Signal Processing

It is rarely the case that we cannot get any information from the subsurface region. The concealing media, while not transparent, can usually be penetrated to a considerable depth by a

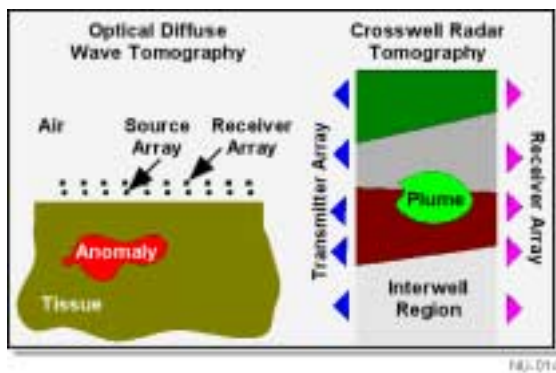


Figure 3. The physical/mathematical framework of diverse problems can be very similar.

variety of acoustic and electromagnetic wave probes. The problem is that the target signal is distorted by complex absorption, dispersion, diffraction, and refraction of the wave through the media and obscured by surface reflection, subsurface clutter, and scattered energy from unknown inhomogenities on many scales. The signal received, y , depends on the target

information x and various signal-dependent clutter and nuisance parameters z through the function C which describes the physics of the probe-wave generation, propagation, and target and clutter interaction:

$$y = C(x, z) + n \quad (1)$$

The inverse problem of un-encoding the signature of the target object x from the received signal y in the presence of unpredictable clutter signals due to z and noise, n , is the challenge of subsurface sensing and imaging.

Since the mapping from the target to the sensor depends on unknown information about the subsurface media and target, the inversion from the scattered wave to the target properties is a nonlinear mathematical problem^{2,3}. The use of appropriate physical models of the probe/surface/media/target/ receiver interaction (C in Equation 1) to assist in the solution of that inverse problem is what is referred to as physics-based signal processing (PBSP). PBSP has been identified in a seminal 1998 review article as a key to progress in image formation in complex media.⁴. Physics-based reasoning through the entire image understanding process and goal-directed processing is the most direct path to produce algorithms which are robust to modeling errors and generate accurate reconstructions of the critical information.

The fundamental problem of subsurface sensing and imaging is to differentiate the target of interest from the background scattered return and irrelevant clutter: to distinguish a landmine from the ground-surface and volume-inhomogeneity reflections, roots, stones, or shell-casings, for example. In the pulse-reflection ground-penetrating radar (GPR) simulation in Figure 4 the signal from the plastic cylinder in the lower figure is obscured by the rough-surface reflection in the upper

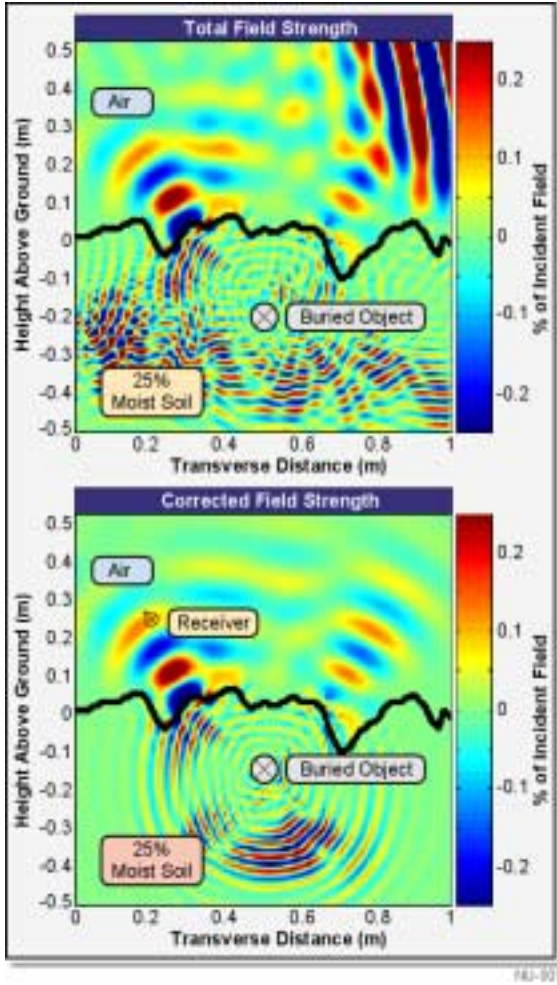


Figure 4. Clutter from rough-surface reflection in the top frame obscures the signature of the buried object in the bottom frame in this pulse GPR simulation.

figure. The task is to extract the signal from the complex scattered field of random surface irregularities. In principle, if the surface profile and the soil dielectric properties were precisely known, one could subtract the calculated background from the received signal to extract the target signature, but a full 3D calculation of the scattered field for a single pulse could take on the order of 10 hours on a 450 MHz desktop Pentium computer.

One of the primary methods of differentiating a target from clutter is through resolution of the target *shape*. Problems where the target distinguishing features are comparable to the clutter size, such as demining, are among the

most challenging subsurface problems that exist. Since resolution in the far-field is limited to order of the wavelength, it is desirable to use probes with wavelengths smaller than the size of identifying features. Unfortunately in most subsurface modalities, absorption increases with increasing frequency (smaller wavelengths). For example, the attenuation of medical ultrasound increases at the rate of 1dB/cm for every megahertz of frequency. Arterial plaque, which can be resolved by inserting catheters containing 30 MHz ($\lambda = 50 \mu\text{m}$) ultrasonic probes, cannot be noninvasively

imaged from outside the body because the attenuation is too severe (about 30 dB per centimeter of depth at 30 MHz). This range/resolution trade-off is a fundamental limitation on many subsurface modalities including underground seismic imaging and underwater sonar imaging.

Alternatively, probes which are sensitive to target *material properties*, such as to material optical response (color), conductivity, permittivity, or magnetic susceptibility, can offer advantages for target differentiation. For example, medical imaging probes such as magnetic resonance imaging or nuclear medicine molecular tags which are sensitive to target chemistry can be used to differentiate targets on the basis of *physiology* (functioning) instead of *anatomy* (structure). Imaging the subtle physiological differences between cancerous cells and normal cells would be a medical breakthrough.

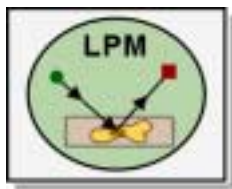
Nonlinear material properties are used for subsurface discrimination in two-photon microscopy^{5,6,7} or ultrasonic harmonic imaging. Harmonic imaging can yield diagnostically useful information on the 25% to 30% of the population that cannot be imaged well by ultrasound due to high clutter levels, distortions, and artifacts. Although harmonic imaging is already commercially available, the physical mechanisms behind it are poorly understood.

Quantitative subsurface imaging, for example measuring low concentrations of contaminants in ground water or determining oxygenation levels in a neo-natal brain, demands a challenging combination of spatial and material differentiation. To image an intensive quantity, such as percent of contaminants, requires 1) a sensing strategy with sufficient material sensitivity to differentiate to the desired measurement precision, and 2) the ability to localize the anomaly to a meaningful region. In this case the target/background differentiation includes the necessity to distinguish, for example,

a ground-water plume with 5% contaminants from an uncontaminated background and also, if a measurement precision of $\pm 5\%$ is desired, from a plume with 10% contaminants. The ability to differentiate such subtle differences in the presence of unknown subsurface inhomogeneity tests the limits of subsurface imaging.

3. Information Extraction Strategies

Despite the bewildering variety of imaging modalities and techniques covered in the Figure 1 taxonomy, subsurface problems can be organized into a relatively small number of *information extraction strategies* which use similar algorithmic tools. Three broad information extraction strategies are discussed here.



Localized probing and mosaicking (LPM) concentrates the probe wave on a local subsurface region by focusing or time-gating and then assembles these individual pieces of information into an information mosaic. Common to these techniques are problems of concentration, aberration, and registration which may use tools as simple as a lens or as complex as three-dimensional image matching and reconstruction. For example, medical reflection ultrasound and confocal microscopy both collect scattered energy from a subsurface target voxel. In both cases, precise focusing assumes a uniform homogeneous wave velocity, rarely the case in subsurface imaging, and resulting aberrations impede accurate imaging. LPM techniques are subject to obstruction by opaque objects (e.g., bones), and

because reflection geometries are sensitive to high spatial frequencies (interfaces) LPM techniques are poor at detecting low-contrast or phase-only objects. One approach to increasing LPM resolution is to produce a localized modification of the object under study, *e.g.*, by introducing a controlled magnetic field gradient in magnetic resonance imaging (MRI) or through radioactive molecular tags in nuclear medicine.

Wide-scan, high resolution LPM imaging usually requires mosaicking of multiple frames. Figure 5 shows an image of an ocean-floor hydrothermal vent assembled in this way by our Woods Hole Oceanographic Institution collaborators. Errors in image registration and

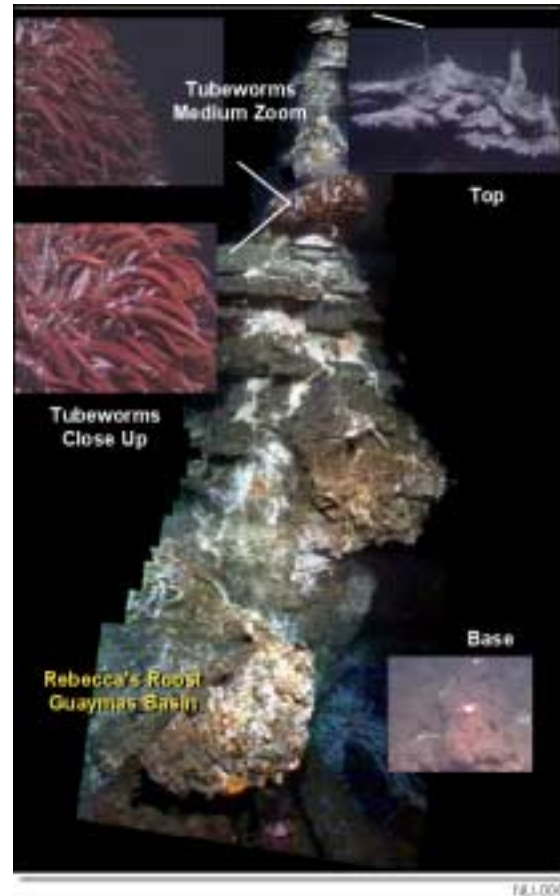
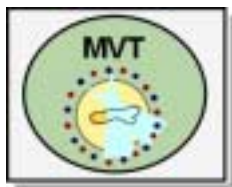


Figure 5. High-resolution underwater optical image of an ocean-floor thermal vent is assembled by 3D photo-mosaicking techniques. (Photo courtesy of Wood Hole Oceanographic Institution.)

composition techniques contribute to errors on the scale of meters in the large-scale representation of imaged objects, reducing their utility in quantitative oceanography. The use of high-resolution sonar maps to register the optical images is a multi-modality path to the desired capability for high-resolution mapping of hundreds of thousands of square meters with an accuracy of centimeters.



In contrast to LPM where the sensor information is spatially isolated, in *multi-view tomography* (MVT) correlated information from multiple sensors is combined mathematically to create a virtual map of the physical properties of

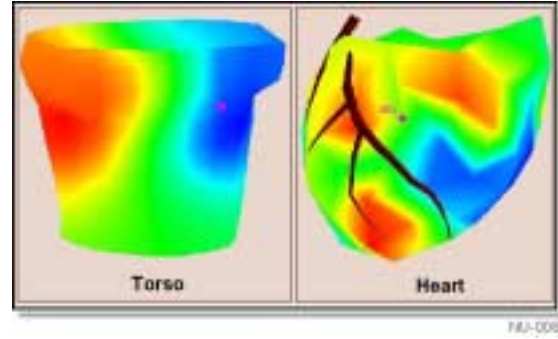
the target. These systems all involve mathematical inversion of integral equations through linearization, regularization, and integral transforms. Examples include x-ray CAT scanning, diffraction tomography, and synthetic aperture radar.

If multiple view angles are possible, MVT techniques can image obstructed/occluded objects and yield quantitative maps of wave velocity as well as absorption, allowing imaging of phase-only objects. For wavelengths that are short compared to feature dimensions, as in CAT scans^{8,9}, Radon convolution-backprojection algorithms combined with Fast Fourier Transforms (FFT) was first used to achieve 3D feature imaging in the 1970's.

Diffraction tomography¹⁰ is the technique of image reconstruction and resolution enhancement by multiple-view imaging when the wavelength is comparable to feature size. The development of the theory of diffraction tomography by the linearization and Fast Fourier Transform (FFT) inversion of the wave diffraction equations using the filtered back-propagation algorithm was pioneered in the early 1980s.^{11,12} Diffraction tomography has been successfully applied for seismic imaging of near-surface objects, including fossil dinosaur bones¹³ and to ultrasonic imaging^{14,15,16,17}, Applications of diffraction tomography with limited or obstructed field-of-view or with higher-order, non-linear models are at the forefront of the state-of-the-art.

When the wavelength is long compared to feature size, *near-field* tomographic techniques can still yield useful information in geophysical, industrial, or medical applications. In Electrical Resistance Tomography and Electromagnetic Induction Tomography, quasi-static probes and models are used to image contaminants in soil and groundwater and leaks from storage tanks on scales much smaller than the electromagnetic wavelength.^{18,19} Capacitance Tomography is used to image the permittivity

of multiphase pipeline flow components for industrial process control.²⁰ Applications in medical imaging include Electrical Impedance Tomography²¹ and Cardiac Electrical Imaging^{22,23,24}. Figure 6 shows the electric potential on the



heart imaged from the measured potential on the torso by Cardiac Electrical Imaging. The potential

Figure 6. In Cardiac Electrical Imaging (CEI), near-field MVT inversion yields the electric potential on the heart from measured voltages on the torso.

benefits of the enhanced information gained by this technique over standard electrocardiograms (ECGs) are enormous. ECGs have a rate of false diagnosis of myocardial infarctions (Aheart attacks®) as high as 30% which results in unnecessary health-care costs in the U.S. estimated at \$4 billion per year²⁵, while up to 25% of actual heart attacks go unnoticed until evidence of cardiac damage is detected in annual checkups.



Multi-spectral discrimination (MSD) adds the element of frequency discrimination to the spatial resolution sought by LPM and MVT giving a 4-dimensional map (3-space plus frequency) of the object. Combinations of MSD with LPM are common (a color photograph or hyperspectral image are examples). Joint methods for MSD and MVT have received little attention. MSD information extraction methods focus on material dispersion, parameter estimation, image registration, and fusion. Multi-sensor fusion can be viewed as an MSD problem involving, in some cases, probes that differ in modality (acoustic and optical, for example) as well as frequency.

For example, the work^{26,27} illustrated in Figure 7 shows that subtractive imaging at two nearby optical wavelengths can map specific chemical concentrations, such as oxygenated /deoxygenated

hemoglobin (Hb).^{28,29,30} This use of optical spectroscopy to detect chemical indicators of physiological function *in vivo* is promising for diagnostic discrimination. The rich spectral interaction of IR-VIS-UV light with biological molecules, however, causes absorption and strong scattering in tissue^{31,32,33} and makes the localization of emergent light difficult. Light will propagate through a hand, but the strong scattering does not allow imaging of the bone structure. Diffusive Wave Imaging^{34,35} in strongly scattering media is the focus of much current research, including optical coherence tomography^{36,37,38} and *dual-wave* acousto-phonic imaging³⁹ which seek to improve spatial resolution from centimeters to millimeters for precise quantitative diagnosis.

Satellite hyperspectral imaging of the Caribbean Basin has been used to determine the health of coral reef ecosystems and measure coastal erosion⁴⁰. Reflected light is strongly scattered in the water column, by the ocean surface, and by the atmosphere, distorting the spectral information as well as the position of underwater objects. This problem is similar to medical diffusive imaging except on a length scale that differs by orders of magnitude.

The physics of both are modeled by the radiative transfer equation (RTE).⁴¹ However, in diffusive medical imaging, the ratio of absorbed to scattered light is assumed to be small leading to the diffusion equation, while in ocean scattering Beer's law is often applied by assuming the ratio is large. In reality, the physical situation in both cases may be intermediate, and there is a need for more rigorous forward models and more robust

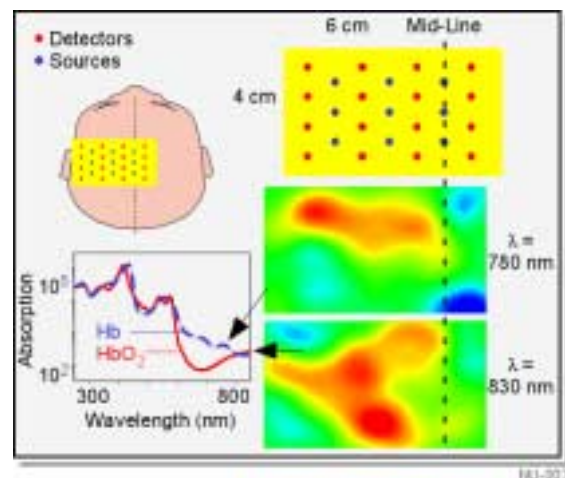


Figure 7 MSD analysis of diffusive optical waves images areas of activity (high blood oxygenation levels) in infant brain.

inversion algorithms. Current spatial resolution from space-based platforms is approximately 1 meter; processing techniques that take advantage of accurate physical models may improve the resolution limit to 10 centimeters.

4. A Systems Approach to Subsurface Sensing and Imaging

Progress in subsurface sensing and imaging approaches within these information extraction strategies has been documented in the feature articles in a recent issue of Science ("Imaging: New Eyes on Hidden Worlds")⁴². Key elements in these advances include the increase in computation power, the application of new mathematical algorithms and advanced sensing strategies, the exploitation of wave coherence, and the fusion of multiple sensing modalities (e.g., microwave and infrared) to extract increasingly detailed information from physical systems.

Still, the need for new technologies is clear. The General Accounting Office has stated, "the dimensions and potential costs of cleaning up our environment are so great that, without innovative technologies, we may find the solution cost prohibitive and impacting on our ability to address other national needs." By using current technologies, the costs of remediating Superfund and Resource Conservation and Recovery Act sites, Federal facilities, and other known hazardous waste sites may total \$750 billion over the next 30 years⁴³. Humanitarian de-mining remains an unsolved problem. No current imaging technique can adequately detect precancerous cell masses in soft tissue or

noninvasively diagnose arteriosclerosis, and there is no way to collect and correlate the images from different modalities to automatically identify incipient health problems.

The path to the solution of these important societal problems lies through the application of the unified framework of subsurface sensing and

imaging to create “end-to-end” systems solutions as shown in Figure 8. A general information

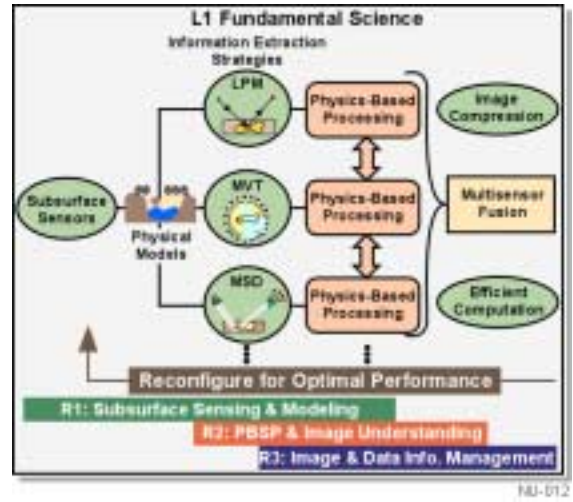


Figure 8. Application of a unified framework of subsurface sensing and imaging will allow the creation of “end-to-end” optimized system solutions.

extraction strategy is based on the common model in Equation 1. This model encompasses all traditional tomographic (MVT) problems: electrical impedance tomography, diffraction tomography, diffuse wave imaging, and the deconvolution and focusing problems associated with LPM and MSD methods such as optical microscopy systems and hyperspectral sensors.

We observe that the number of information extraction strategies is small compared to the number of individual modalities, and there is sufficient commonality in the underlying physical and mathematical structures to develop a unified framework for PBSP techniques. This framework incorporates the relevant physics of wave propagation plus techniques for capturing the effects of random fluctuations in background properties, unknown interfaces, and discrete background clutter.

The end-to-end integration of subsurface sensing mechanisms with advanced algorithms fostered by a unified approach enhances subsurface sensing systems in two ways. First, the physics-based processing and recognition/decision algorithms can exploit the *a priori* knowledge implicit in a deep

understanding of the physics of the wave probe and sensors. Second, the approach to problems from the front-end to final product permits optimization of the entire system--all the way to the sensor modality and configuration--based on the ultimate system objective.

5. Research Needs and Barriers to Progress

Fundamental research on several overlapping levels is necessary to create the desired system solutions of Figure 8. Subsurface Sensing and Modeling research (R1) seeks to elucidate the nature of the physical model contained in C in Equation 1. This includes investigations of fundamental wave interactions in the subsurface media and targets, particularly for non-linear and dual-wave (*e.g.*, acousto-photonic) probes where propagation and scattering mechanisms are poorly understood. Forward modeling plays an important role here. Given a sensor system, the computational model provides the mathematical description of the physics of the device/wave/media/target/receiver interaction problem. We can have reasonable confidence that the physics of the wave interactions is understood only when good agreement can be achieved between a physical model and the sensor data. This forward model then becomes the first step in the inversion process, and may be used repeatedly in iterative reconstructions.

PBSP and Image Understanding (R2) research uses the model C developed from R1 in PBSP algorithms to find x from y . This is the inverse problem which is the core of subsurface sensing and imaging. The difficulties with accurate reconstruction of the required information in x arise from many sources: sparse, limited data collected at surface boundaries, sensor uncertainties, clutter, and the complex nature of the operator C .

To gather enough information to identify subsurface objects in cluttered and noisy environments, large amounts of data may need to be collected from multiple diverse sensors. This can lead to severe computation and data handling tasks. Research in Image and Data Information Management (R3) is needed to develop compression and computation capabilities to link distributed sensors and to implement complex R2 algorithms in near-real-time, allowing sensor reconfiguration to optimize performance.

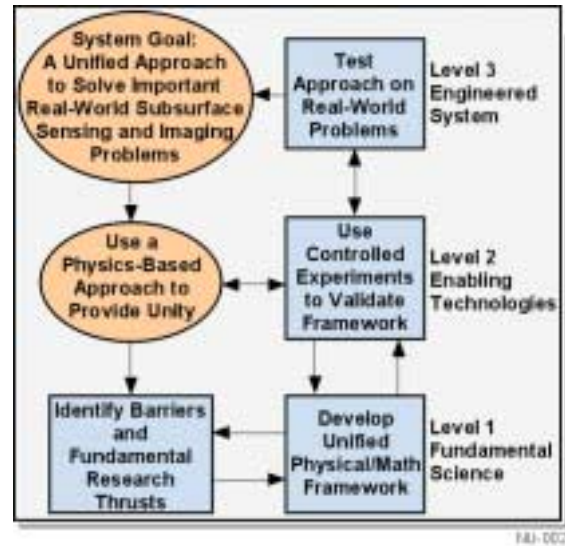


Figure 9. The unified framework enables a top-down approach to sensing and imaging systems solutions.

The path from these fundamental research thrusts to advanced sensing and imaging solutions to societal needs requires a model for goal-directed research and engineering development. In Figure 9 we present one such model based on the National Science Foundation’s Engineering Research Centers Program.⁴⁴ In a top-down focus, we begin with a broad system goal: to create a unified discipline of subsurface sensing and imaging to solve important societal problems. Applying the unified approach of Figure 8 based on the common wave-physics elements, we identify barriers to our goal and select fundamental research thrusts (R1-R3) to address these barriers. The Level 1 fundamental science develops a unified physical/mathematical framework for subsurface sensing and imaging.

At Level 2, experimental testbeds are used to validate the unified framework on well-characterized “ground-truth” scenarios. On this level, the research breakthroughs will be verified

with real data. This will require the development of widely accessible experimental testbeds with reconfigurable sensors in a number of subsurface sensing scenarios. In addition standards for data and metadata need to be developed so that algorithms can be tested in diverse domains at different length scales. As the unified framework is developed and verified it will be used to revise the barriers and research efforts. The verified framework is applied to real-world problems at Level 3, completing the system goal.

Beginning from the unified framework developed in this paper as a first step, we can identify some present barriers to advanced civil-environmental and biomedical detection systems. The major barriers lie both in unsolved fundamental research problems and in lack of adequate technology tools. These barriers are:

Barrier 1: Fundamental knowledge is lacking about nonlinear interactions, dual-wave sensing mechanisms, and coherent imaging in scattering media. While linear acoustic and electromagnetic interactions can be modeled and characterized by well-understood linear response functions, advanced imaging techniques using non-linear or dual-wave (*e.g.*, acoustic/optical) probes require fundamental investigations to determine appropriate physical models.

Barrier 2: The present formulation of coherent inverse scattering is inadequate to quantitatively image objects in highly-scattering random inhomogeneous and cluttered environments. In these situations the non-linear character of the inverse problem defeats tomographic reconstruction and adequate alternatives do not yet exist.

Barrier 3: Recognition strategies for obscured and limited-view subsurface applications are not well developed, and we have no theory for combining different sensor inputs to optimize the information obtained.

Barrier 4: Forward modeling of large complex scattering geometries is too slow for real-time inverse-processing applications. Progress is required in both efficient approximate forward solvers and in hardware/software implementation of processing.

Barrier 5: There are few test facilities with sufficient flexibility and sensor reconfigurability to permit the optimization of sensor modality/configuration and processing strategies based on recognition and decision objectives.

Barrier 6: Techniques for rapid processing, cataloging, storage and retrieval of large image databases are not sufficiently developed. Data and metadata standards will need be instituted so that processing algorithms can be routinely tested on experimental results from diverse experimental domains.

5. Conclusion

The pieces are in place for a major advance in the field of sensing and imaging. The development of a common framework and unified discipline of subsurface sensing and imaging promises to allow the field to emerge as a co-pillar of the Information Age, along with computation and communications. We can look forward to systems-level advances such as integrated, field-tested, algorithmic and computational tools for the entire range of subsurface problems, and standards and criteria for the use of multiple sensing modalities to achieve subsurface sensing system goals. These, in turn, will open the door for the next generation of systems for environmental sensing underground

or under the water, medical imaging and automatic diagnosis inside the body, and biological microscopy to reveal fundamental processes inside living cells.

6. Acknowledgements

The Center for Subsurface Sensing and Imaging (CenSSIS), headquartered at Northeastern University, is a partnership of four universities (Northeastern University, Boston University, Rensselaer Polytechnic Institute, and the University of Puerto Rico at Mayaguez), four affiliated institutions (Lawrence Livermore National Laboratory, Massachusetts General Hospital, Brigham and Women's Hospital, and Woods Hole Oceanographic Institution), and over 30 industry sponsors. The mission of CenSSIS is to create a unified discipline of subsurface sensing and imaging which will enable revolutionary systems solutions for conserving natural resources and providing comprehensive, affordable health care. The CenSSIS logo in Figure 10 represents the model of “diverse problems, similar solutions” which is at the core of a unified discipline. The CenSSIS logo has been adapted for the cover of this initial issue of the Journal of Subsurface Sensing and Imaging.

We would like to acknowledge the invaluable assistance of our many CenSSIS colleagues in developing the concepts presented in this paper. In particular, we would like to thank Bahaa Saleh for the

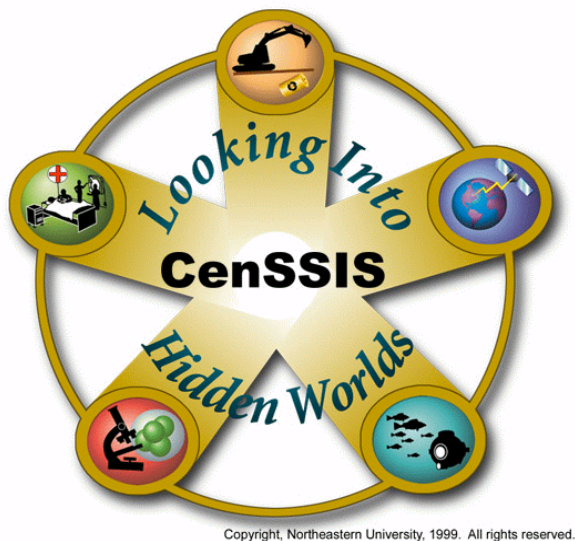


Figure 10. The CenSSIS logo expresses the goal of a unified discipline of subsurface sensing and imaging: “diverse problems, similar solutions.”

concept of Figure 1, Eric Miller for Figure 3 and the related discussion, Hanu Singh and Jon Howland of WHOI for Figure 5, Charles DiMarzio and David Boas for Figure 7, and Dana Brooks for Figure 6. We are especially in the debt of David Castanon for his elucidation of the concept of information extraction strategies. Valuable communications with Ron Roy, Robin Cleveland, and Tom Szabo are gratefully acknowledged.

The research leading to Figure 4 was supported by the Army Research Office Multidisciplinary University Research Initiative (MURI) on Humanitarian Demining under contract number DAAG55-97-1-0013.

1. G. Strada, "Horror of Landmines," *Scientific American*, May 1996, p. 40.
2. D. Colton and R. Kress, "Inverse Acoustic and Electromagnetic Scattering Theory," *Applied Mathematical Science Series, Vol. 93* (Springer-Verlag) 1992.
3. D. Colton and P. Monk, "The Detection and Monitoring of Leukemia Using Electromagnetic Waves: Mathematical Theory," *Inverse Problems* **10**, 1235-1251 (1994).
4. J. A. O'Sullivan, R. E. Blute, D. L. Snyder, "Information Theoretic Image Formation," *IEEE Trans. Information Theory* **44**, 2094-2121 (1998).
5. W. Denk, J. H. Strickler and W. W. Webb, "Two-Photon Laser Scanning Fluorescence Microscopy," *Science* **248**, 73-76 (1990).
6. *Handbook of Biological Confocal Microscopy*, J. B. Pawley, ed. (Plenum, New York, 1995), pp. 445-458.
7. W. Denk, "Two-Photon Excitation in Functional Biological Imaging," *Journal of Biomedical Optics* **1**, 296-304 (1996).
8. G. N. Ramachandran and Lakshminarayan (1971), "Three-Dimensional Reconstructions from Radiographs and Electron Micrographs: Applications of Convolution Instead of Fourier Transforms," *Proc. Nat. Acad. Sci.*, **68**, 2636-2240.
9. L. A. Shepp and B. F. Logan (1974), "The Fourier Reconstruction of a Head Section," *IEEE Trans. Nucl. Sci.*, **NS-21**, 21-43.
10. E. Wolf, "Principles and Development of Diffraction Tomography," *Trends in Optics*, Anna Consortini, ed. (Academic Press, San Diego 1996) Chap. 5.
11. A. J. Devaney, "Filtered Back-Propagation Algorithm for Diffraction Tomography," *Ultrasonic Imag.* **4**, 336 (1982).
12. A. J. Devaney, "A Computer Simulation Study of Diffraction Tomography," *IEEE Trans. on Biomed. Eng. BME-* **30**, 377 (1983).
13. A. J. Witten and W. C. King, "Acoustic Imaging of Subsurface Features," *J. of Env. Eng.* **116**, 166 (1990).
14. N. Sponheim and I. Johansen, "Experimental Results in Ultrasonic Tomography Using a Filtered Back Projection Algorithm," *Ultrasonic Imaging* **13**, 56 (1991).
15. N. Sponheim, I. Johansen and A. J. Devaney, "Initial Testing of a Clinical Ultrasound Monograph," in *Acoustic Imaging*, ed. M. Bertero and E. R. Pike (Adam Hilger, Bristol 1991) 401.
16. K. T. Ladas and A. J. Devaney, "Application of ART Algorithm in an Experimental Study of Ultrasonic Diffraction Tomography," *Ultrasonic Imaging* **15**, 48 (1993).
17. J. J. Stamnes, L.-J. Gelius, I. Johansen and N. Sponheim, "Diffraction Tomography Applications in Seismics and Medicine," in *Inverse Problems in Scattering and Imaging*, ed. M. Bertero and E. R. Pike (Adam Hilger, Bristol, 1992) 268.
18. O. Dorn, H. Bertete-Aguirre, J. Berryman, and G. Papanicolaou, "A Nonlinear Inversion Method for 3D-Electromagnetic Imaging Using Adjoint Fields," to appear in *Inverse Problems*.
19. W. Daily and A. Ramirez, "Environmental Process Tomography in the US," *Chem. Eng. Journal* **56**, 159-165 (1995).
20. M. S. Beck, M. Byars, T. Dyakowski, R. Waterfall, R. He, S. J. Wang, and W. Q. Wang, "Principles and industrial applications of electrical capacitance tomography," *Measurement and Control* **30**, 197-200 (1997).
21. D. Isaacson and M. Cheney, "Current Problems in Impedance Imaging," in *Inverse Problems in Partial Differential Equation*, SIAM, 1990, Chapter 9, pp. 151-149.

22. Y. Rudy and B.J. Messinger-Rapport, "The Inverse Solution in Electrocardiography: Solutions in Terms of Epicardial Potentials", *CRC Crit Rev Biomed Eng* **16**, 215-268 (1988).
23. D.H. Brooks and R.S. MacLeod, "Electrical Imaging of the Heart: Electrophysical Underpinnings and Signal Processing Opportunities", *IEEE Sig. Proc. Mag.* **14**, 24-42(1997).
24. R.S. MacLeod and D.H. Brooks, "Recent Progress in Inverse Problems in Electrocardiology", *IEEE Eng. in Med. & Biol. Soc. Magazine* **17**, 73-83(1998).
25. H. P. Selker, "Coronary care unit triage decision aids: how do we know when the work?", *Am. J. Med.* **87**, 491-493(1989).
26. Siegel, A.M., Marota, J.J.A., and Boas, D.A., "Design and evaluation of a continuous-wave diffuse optical tomography system," *Optics Express* **4**:287-298 (1999).
27. Hintz S., A. M. Siegel, D. A. Benaron, and D. A. Boas, "Bedside functional imaging of the premature infant brain during passive motor activation," in *Optical Tomography and Spectroscopy of Tissue III*, Britton Chance, R. Alfano, Bruce J. Tromberg, Editors, *Proceedings of SPIE* 3597:221-229 (1999).
28. P. J. Dwyer, C. A. DiMarzio and R. R. Anderson (1997), *Imaging of Blood Oxygenation in Skin Using Four-Wavelength Reflectance Spectroscopy*, *Biomedical Sensing, Imaging, and Tracking Technologies II*, 2976, 270-280.
29. P. J. Dwyer and C. A. DiMarzio (1996), *Algorithms for Processing Spectral Reflectance Images of Tissue to Obtain Maps of Oxygen Saturation*, Paper TuAA6, presented at the OSA Annual Meeting in Rochester, NY, October 21-24.
30. Peter J. Dwyer, C. A. DiMarzio, G. Tearney and R. R. Anderson (1997), *Imaging of Blood Oxygen Saturation in Skin Using Four-Wavelength Reflectance Spectroscopy*, submitted to *IEEE Transactions on Biomedical Imaging*.
31. S. Wan, R. R. Anderson and J. A. Parrish (1991), *Analytical Modeling for the Optical Properties of the Skin with in Vitro and in Vivo Applications*, *Photochem. Photobiol.*, **34**, 493-499.
32. R. R. Anderson and J. A. Parrish (1982), *Optical Properties of Human Skin*, *The Science of Photomedicine*, Plenum, New York, 147-194.
33. M.J. C. Van Gemert, Steven L. Jacques, HJ. C. M. Sterenborg and W. M. Star (1989), *Skin Optics*, *IEEE Transactions on Biomedical Engineering*, **36**, 12, 1146.
34. AJ. Knuttel, J. Schmitt and J. Knutson (1993), *Spatial Localization of Absorbing Bodies by Interfering Discursive Photon Density Waves*, *Applied Optics*, **32**, 381.
35. L. Svaasand, R. Tromberg, T. Haskell and M. Berns (1993), *Tissue Characterization and Imaging Using Photon Density Wave*, *Optical Engineering*, **32**, 258.
36. D. Huang, et al (1991), *Optical Coherence Tomography*, *Science*, **254**, 1178-1181.
37. M. E. Brazinski, et al (1996), *Optical Coherence Tomography for Optical Biopsy*, *Circulation*, **93**, 1206-1213.
38. D. Beneron, et al (1997), *Tissue Optics*, *Science*, **276**, 2002-2003.
39. DiMarzio, Charles A., and Thomas J. Gaudette, "A New Imaging Technique Combining Diffusive Photon Density Waves and Focussed Ultrasound," *Optical Tomography and Spectroscopy of Tissue III*, *Proc. SPIE* 3597}, presented at BIOS 99 in San Jose, CA. January 1999. Paper 3597-59.
40. L. O. Jimenez and M. Velez-Reyes, "Subset Selection Analysis for the Reduction of Hyperspectral Imagery," *Proc. IEEE International Geosciences and Remote Sensing Symposium*, Seattle, WA, 1998.
41. Akira Ishimaru, *Wave Propagation and Scattering in Random Media*, IEEE Press, New York, 1997.
42. T. Appenzeller and C. Norman *Imaging: New Eyes on Hidden Worlds*, *Science*, **276**, 1981-1995 (1997).
43. Russel, M., Colglazier, E.W., and English, M.R., 1991, "Hazardous waste remediation--The task ahead" Knoxville, Waste Management Research and Education Institute, University of Tennessee..

44. National Science Foundation Engineering Research Centers Program Announcement NSF 98-146, Supplemental Full Proposal Guidelines and Format.

Laser-Induced Acoustic Imaging of Buried Objects

S. W. McKnight, C. A. DiMarzio, W. Li, and J. Stott

Center for Subsurface Sensing and Imaging Systems

Northeastern University, Boston, MA 02115

Abstract

A 100 mJ, 100ns pulsed CO₂ laser incident on the surface of soil is used as a localized acoustic source for the detection and imaging of underground objects. The acoustic pulse produced by the impulsive heating of the soil surface due to the absorbed 10.6 μm radiation is detected with an acoustic transducer suspended over the surface. Application of a Fourier domain filter enables the separation of the direct acoustic return from the faint echo from an object buried 3-25mm below the surface. Scanning of the laser pulses across the position of a buried object allows the resolution of the shape and the depth of the buried object. The application of this technique to image buried landmines is demonstrated in trials at an outdoor test track.

Keywords: Laser, acoustic, landmine, detection

The detection, location, and ultimate removal of landmines is a challenging and important task. Mines with low metal content pose severe difficulties for the standard electromagnetic induction techniques¹. Often, 300 or more false alarms are observed for every actual landmine detected. A major goal of any new mine-detection technology, then, is to reduce the false alarm

rate of existing techniques. Any technology which provides information on the shape, size, and composition of buried objects may prove useful in this regard. Acoustic detection is attractive in principle because of the large acoustic mismatch between porous soil and solid metallic or non-metallic objects, but difficult to achieve in practice because of the need to couple acoustic energy into and out of the soil^{2,3,4,5}. A laser-induced acoustic imaging system is proposed, and experiments are reported here to validate the technique.

The photo-acoustic effect has been used to produce acoustic pulses for liquid drop studies and materials characterization^{6,7,8}. A short pulse of laser light is absorbed at the surface of the solid or liquid target, resulting in immediate heating and expansion of the material. Provided that the duration of the pulse is shorter than the time required for thermal diffusion to reduce the temperature, the resultant heating will cause expansion, and thus generation of a broad-band acoustic wave. For application to mine detection, the laser pulse would be directed at the surface of the ground over the suspected location of the landmine. The laser-generated acoustic wave reflects and refracts upon encountering buried objects or interfaces. An acoustic receiver placed above the ground will detect the acoustic wave from the surface, and at a later time, will detect the reflection from a buried object, if present⁹.

To test this concept, a pulsed, TEA, carbon-dioxide laser (LSI 150, by Laser Science, Inc.) was used to induce acoustic signals in a box of sand with various buried objects.¹⁰ The laser pulse energy was over 100 mJ, with a typical gain-switched spike about 100 ns long, followed by a tail of several microseconds. The sandbox was moved as shown in the inset in Figure 1 on a grid pattern to produce a raster scan, and the echo at each location was recorded with an acoustic receiver in the air above and an oscilloscope connected to a personal computer.

Two different detectors were used in these measurements. The measurements reported in

the lab used a generic piezoelectric transducer potted in an epoxy cylinder approximately 4cm long by 2cm in diameter. The resonant response of this detector created a high gain signal sharply peaked around 30 kHz. The field trial measurement used a Radio Shack 33-1073A audio microphone with an approximately flat response below 15 kHz. The microphone had a higher sensitivity, but also was much more susceptible to low-frequency noise. The lower frequency microphone also lead to noticeably lower acoustic resolution. The detectors were positioned between 2 and 10cm above the surface of the ground, directly above the laser impact site.

Two typical time histories are shown in Figure 1, one with an ice-hockey puck, sliced in half, buried one centimeter below the surface and one with no buried object. The dominant feature is the ringing of the transducer at approximately 29 kHz, stimulated primarily by the reflection from the air-sand interface, and so the two signals look very similar. The challenge then is to extract the small echo caused by the buried object from the larger surface signal. If the soil surface is level, the laser spot and microphone will remain at the same relative positions, and subtraction of a background signal from a target-free region of ground will reveal the anomaly created by the echo from the subsurface object, albeit strongly modulated by the detector ringing. If the ground is non-uniform, amplitude and phase differences between the target and background time traces will defeat the cancellation by subtraction of the time-domain signals.

Alternatively the signals may be Fourier transformed, the detector response estimated from the no-target signal and removed by a deconvolution. Deconvolution is implemented by dividing the received signal by the no-target signal in the Fourier domain. The problem with this approach is the singularities created at frequencies where the detector response went to zero. To take out these singularities, a deconvolution based on a Weiner filter¹¹ was used.

Under the assumption that the signal consists of the impulse response of the transducer

convolved with a temporal description of the reflectivity as a function of depth, the Fourier transform of the signal, $Y(\omega)$ is the product of the transfer function $H(\omega)$ and the desired reflectivity function $X(\omega)$, plus a noise term, $N(\omega)$.

$$Y=HX+N \quad (1)$$

Straightforward deconvolution would magnify the noise at frequencies where H approaches zero. Instead, the Weiner filter produces an estimate, \hat{X} , of the signal as

$$\hat{X}=Y\frac{H^*}{HH^*+\epsilon} \quad (3)$$

where ϵ is a constant chosen empirically through trial and error. If the magnitude of the transfer function is large, the result is:

$$\hat{X}\approx\frac{Y}{H} \quad (4)$$

which implements a deconvolution, while for small magnitudes of H ,

$$\hat{X}\approx Y\frac{H^*}{\epsilon}$$

which implements a matched filter. The function H can be determined from the Fourier transform of a pulse where the laser is not near the target.

The constant parameter ϵ is chosen by examining the Fourier transform of the data and selecting a value that was equivalent to the signal level at a cut-off frequency about 50 kHz. The data above the cut-off frequency was discarded since the piezoelectric transducer response was

reduced by a factor of 40 from the peak response near 30 kHz. In practice, the selection of the high-frequency cut-off and the value of ϵ did not make a large difference in the quality of the deconvolution.

To test the ability to distinguish depth, the puck was tilted as shown in the Figure 1 inset, and the sand box was moved in the x-direction with respect to the fixed laser and detector so that the laser spot passed directly over the center of the puck. The vertical slices (constant position) through the resulting data processed through the Weiner filter, shown in Figure 2, clearly demonstrated the ability to resolve depth and to effectively filter the signal even under an uneven surface. At our measured speed of sound in dry sand of about 300 m/s, the time delays in Figure 2 correspond to a distance under the surface of between 0.3 to 1.2 cm, consistent with the tilt of the target puck.

A half-hockey puck was buried horizontally less than 1 cm below the surface and horizontal (constant time delay) slices were also developed to show the strength of the signal at each individual depth as the target was scanned in an x-y plane. Each such slice imaged the surface of the puck at a given depth. All the slices between time delays of 45-60 μ s (depth from 0.7-0.9 cm) were summed to produce the projection shown in Figure 3. The outline of the puck is shown by the dashed line to illustrate that the shape is also quite well resolved in two dimensions.

At the 29-kHz resonant frequency of the transducer the acoustic wavelength in dry sand is about 1 cm, and we have measured an pressure-field absorption length ($1/\alpha$) close to 30cm (assuming $1/r$ spreading). Experiments conducted with a microphone with a response from 0-15 kHz produced noticeably poorer resolution. At higher frequencies it is anticipated that resolution will improve and absorption will increase, with an optimal frequency at some yet-to-be-determined value.

To evaluate the usefulness of this technique for application to demining, we conducted a field test of the system at an outdoor test facility. The test track is a 25 x 3 meter section of screened loam traversed by an instrument cart on rails. The laser and electronics were mounted on the cart and pushed over the location of a known mine simulant, a 5.6 cm-diameter plastic M-14 antipersonnel mine with the charge replaced with inert silicone filler and buried 2.5 cm below the surface. The laser was focused with a lens to a spot size about 5mm in diameter and the detector, an audio microphone, was mounted on the cart about 2 cm from the ground surface near the incident laser spot. The cart was moved so the laser pulse crossed the center of the known mine location and stopped every 1.3 centimeter so the acoustic signals from eight laser pulses could be recorded and averaged.

To allow for variations in the ground surface that affected the intensity of the acoustic signal, the signals were Fourier transformed and normalized before applying the Weiner filter. The filtered signal is shown in Figure 4 as a function of laser position with respect to the known position of the mine simulant as indicated. The trace indicates a clear anomaly at a position between about 20 and 30 cm corresponding to the position of the buried mine. The return from the mine at about 0.2 ms time delay is weak, but the modification of the acoustic signal from the plastic mine and the disturbed soil with respect to the undisturbed soil creates a pronounced ringing in the filtered signal that extends for almost a millisecond after the initial pulse.

The Weiner filter was not as successful in reconstructing the impulse echo as in the laboratory test, perhaps due to the limited dynamic range of the sampling oscilloscope, and the image is larger than the diameter of the M-14 simulant, possibly as a result of the imaging of the disturbed soil caused by the simulant emplacement, or the lower resolution of the microphone frequency range. Nevertheless, the existence of an acoustic anomaly at the position of the buried mine simulant is evident in Figure 4.

Additional information could be obtained if multiple acoustic sensors are used, for example, through the use of a multiple-beam laser vibrometer, scanned with the incident laser beam. The shape information attainable with laser-induced acoustic imaging, combined with additional information, has the potential to reduce the false-alarm rate associated with landmine detection.

Acknowledgments

The authors are grateful to Robin Cleveland, Ron Roy, and Jim Sabatier for their insights into acoustic propagation in granular media, to Gerry Sauermann and Dan Hogenboom for assistance in the design and implementation of the experiments, and to Eric Miller for suggestions on the signal processing. This work was supported in part by the Army Research Office Demining MURI under Grant DAAG55-97-1-0013 and by the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821.

References

1. S. Ashley, Mechanical Engineering **118**, 62 (1996).
2. C. T. Schroder and W. R. Scott, Jr., IEEE Trans on Geoscience and Remote Sensing **Vol 38**, 1505 (2000).
3. W. R. Scott, Jr. and J. S. Martin, "Experimental investigations of the acousto-electromagnetic sensor for locating land mines," Proc. of the SPIE, Detection and Remediation Technologies for Mines and Minelike Targets IV, **Vol. 3710**, 204(April 1999).
4. J. M. Sabatier and N. Xiang, "Laser-Doppler-based acoustic-to-seismic detection of buried mines," Proc. of the SPIE, Detection and Remediation Technologies for Mines and Minelike Targets IV, **Vol. 3710**, 215(April 1999).
5. D. M. Donskoy, "Detection and discrimination of nonmetallic land mines," Proc. of the SPIE, Detection and Remediation Technologies for Mines and Minelike Targets IV, **Vol. 3710**, 239(April 1999).
6. S. M. Park, M. I. Khan, H. Z. Cheng and G. J. Diebold, Ultrasonics **29**, p. 63-67 (1991).
7. J.-P. Monchalín and R. Heon, Materials Evaluation **44**, 1231(1986).
8. C. B. Scruby and L. E. Drain, Laser-Ultrasonics: Techniques and Applications (Adam Hilger, Bristol, UK, 1990).
9. C. A. DiMarzio and S. W. McKnight, "Optical Pulse Induced Acoustic Mine Detection", U. S. Patent 6,069,843, May 30, 2000.
10. S. W. McKnight, W. Li, and C. DiMarzio, "Imaging of Buried Objects by Laser-Induced Acoustic Detection," Proc. of the SPIE, Detection and Remediation Technologies for Mines and Mine-like Objects IV, Vol. **3710**, 231 (1999).
11. N. Wiener, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Wiley, New York, 1949.

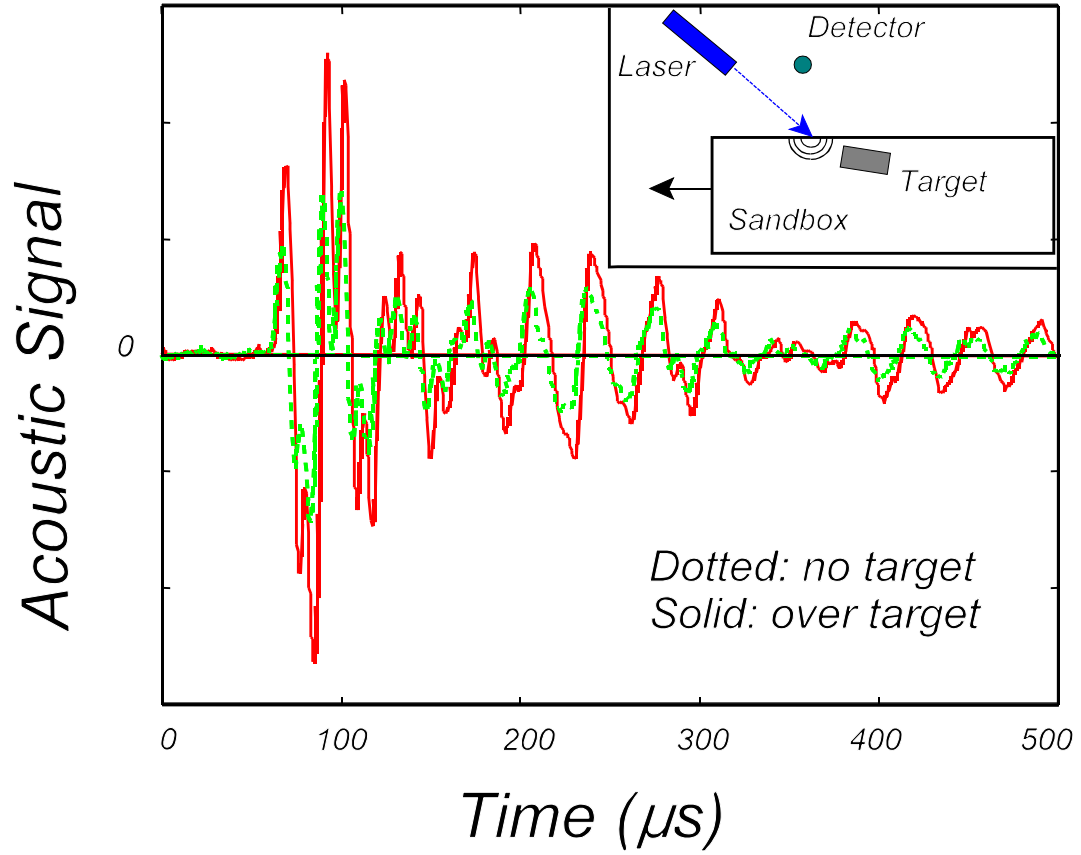


Figure 1. Unprocessed detector signal (PZT transducer voltage) for a single laser pulse incident on the surface of dry sand when the laser is over the target (solid) and when it is not over the target (dotted). Inset: Experimental configuration for laser-induced acoustic experiment.

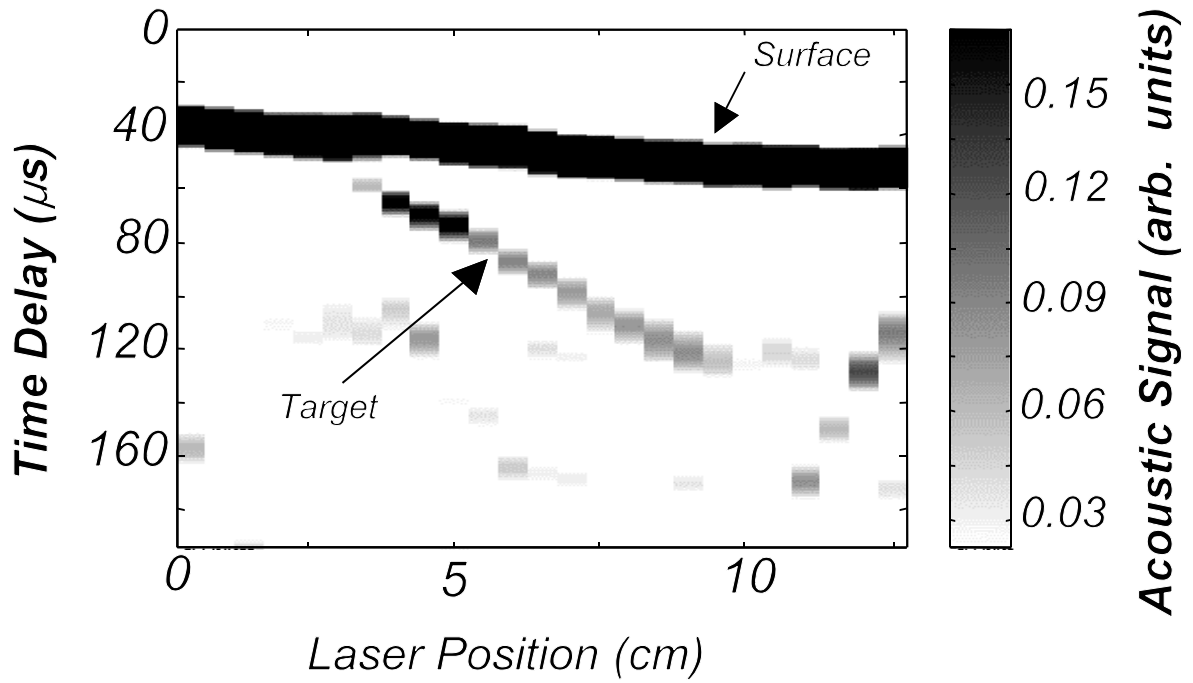


Figure 2. Fourier-filtered acoustic data for tilted hockey puck buried under uneven surface of dry sand. The high intensity return near 40 μs is the ground surface and the return from the puck surface is visible from 50-130 μs .

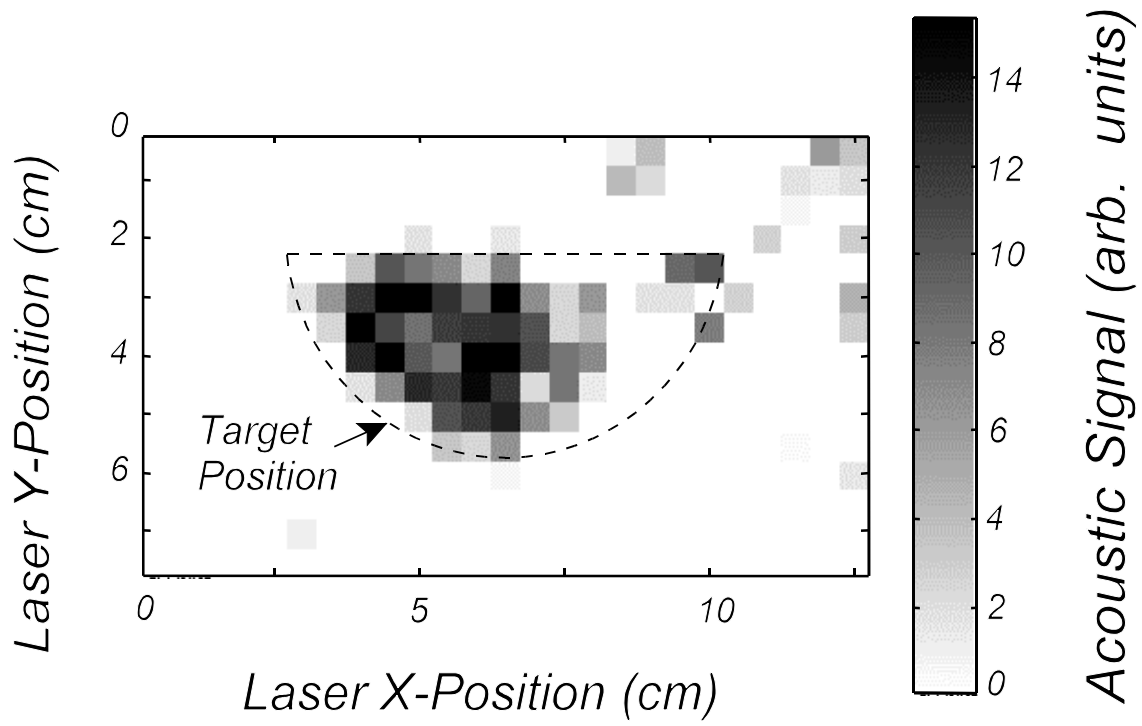


Figure 3. Sum of filtered acoustic return from $\Delta t=45$ to $60\mu\text{s}$ (depths from 0.7 to 0.9 cm) along with shape of object buried in dry sand (a sectioned hockey puck) as a function of laser pulse position, showing imaging of object shape in horizontal plane.

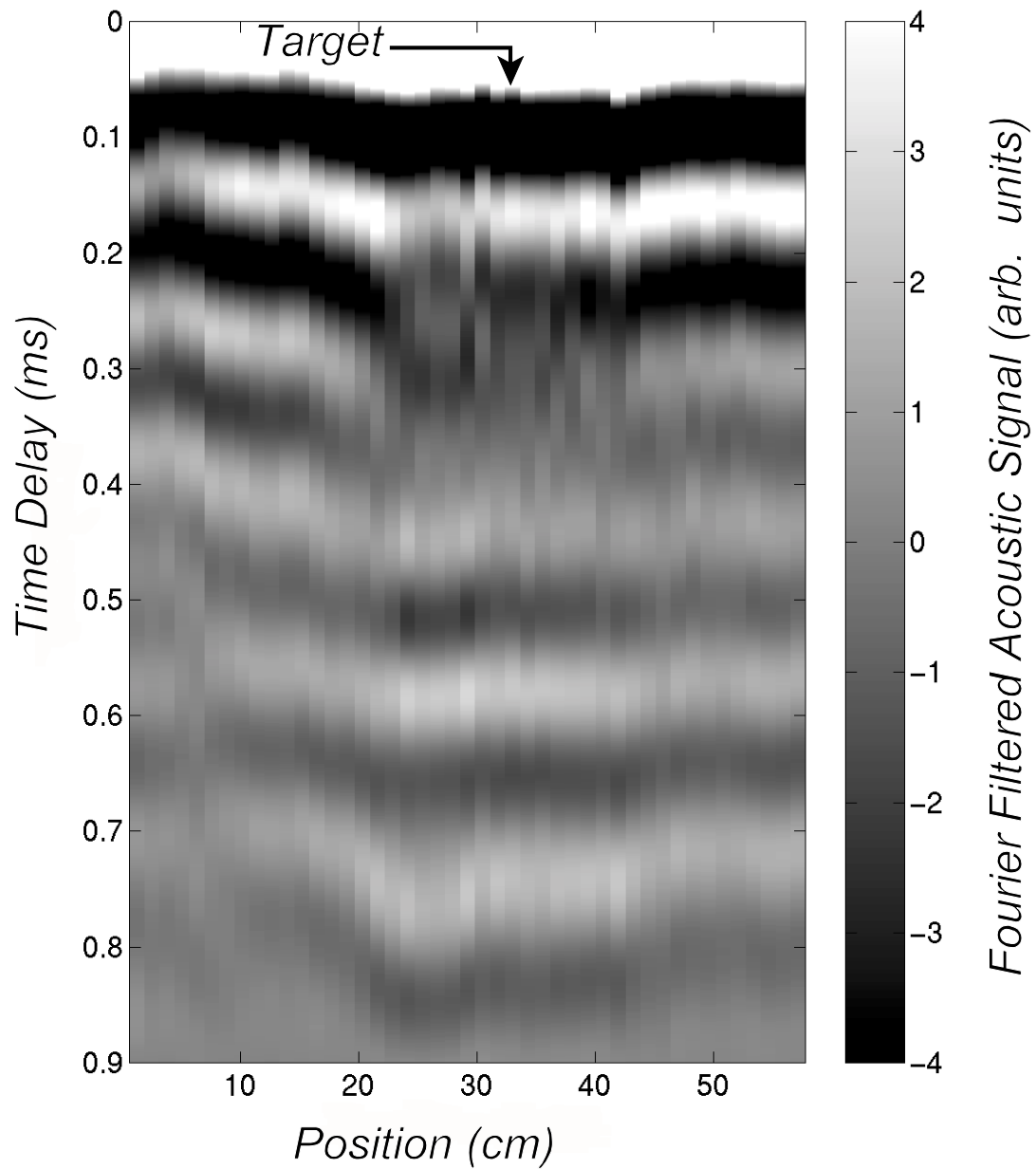


Figure 4. Fourier-filtered signal received from audio microphone as a function of the position on the ground surface of the pulsed laser source, taken at an outdoor test track. An M-14 antipersonnel landmine simulant is buried in screened loam, 2.5 cm under the surface, at the position indicated.

Frequency Domain Simulation of Focused Array Radar Returns from Buried Mines in Clutter

by

Harold Raemer, Carey Rappaport and Eric Miller

Center for Electromagnetics Research
Northeastern University, Boston, MA 02115, U.S.A

Abstract

Simulation code was developed to model monostatic or bistatic radar returns from terrain and discrete objects within the radar's field of view, including subsurface scattering due to complex permittivity inhomogeneities and buried objects, using Born approximations. This code is applied to simulation of return from a mine-like dielectric cylinder buried a few inches below the surface. Clutter sources included are: scattering from the rough surface above the mine and subsurface random permittivity inhomogeneities. Simulated images of received power from a subsurface region are obtained with and without the mine and the results will be compared with experimentally-obtained images of the region. Depth resolution of a few inches is obtained by using a focused linear array of 4 transmitters and a linear array of 4 receivers identical to and parallel to the transmitter array. The illuminated subsurface volume is between the two arrays. The detector scans the volume by varying relative delays between array elements such that, at a given time instant t_1 the signals arrive at a point p_1 and returns from p_1 arrive at the 4 receivers with equal delays. Superposition of received signals at time t_1 favors returns from a small volume around p_1 . At time t_2 all energy is similarly focussed on a different point p_2 . The process continues until the entire volume has been scanned and an image of the region has been generated. Sensitivity to mine dimensions, composition and burial depth and soil parameters is demonstrated.

1 INTRODUCTION

This paper describes a frequency domain simulation of a GPR scenario involving a bistatic focused array radar system attempting to detect and locate a mine in the presence of clutter. The clutter sources are: the rough surface at the air-soil interface above the mine and a random distribution of complex permittivity throughout the subsurface region containing the mine.

In the next section we briefly discuss the modeling of the propagation of the transmitted wave into the soil, its reflection from the interface and scattering from underground objects and regions of deviation of permittivity from its mean value.

In the section following that we describe the focused array scheme and the way in which we have attempted to simulate its application to the mine detection and location scenario of interest in this investigation.

Finally, some results of the simulation are presented in the form of 3- D plots of the amplitude of the received signal against depth and horizontal distance along the array axis. These plots, in showing the distribution of received radar signal energy within the region containing the mine, seem to demonstrate that the mine signals exceed the component of clutter due to subsurface permittivity inhomogeneities. However, the clutter component due to surface reflections is comparable to or larger than the mine signal; hence the latter can not be seen except through time delay resolution.

Some interpretations of the results and some conclusions are presented in the last two sections of the paper. Also discussed are both ongoing and planned enhancements and modifications designed to improve the accuracy of the modeling on which the simulation code is based.

There is a large literature on the modeling of scattering from rough surfaces [1, 2, 3] and volume scattering due to permittivity inhomogeneities [4, 5]. The simulation code used in this study is based on established theoretical models detailed in the cited references. In using these models in the context of this study, we adapted them to the simulation format, i.e. the scattering algorithm is always applied to a small surface element or a small volume element and the incidence and scattering angles for that element depend on the relative positions and orientations of transmitter, receiver and element. Hence, regardless of the specific scattering model used, the code must produce a complex bistatic scattering matrix for arbitrary incidence and scattering angles.

2 SUMMARY OF BACKGROUND ANALYSIS

The theoretical background for the simulation algorithms dealing with received radar signals from earth surface terrain and discrete objects above the surface is given in [6]. The extension to include signals from below the surface [7, 8] will be summarized below.

The algorithm for computation of the received signal voltage in a single receiver at point 2 due to scattering from a small volume centered at point p irradiated by a transmitter at point 1, is based on the vector Born approximation. The geometry is illustrated in Figure 1(a). The 2- D grid of earth surface elements in the x - y plane, numbered 3 to $N+2$, is shown in (b). To include subsurface scattering, this grid is extended to 3- D , where the third dimension is depth $-z$, as illustrated in (c).

Given M depth layers, the volume elements ("cells") in the first (top) layer are numbered $N + 3$ to $2N + 2$, the next layer $2N + 3$ to $3N + 2$ etc. until the deepest layer to be covered by the simulation is reached, its cells being numbered $MN + 3$ to $(M + 1)N + 2$.

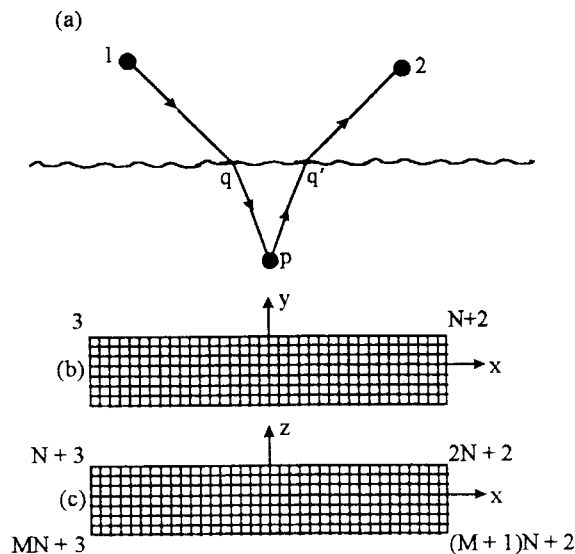


Figure 1: Illustration of problem geometry and 3-D grid

In the generic simulation code described in [6], the single scattering specialization of the received signal voltage is the coherent superposition of the singly-scattered returns from all of the numbered cells. For GPR cases, there are N surface scattering processes and MN volume scattering processes.

The \mathbf{E} -field vector of the transmitted wave at a surface cell q is

$$\mathbf{E}_q = \hat{\Theta}_{1q} E_{\Theta_{1q}} + \hat{\phi}_{1q} E_{\phi_{1q}} \quad (1)$$

where $\hat{\Theta}_{1q}$, $\hat{\phi}_{1q}$ are unit base vectors in the spherical coordinate frame $(r_{1q}, \Theta_{1q}, \phi_{1q})$, centered at 1, and where

$$E_{\Theta_{1q}, \phi_{1q}} = \left[\sqrt{(P_T/4\pi) G_{T0} |f_T(\Theta_{1q}, \phi_{1q})|^2} \right]_{v, h} \exp(-jk_0 r_{1q}) / r_{1q} \quad (2)$$

In (2), P_T , G_{T0} and $f_T(\Theta, \phi)$ are respectively the peak transmitted power, peak antenna gain and complex voltage radiation pattern of the transmitter for vertical ("v") or horizontal ("h") polarization. These polarization directions are defined with respect to the transmitter's local coordinates.

Assuming the air-soil boundary to be horizontal on average and the soil medium to be a LHI half-space on average, the mean \mathbf{E} -field at the center of volume cell p due to downward propagation from surface cell q is (in vector-matrix form)

$$\mathbf{E}(\mathbf{r}_{qp}) = [T_{01}] \cdot \mathbf{E}(\mathbf{r}_{oq}) \exp(-j\mathbf{k}_1 \cdot \mathbf{r}_{qp}) \quad (3)$$

where the tangent plane approximation is used to determine the field just below the surface at q , $[T_{01}]$ is the 3 by 3 matrix of transmission coefficients for downward propagation from air into the soil medium at the point q , and the Stratton-Chu integral [9] evaluated by stationary phase is used to obtain the field at p , thereby excluding wave fields from other surface cells. In (3), \mathbf{k}_1 is the propagation vector for the soil medium, $\mathbf{E}(\mathbf{r}_{qp})$ and $\mathbf{E}(\mathbf{r}_{oq})$ are the field vectors at points p and q respectively, and \mathbf{r}_{qp} and \mathbf{r}_{oq} are respectively the vectors from q to p and from the origin to q .

The "forward path" portion of the computation is the coherent superposition of the fields as given by (3) over all surface cells q and all subsurface volume cells p , resulting in a field vector in every volume cell along the propagation path from a surface cell illuminated by the transmitted wave.

The next phase of the algorithm is the "return path" portion of the computation, whose inputs are the fields given by (3) for all cells p . The first order field vector just below a surface cell q' at position $\mathbf{r} = \mathbf{r}_{oq'}$ due to upward propagation from all volume cells p centered at points $\mathbf{r}' = \mathbf{r}_{op}$, based on the vector Born approximation, is given by

$$\mathbf{E}(\mathbf{r}) = -k_0^2 \int \int \int dx' dy' dz' \bar{G}(\mathbf{r}/\mathbf{r}') \cdot \mathbf{E}(\mathbf{r}') (\Delta \in (\mathbf{r}')) \quad (4)$$

where $\Delta \in (\mathbf{r}')$ is the deviation of complex permittivity from its mean value at \mathbf{r}' , $\mathbf{E}(\mathbf{r}')$ is the field calculated by (3), $\bar{G}(\mathbf{r}/\mathbf{r}')$ is the dyadic Green's function, and in the necessarily digitized algorithm to implement (4), $\mathbf{r}' = \mathbf{r}_{op} = (x', y', z')$ is the vector from the origin to p .

The field vector just above q' is given by

$$\tilde{\mathbf{E}}(\mathbf{r}_{oq'}) = [T_{10}] \cdot \mathbf{E}(\mathbf{r}_{oq'}) \quad (5)$$

where the field components on the right-hand side are those on the left-hand side of (4) and $[T_{10}]$ is the matrix of transmission coefficients for upward propagation from the soil medium back into air.

Again using Stratton Chu, this time to evaluate the field at the receiver due to propagation from q' , we obtain

$$\mathbf{E}(\mathbf{r}_{q'2}) = (K \exp(-jk_0 r_{q'2})/r_{q'2}) [S(\mathbf{r}_{q'2}/\mathbf{r}_{oq'})] \tilde{\mathbf{E}}(\mathbf{r}_{oq'}) \quad (6)$$

where $\mathbf{r}_{q'2}$ and $\mathbf{r}_{oq'}$ represent the vector from q' to the receiver at point 2 and the vector from the origin to the point q' respectively, $\tilde{\mathbf{E}}(\mathbf{r}_{oq'})$ and $\mathbf{E}(\mathbf{r}_{q'2})$ are respectively the field on the left side of (5) and that at the receiver, $[S(\mathbf{r}_{q'2}/\mathbf{r}_{oq'})]$ is a surface scattering matrix derived from Stratton Chu (including the coherent part from the flat mean horizontal surface and the incoherent part due to the surface roughness) and K is a constant given by

$$K = \sqrt{A_{e0} |f_R(\Theta_{2q'}, \phi_{2q'})|^2 \sigma_0 / 4\pi}$$

where A_{e0} is the receiver's peak effective aperture area, $f_R(\Theta, \phi)$ is the receiving antenna's complex voltage receptivity pattern and σ_0 is the peak value of the bistatic scattering cross-section, to whose square root the elements of $[S]$ are normalized. The angles $\Theta_{2q'}$ and $\phi_{2q'}$ are in the receiver's

coordinate frame. They correspond to v and h polarizations with respect to the receiver's coordinate frame, in general different from their counterparts in the transmitter's frame.

Finally we obtain the v and h voltages in the receiver by taking projections of the field vector as given by (6) onto the receiver's aperture plane in the directions along that plane corresponding to those polarizations.

Superposing the computations in (1) through (6) over all MN volume cells and further superposing scatterings from the N surface cells, we obtain the return from the 2- D surface grid and the 3- D volume grid directly below it.

3 THE FOCUSED ARRAY GPR SYSTEM

The focused array scheme [10], shown in Figure 2, is based on two parallel linear arrays, one of them an array of four transmitters parallel to a horizontal axis which we label x . The second array, that of four receivers, is also parallel to the x axis and separated from the transmitter array by a distance $2Y$ in the y direction. The y axis is set midway between the two arrays. In both arrays, the x -separation distance between elements is X ; hence the length of each array is $3X$. The x -origin is at the center of the array. All array elements are at height Z and they are designed to be identical. We can infer from all this that transmitter T_1 is at position $(-1.5X, -Y, Z)$, T_2 at $(-.5X, -Y, Z)$, T_3 at $(.5X, -Y, Z)$ and T_4 at $(1.5X, -Y, Z)$. Corresponding receiver positions are R_1 at $(-1.5X, Y, Z)$, R_2 at $(-.5X, Y, Z)$, R_3 at $(.5X, Y, Z)$ and R_4 at $(1.5X, Y, Z)$. The current values of these parameters are: $X = 10''$, $Y = 5''$, $Z = 15''$.

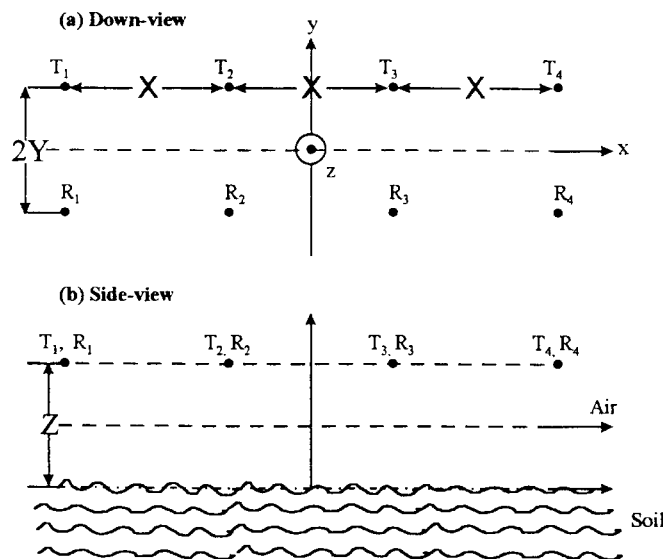


Figure 2: Geometry of Focused Array System

The vehicle-mounted mine detection assembly is moved in the y direction. Every three seconds it scans a specified subsurface volume directly below the pair of arrays, as shown in Figure 2. The volume scanned is envisioned as a rectangular 3-D grid of volume elements containing 16, 1 and 64 in the x , y and z directions respectively. To focus on a particular volume cell centered at point p , pulses are triggered at transmitters T_1 , T_2 , T_3 and T_4 at time instants t_1 , t_2 , t_3 and t_4 respectively in such a manner that they arrive at p at precisely the same time. (Ideally these pulses are conceived as impulse waveforms so that it is possible to define their times of arrival exactly.) The receiver R_1 sees the superposition of the signals from p due to all four transmitters at a time instant $t_{1'}$. R_2 , R_3 and R_4 see the superposition of signals from p at $t_{2'}$, $t_{3'}$ and $t_{4'}$ respectively. The received pulses that the timing algorithm indicates to be from p are then delayed in each receiver such that the effective arrival times for the pulses from p are exactly the same for all four receivers (i.e. the induced delay in R_2 is $t_{1'} - t_{2'}$, that in R_3 is $t_{1'} - t_{3'}$ and that in R_4 is $t_{1'} - t_{4'}$). Then the artificially delayed signals from p are superposed, resulting in a total signal from p consisting of a sum of sixteen signals whose path delays are all equal, i.e. a sharply focused return from p in the ideal case where the transmitted signals are impulses. The sixteen returns from other points within the grid arrive with different delays and hence should be weaker than the signal from p .

The scanning process performed by the system consists of application of the timing protocol as described above to each cell in the grid during the three second scan period, searching for especially strong signals from a particular region of the grid, indicating the presence of a mine.

The task of computing the radar return from all of the cells in the subsurface grid for all sixteen transmitter-receiver position pairs is straightforward for our simulation software. To execute that phase of the simulation we choose the subsurface grid of numbered cells, as described in Section 2, to closely match the grid defined by the timing algorithm in the actual system. This cannot be done exactly without incorporating that algorithm into the simulation code. It is done approximately, maintaining the feature that the numbered grid cells in the code used to compute the radar signals are all of the same size. The vertical dimensions of the cells of the scanned grid in the system itself are different, being dependent on the timing protocol. Because of refraction at the interface and the phase velocity difference between soil and air, equal time delays obviously do not correspond to equal depths and hence the vertical dimension of a grid cell is depth-dependent.

The horizontal dimensions of the grid cells used in the radar signal computations are roughly the same as those of the scanning grid. There are 16 cells in the x -direction each 2" wide and 1 cell in the y direction, of dimension 5". There are 64 cells in the z direction, each of dimension .1875", thus covering depths from zero down to 12". The total coverage region of the computation grid is about the same as that of the scanning grid, from $x = -16''$ to $+16''$, from $y = -2.5''$ to $y = +2.5''$ and from $z = 0$ to $z = -12''$.

The computation of the returns from all 16 surface cells and 1024 volume cells for each pair of transmitter-receiver positions includes the evaluation of the electrical path length τ for each return, using an algorithm based on

$$\tau = (1/c) [(r_{1q} + r_{q'2}) + (c/v)(r_{qp} + r_{pq'})] \quad (7)$$

where c and v are phase velocities in free space and the medium respectively and r_{jk} is the distance between points j and k .

In the system used in the field, the probing signal is a 500 picosecond "time domain" pulse. Our simulation code is designed for frequency domain modeling. The program can model the transmitted pulse spectrum, compute the scattering at sampled frequencies throughout the signal band and IFT the resulting set of frequency samples of the return to construct the received signal pulse. This would be the ideal way to apply the code to this problem. However, that process is very CPU time-intensive and hence was not invoked for this first investigation. The simulations were all executed at 1 GHz.

4 RESULTS

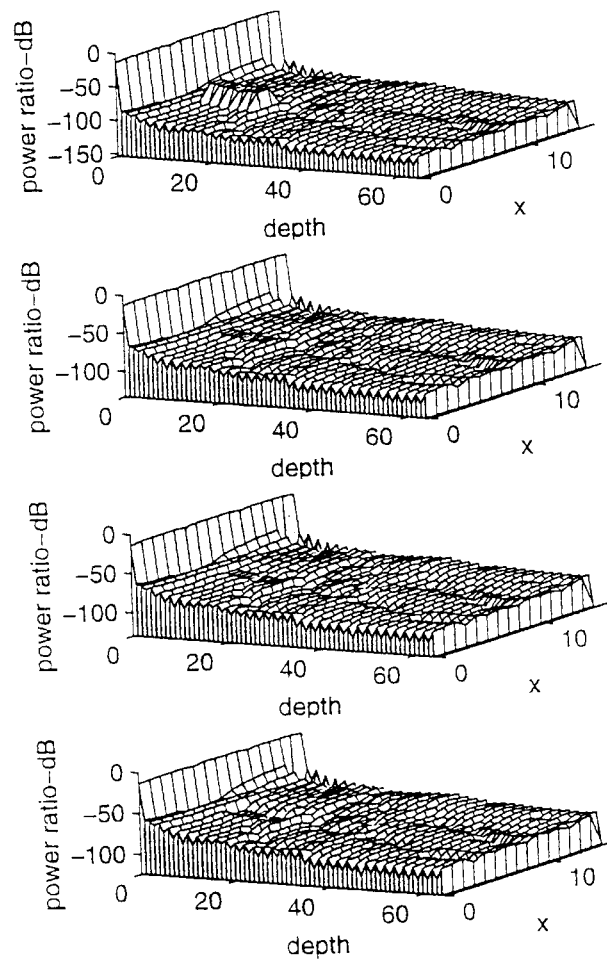
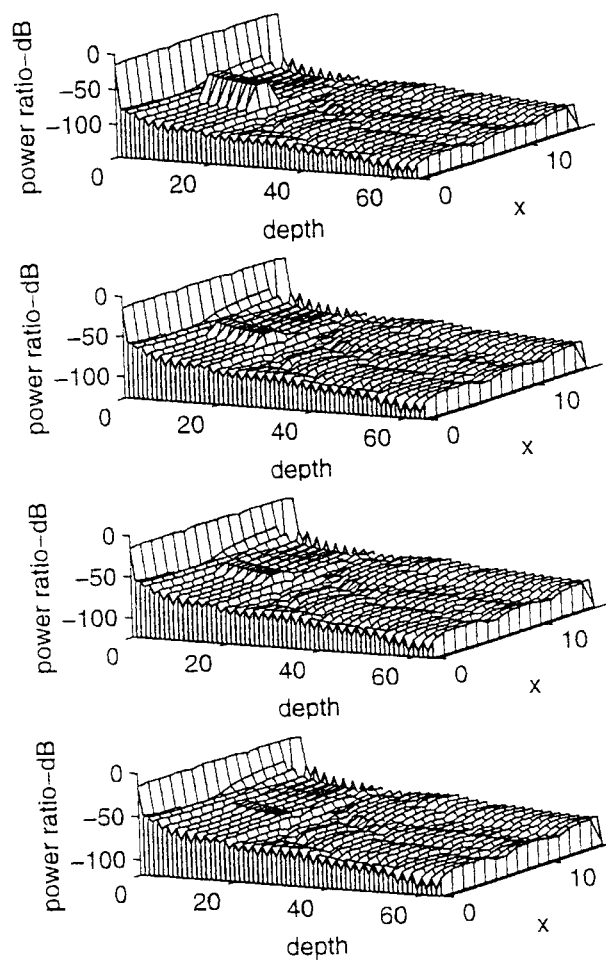
The results to be shown below were obtained by computing the returns from the subsurface grid described in Section 3 for each of the sixteen transmitter-receiver position pairs, then superposing the sixteen amplitudes of returns from each of the 1040 grid cells. This is not an exact simulation of the focused array system's scanning procedure. However, it captures the essential feature of its operation, the superposition of returns from a small volume at various look angles synchronized in time delay so that the return from that region is enhanced relative to that from other regions. The time synchronization can also be achieved in our code by equalizing path delays for all 16 $T - R$ pairs and superposing the resulting returns coherently. In some respects that is closer to the system's true operation than summation of amplitudes. However, it also retains phase differences due to causes other than path delay, so that all 16 sets of returns from the same cell are not really in phase. That detracts from the desired effect and doesn't happen in the actual system, so it would not be a realistic simulation.

The output plots shown here are not in the same format as the experimental results obtained with the time domain system. The latter are acquired by averaging returns from volume cells similar to those generated in our simulation; hence, they contain much less information. An output format like the one we are using allows us to examine in detail the spatial distribution of radar energy to be expected from the illuminated underground region as predicted by our model.

On all of the output plots, the x-axis is along the array axis and is calibrated in cell index numbers in the x-direction, each cell being 2 inches wide in that direction. Thus the x-axis region covered is between $-16''$ and $+16''$. The other horizontal axis is that depth, calibrated in vertical cell index numbers, where each cell is .1875 inches deep, which implies that 64 cells cover depths from 0 to 12 inches. The cell whose vertical index number is 1 is that for the surface return, which is assumed to be resolved by the timing algorithm from the returns from the first depth layer (vertical cell index number 2); hence, the results show the surface return as separated from the underground volume grid returns. In the actual system, the surface return is separated from the total return because it would overwhelm the latter if it were included in the final output display. The ability to resolve surface return from underground returns is a requirement for the feasibility of subtracting it out.

Figures 3,4,5,6:CL,1%,10%,15%,30%

Figures 7,8,9,10:SS,1%,10%,15%,30%



% means percent clutter level
depth in units of 3/16"

x = array axis; in units 2"

CL - clay loam

SS - sandy soil

The vertical axis on the plots is the power received due to scattering from a single cell, where the amplitudes of the returns have been summed over all 16 T - R pairs and the sum has been divided by 16, i.e., the average amplitude (not the average power) has been computed, then squared and plotted in dB relative to the peak transmitted power.

The computed outputs are vpol and hpol components of the returns from all the cells, but the output plots show the square root of the sum of the squares of vpol and hpol returns, i.e., polarization information for each cell has been discarded and we show only the total power for both polarizations.

The source of the clutter model used here is a set of zero mean complex random numbers whose real and imaginary parts have a bivariate Gaussian PDF. The rms values are weighted with assigned fractions of the magnitude of relative complex permittivity of the soil medium. The weighted random numbers are entered into the code as the random variations about the mean of the complex relative permittivity of the underground volume cells. The assigned fractions, multiplied by 100, are designated as "percent clutter levels", e.g., a 1% level means a weighting fraction of .01.

In figure 3-6 and 7-10 respectively, plots are shown for clay loam soil (dielectric constant = 6.2, loss tangent = .145 at 1 GHz) and sandy soil (dielectric constant = 2.55, loss tangent = .03) [11, 12] for a cylindrical dielectric mine (dielectric constant = 3.00, loss tangent = .00033) buried 2" deep, whose vertical dimension and radius are 2" and 1.5" respectively. The surface clutter is shown in all of the plots. Figures 3-6 and 7-10 show how the mine signal becomes less discernible as the clutter level is increased, being clearly distinguishable from the clutter for a 1% clutter level for both soils, still visible at 10% and 15% for clay loam and below 10% for sandy soil. The reason for this difference in mine visibility in the two media is the difference between their dielectric constants relative to that of the mine. This difference is 3.2 for clay loam and -.45 for sandy soil. The ratios of mine contrast to rms clutter are 34 dB and 25 dB respectively for a 1% clutter level. For 10% clutter level, these ratios are 14.3 dB and 5 dB respectively, and they are 10.7 and 1.4 dB for a 15% clutter level. The mine contrast does not descend below the clutter (i.e., the ratio defined above = 0 dB) until the clutter level exceeds 51% in clay loam, but does so in sandy soil when the clutter level exceeds 17.6%.

5 CONCLUSIONS AND DISCUSSION OF CONTINUING WORK

At this stage of the work definitive conclusions about the realism of the results presented here would be premature, since we do not yet have comparisons with experimental results. However, some tentative observations can be made about the potential for use of this simulation code as a forward model for study of mine detection algorithms. The code is based on high frequency approximations and hence cannot be expected to produce images whose accuracy is competitive with purely numerical techniques. It is designed as a fast and easily implemented algorithm to simulate entire systems such as the focused array and to capture the essential features of the

returns from a subsurface region containing a mine.

Our 3-D plots show that the focusing scheme achieves the necessary spatial resolution to provide significant contrast between the region containing a mine and the remaining volume that contains only clutter. The mine image would be still sharper if we fully accounted for the time domain pulse spectrum, which is one of the planned improvements alluded to in the discussion below.

There are a number of ways in which the accuracy of the modeling used to obtain these results can be greatly improved. Work is currently in progress on generalizations accounting for the fact that mines buried only a few inches below the surface are actually in the near zone of points on the surface at frequencies from 0.7 GHz to 1.3 GHz. Hence, the interactions between points on the surface and those on the mine are not really far zone scattering processes. The code is being extended to include some near-zone effects for both forward and return paths. In this approach, the surface integrals are not evaluated by stationary phase; hence waves (not necessarily transverse) propagate in all directions between ground surface and subsurface points, not only in those directions dictated by Snell's law at the air-interface boundary. This generalization results in significantly greater accuracy but greatly increases CPU time. The tradeoff between speed and accuracy for this application will need to be evaluated ultimately by comparing the results with experimental data and determining whether the enhanced modeling accuracy has a significant positive effect on the comparison.

The same remarks would apply to other generalizations that are either in the process of being written or already exist as options in the code. Among these are the time domain modeling of the returns through frequency domain computation of the scattering over the spectrum of the transmitted waveform followed by inverse Fourier transformation, already feasible with the existing code but very CPU time-intensive. Other planned generalizations not yet fully implemented are: inclusion of a distribution of rocks as an additional clutter source, inclusion of higher order interactions between the mine and air-soil boundary, which may have a significant effect on the results for shallow mines, and finally: generalizations beyond the first order Born approximation [13], e.g., first order renormalization [13, 14] and possibly the extended Born approximation [15] (although the latter is most suitable for frequencies below 100MHz). These extensions would decrease the limitations in maximum allowable permittivity contrast between mine and soil that still preserve the validity of the analysis.

References

- [1] Ruck, G. T., D. E. Barrick, W. D. Stuart and C. K. Krichbaum, "Radar Cross-Section Handbook", Volumes 1 and 2, Plenum Press, New York, 1970, Chapter 9, pp 671-770.
- [2] Ulaby, F. T., R. K. Moore and A. K. Fung, "Microwave Remote Sensing: Active and Passive", Volume 1, Addison Wesley, Reading, MA, 1981; Volume 2, Artech House, Norwood, MA, 1982; Volume 3, Artech, 1986; Volume 2, Chapter 12, pp 922-995 and Appendices 12A-12L, pp 996-1032.

- [3] Tsang, L., J. A. Kong and R. T. Shin, "Theory of Microwave Remote Sensing", Wiley, New York, 1985, Chapter 2, Section 6, pp 70-115.
- [4] Ulaby et al, op cit [2] Volume 3, Chapter 13, Sections 13-2 through 13-13-4, pp 1066-1072 and Appendix 13B, pp 1174-1179.
- [5] Tsang, Kong and Shin, op cit[3], Chapter 2, Sections 1, 2, 3, pp 24-41.
- [6] Raemer, H. R., "Analytical Modeling and Computer Simulation of Radar Clutter and Multipath Signals", monograph and project report, Center for Electromagnetics Research, Northeastern University, Boston, MA; Volume 1, May 1989, revised version, 1992; Volume 2, 1994 (chapter contributed by A. K. Bhattacharya), Appendix 3, 1996. Revised version of entire monograph, in preparation.
- [7] Raemer, H. R. and R. Bilotta, "Simulation of Imaging Radar Returns Including Underground Volume Scattering", URSI Symposium, Baltimore, MD, July 1996.
- [8] Raemer, H. R. and E. Miller, "Signal Processing Methods for Subsurface Detection", URSI Symposium, Montreal, July, 1997.
- [9] Stratton, J. A., "Electromagnetic Theory", McGraw Hill, New York, 1941, Chapter 8, Section 8.4, pp 464-470.
- [10] Rappaport, C. M. and D. M. Reidy, "Focused Array Radar for Real Time Imaging and Detection", SPIE Volume 2747, April 1996, pp 202-213.
- [11] Von Hippel, A. R., "Dielectric Materials and Applications", M.I.T. Press, Cambridge, MA, 1954.
- [12] Hipp, J. E., "Soil Electromagnetic Parameters as a Function of Frequency", Proc. IEEE, Vol.62, 1974, pp98-103.
- [13] Tsang, Kong and Shen, op cit [3], Chapter 5, Sections 1-4, pp 317-390.
- [14] Ulaby et al, op cit [2], Section 13-4, pp 1072-1084, Appendix 13-A, pp 1164-1173.
- [15] Torres-Verdin, C. and T. M. Habashy, "Rapid 2.5-Dimensional Forward Modeling and Inversion via a New Nonlinear Scattering Approximation", Radio Science, Vol.29, No.4, July/August 1994, pp 1051-1079.

Signal processing methods for sub-surface detection

Harold Raemer and Eric Miller

ECE Department and Center for Electromagnetics Research,
Northeastern University, Boston, MA 02115

Sophisticated signal processing methods for the detection and localization of buried objects must be built around a scattering model describing the interaction of the probing energy with the medium parameters of interest (i.e., complex electrical permittivity). These algorithms can assume one of two forms. An imaging-type algorithm would use the forward model as part of an optimization routine to construct a pixel-by-pixel image or volumetric rendering of the subsurface permittivity. Alternatively, in the event that one's primary objective is the localization of known objects such as mines or buried drums of hazardous waste, one would make use of this forward model as part of a more structured object-detection algorithm. This class of algorithms typically employs methods such as template matching, matched filtering, or sequential hypothesis testing and is distinct from an imaging algorithm in that these unknown quantities are the number and spatial locations of the objects rather than a large number of pixel or voxel values.

The forward model to be used in this application is an extension of a simulation program developed by one of the authors, whose purpose was to model radar returns from the terrain surface and discrete objects in the radar's field of view for a wide variety of radar scenarios. The extension, discussed in a paper presented at AP-S/URSI in July 1996 (H.R. Raemer and R. Bilotta, "Simulation of imaging radar returns including underground volume scattering"), adds to the existing capabilities of this software the option of modeling the scattered returns due to a specified volume distribution of complex permittivity deviation from the mean value below the ground surface and spatially resolving these contributions. The simplest option uses the first-order Born approximation for vector fields, accounting for the effects of transmission through the air-earth interface. This can be used to model both subsurface clutter and returns from some buried objects. More advanced options include layered media and hence the effects of reflections from interfaces below the ground surface. Still more advanced options not yet fully implemented involve higher-order vector Born approximations. Some examples of the use of this forward model in the context of the signal processing algorithms referred to above are shown in the presentation. The basis of the simulated scenario is some SAR data on buried objects in desert terrain. The case studies include both buried objects and subsurface clutter due to random spatial variations of complex permittivity.

Effects of wavefield interactions on simulated GPR signals

by Harold Raemer, Carey Rappaport and Eric Miller
Northeastern University, Boston, MA, 02115

In previous papers by the authors (Proc. SPIE 3392, April, 1998, pp 754-765 and Proc. SPIE, April, 1999), work was reported on a frequency domain simulation of a bistatic GPR scenario in which dielectric objects, (e.g. mines) are buried a few inches below the ground surface. Received signals from the objects must compete with surface return, the direct signal, scattered signals from random permittivity fluctuations and clusters of rocks within the illuminated subsurface region. Near-field effects and the influence of surface roughness were partially accounted for in our earlier work. However, the reflections from the underside of the air-ground interface due to first-order scattering from the buried objects and permittivity inhomogeneities were neglected. In the present work, this effect is approximately accounted for in determination of the illumination on a subsurface scatterer. Since the distance between the underground scatterers and the interface are within a wavelength in the frequency region of interest (0.7 to 1.3 GHz), this interaction between them can significantly influence the received signal within specified delay gates from which the object's location is inferred in the signal processing. Simulated image plots of the illuminated subsurface region are shown with the above-mentioned effects included in the algorithm and compared with previous results in which these effects have been neglected.

ABSTRACT

THE CONICAL SPIRAL ANTENNA PROBE FOR UNDERGROUND OBJECT DETECTION

Harold Raemer and Carey Rappaport
Department of Electrical and Computer Engineering and
Center for Subsurface Sensing and Imaging Systems
Northeastern University, Boston, MA 02115

In underground object detection by ground penetrating radar, it is necessary to use an antenna with a bandwidth sufficiently large to accommodate wideband transmitted signals and wide receiver passbands. The high spatial resolution attainable with system bandwidths of the order of Gigahertz is limited by employment of antennas with insufficient bandwidths.

Recently the authors initiated an investigation of the use of a particular class of "frequency independent antenna", the conical spiral, in a GPR application. The antenna is used as a probe, thrust into the ground apex first at an angle of about 60 degrees from vertical and radiating primarily along the cone axis. This antenna offers promise of large bandwidth within the constraints of size and portability. The planned deployment method should provide significant enhancement of the information about nearby buried objects (e.g. mines) than the mechanical probes routinely used in mine detection work.

The basic reference sources are a series of papers published in the 1960's in AP Transactions by Rumsey, Dyson, Yeh, Mei and others, on the theory of frequency independent antennas in general and the conical spiral in particular, in free space. Our investigation extends this work to include the partial immersion of the antenna in a soil medium, the interactions between the antenna and the possibly rough air-ground interface and interactions between the antenna and the buried object that the radar is attempting to detect. The frequency range of interest is 500 MHz to 8GHz.

A MOM computation of the currents on a conical spiral antenna 52 cm. long and with a maximum diameter of 12 cm. immersed in dry sand, followed by evaluation of the near field pattern, was carried out by the authors and reported in a recent conference paper (SPIE Aerosense, Orlando, April, 2001). The present paper is a continuation of that work to include other soil media and to begin to address some of the effects of the air-soil interface and the buried object on the operation of the antenna.

Center for Electromagnetics Research

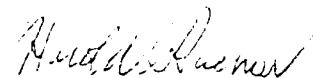
March 29, 1999

Ms. Ellen Kramer
Editor, Proceedings of SPIE
The International Society of Optical Engineering
1000 20th Street
Bellingham, WA 98225

Dear Ms. Kramer:

Enclosed is the revised manuscript of the paper "Near-Field and Timing Effects in Simulation of Focused Array Radar Signals from a Mine in Subsurface Clutter," by Harold Raemer, Carey Rappaport and Eric Miller, whose code number for the SPIE Conference in Orlando in April 1999 is 3710-135. In our recent telephone conversation, I informed you that significant revisions were being made and that I would send you a revised copy before the conference. Since the corrections were made on several pages (Pages 6, 8, 9, 10, 11, 12), please discard the paper you now have and replace it with the enclosed revised version. Thank you.

Sincerely,



Harold Raemer

Near-Field and Timing Effects in Simulation of Focused Array Radar Signals from a Mine in Subsurface Clutter

by

Harold Raemer, Carey Rappaport and Eric Miller

Department of Electrical and Computer Engineering and
Center for Electromagnetics Research
Northeastern University, Boston, MA 02115, U.S.A

Presented at Radar 4, Proc. SPIE 3392, Orlando, FL, April, 1999.

Abstract

In a previous paper (Raemer, Rappaport and Miller, Proc. SPIE 3392, Orlando, FL, April 1998) we discussed a frequency-domain simulation of GPR returns from buried mines in clutter due to random permittivity inhomogeneities, using a focused array radar system (Rappaport and Reidy, SPIE 2747, Orlando, FL, April 1996, pp. 202-213). 3-D image plots of the illuminated volume resulting from this simulation were presented and showed that a mine buried a few inches deep in clay loam or sandy soil appears distinguishable from the clutter if the rms deviation of the permittivity from its mean is less than ten percent of the mean permittivity.

The simulation is designed to be a forward model for signal processing algorithms for mine detection and location. Hence, both accuracy and running speed are important considerations. The code discussed in our previous paper is very fast, but contains approximations that compromise the accuracy of the electromagnetic modeling. The recent work on which the present paper is based addresses improvements in accuracy, emphasizing inclusion of near-field effects and more accurate depiction of the timing algorithm that is the basis of the focused array system. The results obtained from the more accurate algorithms require more running time to obtain but are still sufficiently fast for use as a forward model for signal processing.

1 INTRODUCTION

This paper reports continuing work on the subject of a previous SPIE paper [1] on a frequency domain simulation of a GPR scenario involving a focused array system [2] attempting to detect and locate mines in the presence of clutter. The clutter sources are: the rough surface at the air-soil interface above the mines, a random distribution of complex permittivity throughout the subsurface region containing the mines and an aggregate of rocks in that same region.

The modeling of the propagation of the transmitted wave into the soil, its reflection at the interface and scattering from underground objects and permittivity fluctuations were discussed in [1], together with the focused array system and our method of simulating its use in the mine detection and location scenario of interest in this study.

In the work reported in [1] each ground-patch in the surface grid that covers the radar's field of view is assumed to be in the far-zone of both the transmitted and receiver and hence the field of the wave from transmitter to patch and that scattered from the patch toward the receiver have inverse distance dependence. The far-zone approximation also applies to the wave propagating downward from the patch to a volume element of the subsurface grid and that scattered upward from that element toward another surface patch from which it propagates toward the receiver. These may be valid approximations for the deeper regions of the subsurface coverage volume, but tend to become poorer for the more shallow regions, which are in the near-zones of the surface patches.

The objective of the major enhancement of the code for the present paper is to more accurately model the fields in the near zones of their sources. In the general expressions that describe a field from a source in an LHI medium (those specifically used in this study are the Stratton-Chu integral [3] and the Green's Function for the soil medium [4]) the factor $(\exp(-jkR)/R)F(R)$ always occurs. In this expression R is the distance between source and observation point, the real part of k is $2\pi/\lambda$, where λ is the wavelength, and $F(R)$ is a vector-phasor of the form

$$F(R) = c + (a/jkR) + b(1/jkR)^2 \quad (1)$$

where $|a|/|c|$ and $|b|/|c|$ are quantities of the order of unity (no greater than about 3).

The media assumed in this study are complex in general. The factor $|F(R)|$ is approximately unity in the far zone, where $|1/kR| \ll 1$, and hence the last two terms are negligible in (1). However, if we examine the wavelengths in the media assumed in this study at 1 GHz, and the distances between source and observation points where the most important effects occur, we find that these terms are not necessarily negligible and neglecting them can significantly reduce the accuracy of our simulation code.

Far zone analysis predicts only transverse waves. But longitudinal field components are always present in a general analysis. They are negligible in cases where $|F(R)|$ is approximately unity but begin to appear when $|1/kR|$ becomes large enough to compare significantly with 1, since they only exist in the terms proportional to $(1/jkR)^2$ and never in those proportional to $(1/jkR)$, the latter being the far zone terms.

At 1 GHz, the central frequency of the transmitted signal spectrum in this investigation, the free-space wavelength is 30 cm., or equivalently 11.8 inches. In the two media assumed in the study,

clay loam and dry sand, the wavelengths are about 12 cm. (4.7 ") and 19 cm. (7.5 ") respectively. For distances $R = 1$ inch and $f = 1\text{GHz}$, $|1/kR|$ is about .75 for clay loam and about 1.18 for dry sand. Thus for these media and a wide class of other soils whose constitutive parameters have the same orders of magnitude, the second and third terms of $F(R)$ in (1) may have magnitudes comparable to the first term and should not be neglected in the field computation code. This is particularly true for burial depths of the order of 4 inches or less, where the subsurface scattering points are in the near zone of the air-ground interface.

Consider the waves propagating from the transmitter to the ground surface and from ground to receiver. In this scenario the transmitter and receiver heights are 15 inches. The minimum value of R is 15 " and hence the largest value of $1/kR$, based on the free space wavelength of 11.8" at 1 GHz, is about 0.1. This means that the terms beyond the first in $F(R)$ can probably be neglected in most cases without much compromise in accuracy. In the original code, where all scatterers and receiving points were assumed to be in the far-zones of the sources, these terms were always neglected. In the work we are discussing here, no changes have been made in the portions of the code that model the propagation between transmitter and ground and between ground and receiver.

2 BACKGROUND ANALYSIS

The geometry of a subsurface scattering process was illustrated in Figure 1 in Section 2 of [1], which is repeated here as Figure 1.

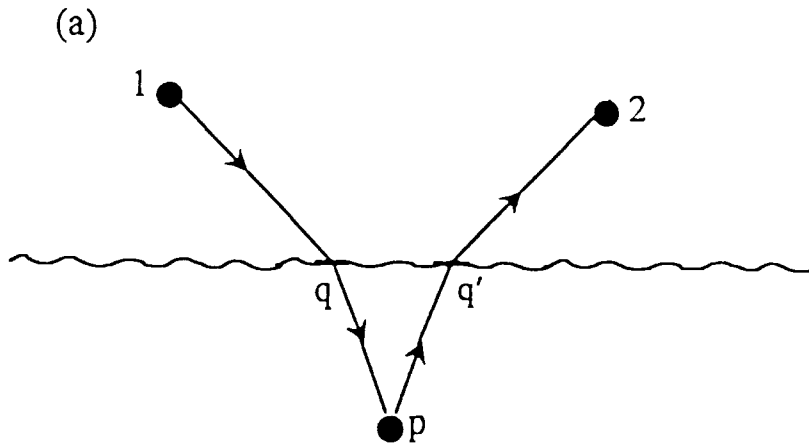


Figure 1: Illustration of Problem Geometry.

Referring to that figure, transmitter and receiver are at points 1 and 2 respectively and the subsurface scatterer is at point p . The transmitted wave propagated downward toward p intersects the air-ground interface at point q . The wave propagating upward from p toward the receiver intersects the interface at q' .

The surface patch centered at q is small, allowing the incoming wave from point 1 to be approximated locally as a plane wave since the curvature of the phase-front is negligible within the patch. Its traversal of the boundary is treated at point q as if it were a plane wave incident on an infinite horizontal plane and is modeled accordingly. The vertical propagation plane of that wave is the x' , z' plane of a Cartesian coordinate system. The y' components of it's electric ("E") and magnetic ("H") fields immediately below the surface are:

$$E_{qv'}^- = T_{01}^{TE} E_{qv'}^+ \quad (2-a)$$

$$H_{qv'}^- = T_{01}^{TM} H_{qv'}^+ \quad (2-b)$$

where T_{01}^{TE} and T_{01}^{TM} are respectively the transverse electric and transverse magnetic downward transmission coefficients and the E^+ and H^+ on the right hand sides are those immediately above the surface. The arguments of E and H are the vectors from the origin of a global coordinate system to the center of the surface patch at point q .

Since there are many processes of the kind described here in each simulation run and they must be superposed, they must all be in the same coordinate frame. The global Cartesian frame used for this purpose has its x , y plane on the horizontal air-ground interface. Its x axis is at an angle ϕ_{1q} with respect to the x' axis of the primed system alluded to above. Obviously $z = z'$. The field vectors in the global frame are:

$$\mathbf{E}_q^- = \hat{x} \left[-Z_1 H_{qv'}^-, \cos \Theta_t \cos \phi_{1q} - E_{qv'}^-, \sin \phi_{1q} \right] \quad (3-a)$$

$$+ \hat{y} \left[-Z_1 H_{qv'}^-, \cos \Theta_t \sin \phi_{1q} + E_{qv'}^-, \cos \phi_{1q} \right] + \hat{z} \left[-(Z_0/\epsilon_{1R}) H_{qv'}^-, \sin \Theta_i \right]$$

$$\mathbf{H}_q^- = \hat{x} \left[Y_1 E_{qv'}^-, \cos \Theta_t \cos \phi_{1q} - H_{qv'}^-, \sin \phi_{1q} \right] \quad (3-b)$$

$$+ \hat{y} \left[Y_1 E_{qv'}^-, \cos \Theta_t \sin \phi_{1q} + H_{qv'}^-, \cos \phi_{1q} \right] + \hat{z} \left[Y_0 E_{qv'}^-, \sin \Theta_i \right]$$

where

$$\cos \Theta_t = \sqrt{1 - \frac{\sin^2 \Theta_i}{\nu_1^2}} = \frac{1}{\sqrt{\epsilon_{1R}}} \sqrt{\epsilon_{1R} - \sin^2 \Theta_i}$$

where Z_1 and Y_1 are respectively the complex wave impedance and admittance of the medium, \hat{x} , \hat{y} , \hat{z} are the unit base vectors along the global coordinate axes, Θ_i is the angle of incidence of the transmitted wave at q , ν_1 is the complex refractive index, equivalent to the square root of ϵ_{1R} the complex relative permittivity. Z_0 and Y_0 are wave impedance and admittance of free space, where $Z_1 = Z_0/\sqrt{\epsilon_{1R}}$ and $Y_1 = Y_0\sqrt{\epsilon_{1R}}$.

After evaluation of \mathbf{E}_q^- and \mathbf{H}_q^- , these field vectors are viewed as sources of waves propagating downward from point q . More precisely, the surface patch whose center is at q is treated as an

incremental element of the Stratton-Chu integral over the entire ground surface assumed to be illuminated by the transmitting antenna beam. Based on this, the electric field vector at an arbitrary point p within the underground volume is:

$$\mathbf{E}(\mathbf{r}_p) = \frac{1}{4\pi} \int_{\Delta S_q} \int d^2\mathbf{r}_q \left(\frac{e^{-jk_1 r_{qp}}}{r_{qp}} \right) \tilde{\mathbf{E}}(\mathbf{r}_p/\mathbf{r}_q) \quad (4)$$

where

$$\tilde{E}_x(\mathbf{r}_p/\mathbf{r}_q) = -j\omega\mu_0 H_y(\mathbf{r}_q) + d(\mathbf{r}_{qp}) [E_x(\mathbf{r}_q) \cos \Theta_{qp} + E_z(\mathbf{r}_q) \sin \Theta_{qp} \cos \phi_{qp}] \quad (4-a)$$

$$\tilde{E}_y(\mathbf{r}_p/\mathbf{r}_q) = j\omega\mu_0 H_x(\mathbf{r}_q) + d(\mathbf{r}_{qp}) [E_y(\mathbf{r}_q) \cos \Theta_{qp} + E_z(\mathbf{r}_q) \sin \Theta_{qp} \cos \phi_{qp}] \quad (4-b)$$

$$\tilde{E}_z(\mathbf{r}_p/\mathbf{r}_q) = d(\mathbf{r}_{qp}) [\sin \Theta_{qp} (E_x(\mathbf{r}_q) \cos \phi_{qp} + E_y(\mathbf{r}_q) \sin \phi_{qp}) + E_z(\mathbf{r}_q) \cos \Theta_{qp}] \quad (4-c)$$

and where

$d(\mathbf{r}_{qp}) = jk_1 \left(1 + \frac{1}{jk_1 r_{qp}} \right)$, ΔS_q is the area of the surface patch, (Θ_{qp}, ϕ_{qp}) are the spherical polar angles of point p as viewed from point q in the global coordinates and $d^2\mathbf{r}_q$ is a surface element on the patch.

The next phase of the computation is the vector Born scattering [5] from a small volume element centered at p . The general expression for this is given by Eq. (4) of [1], repeated below as Eq. (5).

$$\mathbf{E}(\mathbf{r}) = -k_1^2 \int \int \int d\mathbf{r}' \bar{\bar{G}}(\mathbf{r}/\mathbf{r}') \cdot \mathbf{E}(\mathbf{r}') (\Delta \in (\mathbf{r}')) \quad (5)$$

where $\Delta \in (\mathbf{r}')$ is the deviation of complex permittivity from its mean value at \mathbf{r}' , $\mathbf{E}(\mathbf{r}')$ is the field calculated by (3), $\bar{\bar{G}}(\mathbf{r}/\mathbf{r}')$ is the dyadic Green's function, and in the necessarily digitized algorithm to implement (5), $\mathbf{r}' = (x', y', z') = \mathbf{r}_{op}$ is the vector from the origin to a point within the volume cell centered at p .

The dyadic Green's function in (5) [6] is a symmetric 3×3 matrix given by

$$\bar{\bar{G}}(\mathbf{r}/\mathbf{r}') = -g(R) \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ G_{12} & G_{22} & G_{23} \\ G_{13} & G_{23} & G_{33} \end{bmatrix} \quad (6)$$

where

$$\begin{aligned} g(R) &= e^{jk_1 R}/4\pi R \\ R &= |\mathbf{r} - \mathbf{r}'| \end{aligned}$$

and where the matrix elements are given generically by

$$G_{jk} = BU_j U_k - A\delta_{jk}$$

where δ_{jk} is the Kroneker delta (1 if $j = k$, 0 if $j \neq k$),

$$\begin{aligned} U_1 &= (x - x')/R \\ U_2 &= (y - y')/R \\ U_3 &= (z - z')/R \end{aligned}$$

$$A = \frac{1}{jk_1 R} + \left(\frac{1}{jk_1 R}\right)^2; \quad B = 1 + 3A$$

The observation point in (5) is specialized to the point q' as illustrated in Figure 1. Then counterparts of the steps implemented in modeling the downward boundary traversal at q are executed to model the upward boundary traversal at q' . First, we postulate another primed coordinate system whose x' , z' plane is the propagation plane from q' to the receiver at point 2. The y' component of the electric field immediately below the boundary at q' is obtained from (5) and is given by

$$E_{y'}^-(\mathbf{r}_{q'}) = -E_x^-(\mathbf{r}_{q'}) \sin \phi_{pq'} + E_y^-(\mathbf{r}_{q'}) \cos \phi_{pq'} \quad (7)$$

where $\phi_{pq'}$ is the azimuthal angle of the point q' relative to p in the global frame.

The y' component of the electric field immediately above the patch centered at q' is given by

$$E_{y'}^+(\mathbf{r}_{q'}) = T_{10}^{TE} E_{y'}^-(\mathbf{r}_{q'}) \quad (8)$$

where T_{10}^{TE} is the upward transmission coefficient for the TE mode.

To obtain the TM field above the boundary, we must first evaluate the y' component of the magnetic field immediately below the boundary. Again viewing the upward-propagating wave from p toward q' in the vicinity of q' as approximated by a plane wave, we can calculate the y' component of the magnetic field from the electric field given by (5). The result is

$$\begin{aligned} H_{y'}^-(\mathbf{r}_{q'}) = \\ Y_1 \left[E_x(\mathbf{r}_{q'}) \cos \Theta'_{pq'} \cos \phi_{pq'} + E_y(\mathbf{r}_{q'}) \cos \Theta'_{pq'} \cos \phi_{pq'} - E_z(\mathbf{r}_{q'}) \sin \Theta'_{pq'} \right] \end{aligned} \quad (9)$$

where $\Theta_{pq'}$ is the spherical polar angle of q' relative to p' in the global frame.

Next we calculate the y' component of the magnetic field immediately above q' from that immediately below q' , with the result

$$H_{y'}^+(\mathbf{r}_{q'}) = T_{10}^{TM} H_{y'}^-(\mathbf{r}_{q'}) \quad (10)$$

where T_{10}^{TM} is the upward transmission coefficient for the TM mode.

From (8) and (10), the Maxwell equations with the plane wave approximation and an expression for $(\hat{x}, \hat{y}, \hat{z})$, the unit base vectors in the global frame, in terms of $(\hat{x}', \hat{y}', \hat{z}')$, the unit base vectors in the primed coordinate system, we obtain the components of $\mathbf{E}^+(\mathbf{r}_{q'})$ and $\mathbf{H}^+(\mathbf{r}_{q'})$ in the global frame, given by

$$\begin{aligned} \mathbf{E}^+(\mathbf{r}_{q'}) = & \hat{x} \left[-E_{y'}^+(\mathbf{r}_{q'}) \sin \phi_{q'2} + Z_0 H_{y'}^+(\mathbf{r}_{q'}) \cos \Theta_{q'2} \cos \phi_{q'2} \right] \\ & + \hat{y} \left[E_{y'}^+(\mathbf{r}_{q'}) \cos \phi_{q'2} + Z_0 H_{y'}^+(\mathbf{r}_{q'}) \cos \Theta_{q'2} \sin \phi_{q'2} \right] \\ & + \hat{z} \left[-Z_0 H_{y'}^+(\mathbf{r}_{q'}) \sin \Theta_{q'2} \right] \end{aligned} \quad (11-a)$$

$$\begin{aligned} \mathbf{H}^+(\mathbf{r}_{q'}) = & \hat{x} \left[-H_{y'}^+(\mathbf{r}_{q'}) \sin \phi_{q'2} - Y_0 E_{y'}^+(\mathbf{r}_{q'}) \cos \Theta_{q'2} \cos \phi_{q'2} \right] \\ & + \hat{y} \left[H_{y'}^+(\mathbf{r}_{q'}) \cos \phi_{q'2} - Y_0 E_{y'}^+(\mathbf{r}_{q'}) \cos \Theta_{q'2} \sin \phi_{q'2} \right] \\ & + \hat{z} \left[Y_0 E_{y'}^+(\mathbf{r}_{q'}) \sin \Theta_{q'2} \right] \end{aligned} \quad (11-b)$$

where $\Theta_{q'2}$ and $\phi_{q'2}$ are the spherical coordinate angles of the receiver at point 2 relative to the point q' as origin.

The fields given by Eqs. (11-a, b) at the point q' act as sources for the wave propagating from q' toward the receiver.

The portions of the analysis not explicitly included in the discussion above are:

- (a) the process by which the field components $E_{y'}^+$ and $H_{y'}^+$ in Eqs. (2-a, b) are obtained from the fields generated at the transmitting antenna at point 1;
- (b) the determination of the fields at the receiver due to the wave propagating from q' ;
- (c) explicit indication of the transmission coefficients in Eqs. (2-a, b), (8) and (10);
- (d) evaluation of the surface integral indicated by (4);
- (e) simulation of the focused array system.

Items (a) and (b) are part of the radar simulation program that serves as a framework for the GPR code [7]. Bistatic radar returns from a grid of small ground surface patches are modeled in that program. It includes the effect of the antenna patterns for both transmitter and receiver and the direct signal from transmitter to receiver. The theoretical background for all of the received

signals from scattering regions on or above the ground (i.e. the direct signal and the ground-reflected signals) is covered in [7] and will not be given here, the focus of this paper being returns from below the ground surface, as illustrated in Figure 1. The enhancement of the original code to include subsurface returns begins with the transmitted wave field impinging on the surface patch centered at point q . That field is the source for both the wave scattered toward the receiver from that patch and the wave transmitted across the boundary toward the subsurface point p as described by Eqs. (2)-(11). The patch centered at point q' in Figure 1, which receives the subsurface wave scattered from p , also intercepts a wave from the transmitter and scatters it toward the receiver. All of these processes, executed thousands of times, are the constituents of the overall simulation program. Turning to Item (c), the transmission coefficients in Eqs. (2-a, b) are given by

$$T_{01}^{TE} = \frac{2 \cos \Theta_i}{\cos \Theta_i + \sqrt{\epsilon_{iR} - \sin^2 \Theta_i}} \quad (12-a)$$

$$T_{01}^{TM} = \frac{2\epsilon_{iR} \cos \Theta_i}{\epsilon_{iR} \cos \Theta_i + \sqrt{\epsilon_{iR} - \sin^2 \Theta_i}} \quad (12-b)$$

and those in (8) and (10) are

$$T_{10}^{TE} = \frac{2\sqrt{\epsilon_{iR}} \cos \Theta'_i}{\sqrt{\epsilon_{iR}} \cos \Theta'_i + \sqrt{1 - \epsilon_{iR} \sin^2 \Theta'_i}} \quad (13-a)$$

$$T_{10}^{TM} = \frac{2 \cos \Theta'_i}{\cos \Theta'_i + \sqrt{\epsilon_{iR}} \sqrt{1 - \epsilon_{iR} \sin^2 \Theta'_i}} \quad (13-b)$$

where Θ'_i is the angle of incidence for the wave propagating upward from p toward q' .

Item (d), the surface integration in (4), can be treated in one of two ways, and both options are available in the code. The option used in the computations done for this paper is a straightforward integration over the patch centered at q under the assumption that the field is constant over the patch surface and equal to its value at the center. The patches are small enough to use the standard approximation

$$\exp(-jk_1 r_{qp})/r_{qp} \simeq [\exp(-jk_1 r_{qp})/r_{qp}]_0 \exp(jk_1 \mathbf{r}'_q \cdot (\hat{\mathbf{r}}_{1q} + \hat{\mathbf{r}}_{qp})) \quad (14)$$

where subscript 0 indicates that the point q is at the patch center. This leads to a product of complex expressions for the x and y integrals, each of which degenerates into a sinc function in the limit of zero conductivity.

The other option is to find stationary phase points and to obtain the stationary phase approximation of the integral at only those points. The relative merit of these two approaches is the subject of continuing work but will not be discussed further here. Item (e), the operating principle of the focused array system, was covered in detail in [2] and further summarized in [1] and

will not be repeated here. To simulate the action of the focused array we are required to execute the computation of the received radar signals from the illuminated region 16 times, once for each transmitter-receiver position pair, and combine the outputs of each run of the code for every one of the illuminated volume cells, as explained in [1] (Section 3, "The focused array GPR system").

3 RESULTS AND CONCLUSIONS

The 3D plots in Figures 2 through 9 show outputs of simulations of received signals from a volume directly below the focused array. The horizontal locations of the four transmitters $T_k = 1, 2, 3, 4$ and the four receivers $R_k = 1, 2, 3, 4$ are shown in Figure 2 of [1]. This assembly moves along the y axis ("alongtrack" direction). The x axis ("crosstrack" direction) is parallel to transmitter and receiver array axes, which are at $y = -5''$ and $+5''$ respectively and both of which extend from $x = -15''$ to $x = +15''$ and are at heights of $15''$ above ground. This configuration corresponds roughly to a version of the focused array system used to obtain some of the experimental results we have received from Geo-Centers. The underground volume grid covered in these simulations extends from $-15''$ to $+15''$ in the crosstrack direction, from $-5''$ to $+5''$ in the along-track direction and from the ground surface to a depth of $8''$. Figures are plots of received power (in dB over a given reference level) from each small volume element in the grid, averaged over all 16 transmitter-receiver position pairs, vs. the depth and the crosstrack coordinate (where the latter is calibrated in inches from 0 to 30, rather than from -15 to $+15$, as depicted in Figure 2 of [1]). The scattering computations whose background theory was discussed in Section 2 were performed for each of 3072 volume elements of 1 inch dimension in all 3 directions. There are 32 x elements, 12 y elements and 8 depth elements. For each x and depth value, the received powers from all 12 y elements are summed to generate the total return from a single y element extending from $-5''$ to $+5''$. This refinement of the y scale for the computations addresses a problem that existed in the results presented in [1]. In those earlier simulations the grid used for the computations was the same as the grid used for the 3D output plots, which in turn is the same grid as used to collect experimental data as the moving detector assembly stops to examine a region directly below it (16, 1 and 64 in x , y and z directions respectively, totaling to 1024 elements). The physics of the scattering computations is less accurately depicted if the scatterer dimension in the y direction is 10 inches than if that dimension is only 1 inch. A better simulation of the scattering process is realized by computing the returns from 12 small elements and summing them to generate an output from a single element.

In all of the plots shown, the mine is a slightly lossy dielectric cylindrical object $3''$ high and $3''$ in diameter, with complex relative permittivity of TNT (3, $-.001$) at 1 GHz. Its burial depth is indicated on each figure. The rocks are slightly lossy dielectric spheres with radii between $.75''$ and $1.5''$, burial depth between $1.5''$ and $3''$ and distributed horizontally around the illuminated region, with dielectric constant between 2 and 5. The soil is clay loam or dry sand, with mean relative complex permittivity (6.2, -0.9) or (2.55, $-.077$) respectively at 1 GHz. The clutter from each element is modeled as a bivariate Gaussian distribution of the soil's complex permittivity around its mean. A given "percent clutter" is the ratio of average deviation to the mean value times 100.

Figures 2 and 3 show image plots for a mine buried 2" deep in clay loam soil for 1 and 10 percent level of permittivity fluctuation clutter respectively and Figures 4 and 5 are analogous plots for dry sand. There is no clutter due to rocks in any of these cases. With either soil type, the mine is easily discernible in 1 percent clutter and imperceptible in 10 percent clutter based on modeling not accounting for near field effects. Inclusion of these effects does alter the shapes of the image plots somewhat but does not greatly change the dependence of mine visibility on clutter level on these plots. The near field effects appear to influence the mine signals and clutter signals to about the same extent. In Figures 6 and 7, the effects of 10 rocks in the vicinity of the mine is shown for both soil types, where clutter due to permittivity fluctuations has been suppressed. Even in the absence of such clutter, and more prominently when both rocks and permittivity fluctuation clutter are present (Figures 8 and 9), there appears to be no way to distinguish a mine from one of the rocks with the spatial resolution in our output displays. That resolution is set to be comparable to that in the actual system as provided by the focused array principle in the crosstrack and depth directions, which partially offsets the delay resolution limitations of a 500 picosecond pulse. It would be feasible to improve the resolution in our displays by decreasing the size of the elements in the volume grid used for the output displays and thereby enable discrimination on the basis of shape, but that would reduce the realism of the model and therefore is not a permissible option.

Figure 2: 1 mine,1% clutter,clay loam

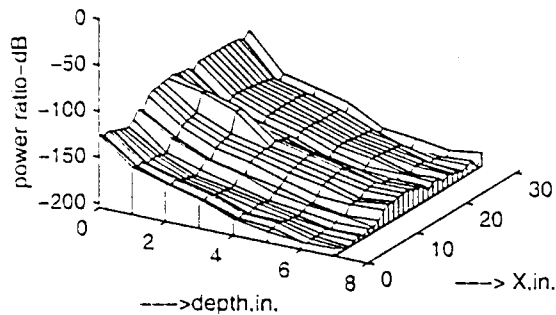


Figure 4:1 mine,1% clutter,dry sand

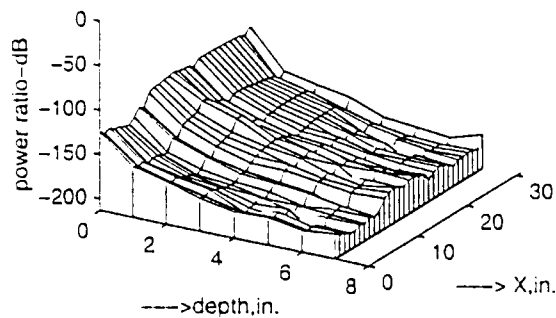


Figure 3: 1 mine,10% clutter,clay loam

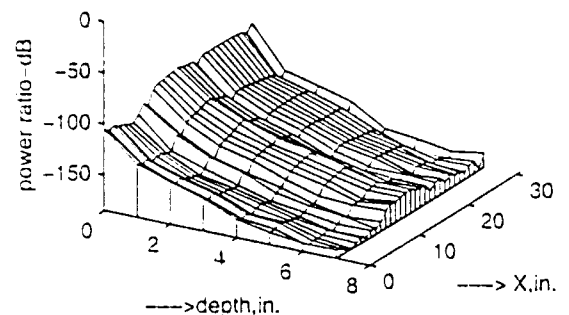


Figure 5:1 mine,10% clutter,dry sand

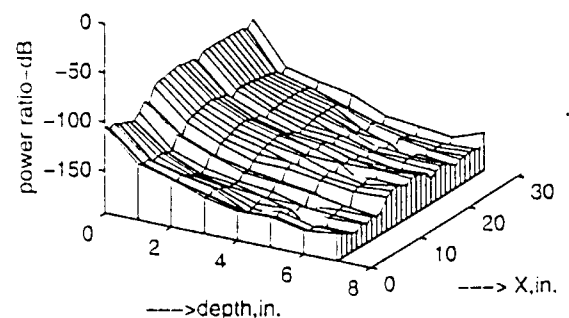


Figure 6: 1 mine,10 rocks,0% clutter,clay loam

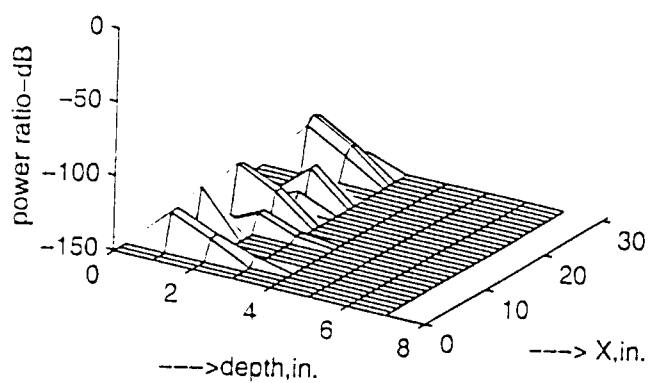


Figure 7: 1 mine,10 rocks,0% clutter,dry sand

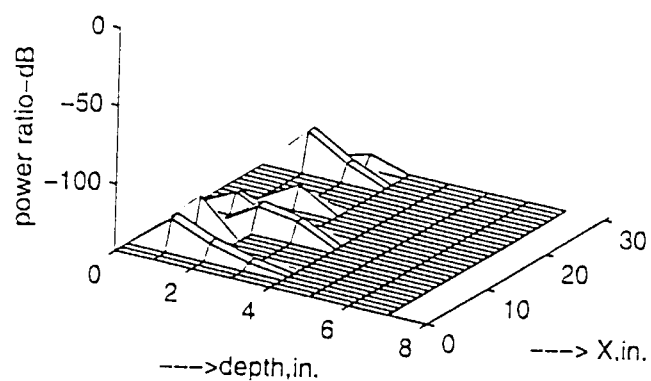


Figure 8: 1 mine,10 rocks,3% clutter,clay loam

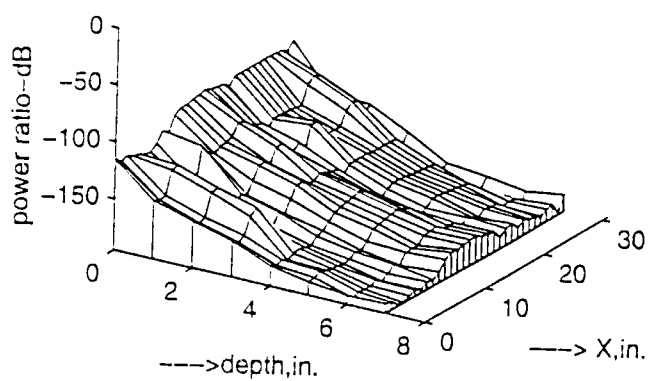
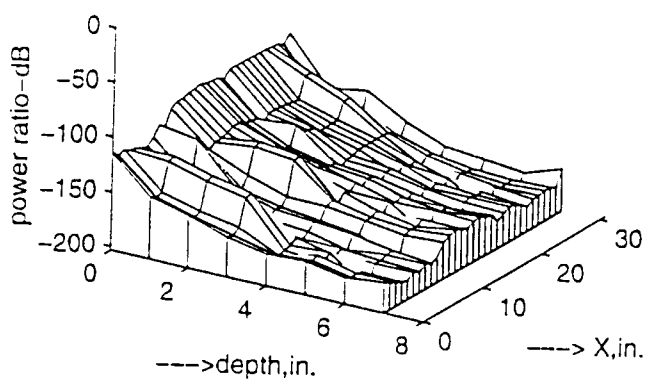


Figure 9: 1 mine,10 rocks,3% clutter,dry sand



4 CONTINUING WORK

In Section 5 of [1], we cited ongoing efforts to improve the realism and scope of the code. One task was inclusion of near-field effects, which is the major theme of this paper. Another was study of rocks as clutter sources, also included in this paper. Still not implemented are generalizations beyond the first order Born theory, especially effects of reflection of scattered waves from subsurface volume elements from the underside of the air-soil interface, resulting in second order contributions to the illumination of these elements. Another study, now in progress but not mentioned in [1], is the effect of local surface tilt due to large-scale roughness on scatterer locations indicated by the timing algorithm. This could be an important source of distortion of image plots of the return from the illuminated volume. In addition to carrying out the studies indicated above, we are reconstructing the output formats to simulate a scenario that matches the geometry consistent with minefield data received from Geo-Centers, a step required for comparisons with experimental results.

References

- [1] Raemer, H. R., C. M. Rappaport and E. L. Miller, "Frequency domain simulation of focused array radar returns from buried mines in clutter", SPIE Volume 3392, Orlando, FL, April 1998, pp 754-765.
- [2] Rappaport, C. M. and D. M. Reidy, "Focused array radar for real time imaging and detection", SPIE Volume 2747, Orlando, FL, April 1996, pp 202-213.
- [3] Stratton, J. A., "Electromagnetic Theory", McGraw Hill, New York, 1941, Chapter 8, Section 8.4, pp 464-470.
- [4] Tsang, L., J. A. Kong and R. T. Shin, "Theory of Microwave Remote Sensing", Wiley, New York, 1985, Chapter 2, Section 3, pp 32-41 and Chapter 5, Section 2, pp 319-337.
- [5] Tsang, Kong and Shen, op cit [4], pp 319-328.
- [6] Tsang, Kong and Shen, op cit [4], pp 32-33.
- [7] Raemer, H. R., "Analytical Modeling and Computer Simulation of Radar Clutter and Multipath Signals", monograph and project report, Center for Electromagnetics Research, Northeastern University, Boston, MA; Volume 1, May 1989, revised version, 1992; Volume 2, 1994 (chapter contributed by A. K. Bhattacharya), Appendix 3, 1996.



Center for Electromagnetics Research

June 23, 1999

Ms. Kristina Steger
Editor, Proceedings of SPIE
SPIE – International Society for Optical Engineering
P. O. Box 10
Bellingham, WA 98227-0010

Dear Ms. Steger:

Enclosed is the manuscript for Paper 3752-54, which was officially due on June 21, 1999. In our telephone conversation last week, you agreed to extend the deadline to June 28, 1999. I believe you will receive this before that date.

Thank you for the extension.

Sincerely,

A handwritten signature in cursive script that reads 'Harold Raemer'.

Harold Raemer
Professor

Paper Number 3752-54
Surface Roughness Effects in Simulation of Mine Signals
Using a Focused Array GPR System

by

Harold Raemer, Carey Rappaport and Eric Miller

Department of Electrical and Computer Engineering and
Center for Electromagnetics Research
Northeastern University, Boston, MA 02115, U.S.A

Presented at SPIE International Symposium, Conference 3752,
Denver, CO, July, 1999.

Abstract

In two previous papers, the authors discussed a frequency-domain simulation of received signals in a focused array radar system (Rappaport and Reidy, SPIE 2747, April 1996, pp. 202-213) illuminating a swath of flat terrain with one or more small plastic mines buried a few inches below the surface and a clutter background consisting of returns from random permittivity fluctuations, aggregates of rocks and the ground surface (SPIE 3392, pp 754-765, April 1998 and "Near-field and timing effects in simulation of focused array radar signals from a mine in subsurface clutter", SPIE, April 1999). The second paper emphasizes generalization of the algorithm to take proper account of the fact that distances between participants in the scattering processes are within fractions of wavelengths, enhancing the importance of near field effects in the modeling.

In the present work, further generalizations are made which enhance the realism of the model. In particular, we investigate the role of surface height variations in changing the apparent propagation delay of the signal from a subsurface region. Since delay is the major discriminant in the processing of received signals in this system, this has a possibly significant effect on the simulated images. Still another generalization under investigation is that of inclusion of surface reflections from the interface's underside on the effective illumination of subsurface regions.

Simulated image plots will be shown and compared with some previous results in order to assess the effects of these enhancements in the model.

1 INTRODUCTION

In this paper we discuss further progress on a study reported in two previous SPIE papers by the authors [1, 2]. The objective of the study is to develop an effective simulation of the received signals in a focused array radar system [3] illuminating a swath of land terrain with small plastic mines buried within a few inches below the ground surface. The ultimate purpose of this simulation is to provide a forward model for evaluation of signal processing algorithms for improvement of detection of the mines in a clutter background. The clutter consists of random fluctuations of the complex permittivity of the soil, aggregates of rocks and returns from the ground surface.

Previous improvements in the accuracy of the modeling [2] were directed toward accounting for the fact that the buried mines, rocks and the permittivity fluctuations from which scattering occurs are within fractions of a wavelength from the ground surface. Hence, near field effects were addressed in [2].

In [2], the air-ground interface was assumed to be horizontal on the average, with the possibility of small-scale fluctuations around the average height.

In the present paper, both small scale and large scale height variations around the horizontal are allowed, with a significant effect on the timing algorithm on which the focused array system is based. Application of that algorithm, which is based on the assumption of a flat horizontal ground surface, results in a distortion of the spatial distribution of the illumination of the underground volume of interest and that of the returns from that volume if the true surface is rough.

2 BACKGROUND ANALYSIS

In Section 2 of [2], aided by a diagram in Figure 1 of [2], we described the background analysis for simulation of the propagation of a ray from a transmitter at point 1 to a subsurface point label p , via a refracting surface point q . The diagram of Figure 1 of [2] then shows a scattered wave from q propagating toward a different surface point q' , from which it propagates toward the receiver at point 2. In that diagram the ground surface is shown as nearly flat and horizontal with very small-scale height fluctuations.

In the code that generated the simulation discussed in [2], the refracting angles at points q and q' are based on the average surface height and are independent of the small-scale height fluctuations. The only assumed effect of these fluctuations is on the scattered wave from the ground surface to the receiver and they have no effect on the returns from the illuminated subsurface region.

Part (a) of Figure 1 in this paper is Figure 1 of [2]. Part (b) of Figure 1 illustrates the generalization that is the subject of the most recent work. The surface in Figure 1(b) is shown to have large-scale height variations on which are superposed small-scale variations like those shown in Part (a). The surface patches centered at q and q' now have an average tilt with respect to the horizontal. As in [2], the small-scale surface fluctuations on the now tilted patch have no effect on the refraction angles, which are based on the average height and tilt angles. However, they do have a significant effect on the surface return, whose ratio of incoherent-to-coherent scattering increases with the small-scale rms height variation.

This generalization complicates the geometry considerably and requires a re-definition of the primed coordinate system used in [1] and [2]. To explain this, some of the equations and associated text will be repeated below, beginning with Eqs. (1) and (2) of [1].

The E-field vector of the transmitted wave at a surface cell q is

$$\mathbf{E}_q = \hat{\Theta}_{1q} E_{\Theta_{1q}} + \hat{\Phi}_{1q} E_{\Phi_{1q}} \quad (1)$$

where $\hat{\Theta}_{1q}, \hat{\Phi}_{1q}$ are unit base vectors in the spherical coordinate frame $(r_{1q}, \Theta_{1q}, \Phi_{1q})$, centered at 1, and where

$$E_{\Theta_{1q}, \Phi_{1q}} = \left[\sqrt{(P_T/4\pi) G_{T0} |f_T(\Theta_{1q}, \Phi_{1q})|^2} \right]_{v, h} \exp(-jk_0 r_{1q})/r_{1q} \quad (2)$$

In (2), P_T, G_{T0} and $f_T(\Theta, \Phi)$ are respectively the peak transmitted power, peak antenna gain and complex voltage radiation pattern of the transmitter for vertical (“v”) or horizontal (“h”) polarization. These polarization directions are defined with respect to the transmitter’s local coordinates.

The magnetic field components of the transmitted wave at q , obtained from \mathbf{E}_q , are

$$H_{\Phi_{1q}} = -Y_0 E_{\Theta_{1q}}, \quad H_{\Theta_{1q}} = Y_0 E_{\Phi_{1q}} \quad (3)$$

where Y_0 is the wave admittance of free-space.

As in [2], we consider the surface patch centered at q to be small enough to allow the incoming wave to be approximated locally as if it were a plane wave incident on an infinite plane (i.e. the “tangent plane approximation.”) But the departure from the analysis in [2] occurs at this point, wherein the plane of the patch is tilted with respect to the horizontal direction. To determine the tilt angles, we first note that the unit normal to the patch at q , pointing into free space, is given by

$$\hat{n}_q = \frac{-(z_x)_q \hat{x} - (z_y)_q \hat{y} + \hat{z}}{\sqrt{1 + (z_x)_q^2 + (z_y)_q^2}} \quad (4)$$

where $(\hat{x}, \hat{y}, \hat{z})$ are the unit base vectors for the global coordinate system and $(z_x)_q$ and $(z_y)_q$ are the partial derivatives of the surface height with respect to x and y respectively at the point q , both assumed constant within the patch.

The orientation angles of the surface normal \hat{n} in the global frame are (Θ_n, Φ_n) , given at q by

$$\Theta_{nq} = \cos^{-1} \left(\frac{1}{\sqrt{1 + (z_x)_q^2 + (z_y)_q^2}} \right), \quad \Phi_{nq} = \cos^{-1} \left(\frac{-(z_x)_q}{\sqrt{(z_x)_q^2 + (z_y)_q^2}} \right) \quad (5)$$

A local coordinate system (x'_q, y'_q, z'_q) is constructed such that the z'_q axis is in the direction of the surface normal \hat{n}_q as given by (4) and the x'_q axis is along the projection of the incident wave’s propagation vector \mathbf{k}_i (from the transmitter at point 1 to q) onto the $(x'_q - y'_q)$ plane (Figure 2). The plane of propagation, no longer necessarily vertical as it was in [2], is the $x'_q - z'_q$ plane. The y'_q axis is normal to the plane of propagation. Thus the unit base vectors are given by

$$\hat{z}'_q = \hat{n}_q, \hat{y}'_q = \frac{\hat{n}_q \times \hat{k}_{1q}}{|\hat{n}_q \times \hat{k}_{1q}|}, \hat{x}'_q = \hat{y}'_q \times \hat{z}'_q \quad (6)$$

The angle of incidence Θ'_i in this local primed coordinate frame is given by

$$\Theta'_i = \cos^{-1}(-\hat{k}_{1q} \cdot \hat{n}_q) \quad (7a)$$

and it is evident from (6) and (7a) that

$$\sin \Theta'_i = |\hat{n}_q \times \hat{k}_{1q}| \quad (7b)$$

To calculate the TE and TM fields of the wave immediately below the boundary at the point q , we first find the y' components of those immediately above the boundary, which are given by

$$E_{qv'}^+ = \mathbf{E}_q \cdot \hat{y}'_q, \mathbf{H}_{qv'}^+ = \mathbf{H}_q \cdot \hat{y}'_q \quad (8)$$

where \mathbf{E}_q is obtained from (1) and (2), \mathbf{H}_q from (3) and \hat{y}' from (5) and (6). The fields immediately below the boundary are given by

$$E_{qv'_q}^- = T_{01}^{TE} E_{qv'_q}^+, \quad H_{qv'_q}^- = T_{01}^{TM} H_{qv'_q}^+ \quad (9)$$

where T_{01}^{TE} and T_{01}^{TM} are respectively the transverse electric and transverse magnetic downward transmission coefficients.

The field vectors at the point q immediately below the boundary act as sources of the wave propagating downward toward the subsurface point p . As in Eqs. (4) and (4-a, b, c) of [2], we use the Stratton Chu integral including near field terms to evaluate the fields at p , using the fields $\mathbf{E}_q^-, \mathbf{H}_q^-$ as the source fields on the bottom surface of the patch. The latter fields were given in the global coordinate frame by Eqs. (3-a, b) of [2], but are specialized there to the case of a flat horizontal patch. For the tilted patch, these equations are more complicated. In the primed (i.e. tilted) coordinated frame, they are

$$\begin{aligned} \mathbf{E}_q^- &= \hat{x}'_q \left[-Z_1 H_{qv'_q}^- \cos \Theta'_t \right] \\ &+ \hat{y}'_q \left[\mathbf{E}_{qv'_q}^- \right] + \hat{z}'_q \left[-(Z_0/\epsilon_{1R}) H_{qv'_q}^- \sin \Theta'_i \right] \end{aligned} \quad (10-a)$$

$$\begin{aligned} \mathbf{H}_q^- &= \hat{x}'_q \left[Y_1 E_{qv'_q}^- \cos \Theta'_t \right] \\ &+ \hat{y}'_q \left[H_{qv'_q}^- \right] + \hat{z}'_q \left[Y_0 E_{qv'_q}^- \sin \Theta'_i \right] \end{aligned} \quad (10-b)$$

where

$$\cos \Theta'_t = \sqrt{1 - \frac{\sin^2 \Theta'_i}{\nu_1^2}} = \frac{1}{\sqrt{\epsilon_{1R}}} \sqrt{\epsilon_{1R} - \sin^2 \Theta'_i}$$

where Z_1 and Y_1 are respectively the complex wave impedance and admittance of the medium, \hat{x}'_q , \hat{y}'_q , \hat{z}'_q are the unit base vectors along the tilted coordinate axes, Θ'_i is the angle of incidence of the transmitted wave at q , relative to the tilted plane of the patch, ν_1 is the complex refractive index, equivalent to the square root of ϵ_{1R} , the complex relative permittivity. Z_0 and Y_0 are wave impedance and admittance of free space, where $Z_1 = Z_0/\sqrt{\epsilon_{1R}}$ and $Y_1 = Y_0\sqrt{\epsilon_{1R}}$.

To determine \mathbf{E}_q^- and \mathbf{H}_q^- in the global coordinates from (10-a, b), a rotational transformation is required, i.e. in vector-matrix notation

$$[\mathbf{E}_q^-] = [B_q] [\mathbf{E}_q^-]' , [\mathbf{H}_q^-] = [B_q^-] [\mathbf{H}_q^-]' \quad (11-a)$$

where $[\mathbf{E}_q^-]$ and $[\mathbf{E}_q^-]'$ are the vectors of the electric field components in global and local coordinates respectively and $[\mathbf{H}_q^-]$ and $[\mathbf{H}_q^-]'$ have analogous meanings for magnetic fields and where

$$(B_q)_{jk} = \hat{\mathbf{u}}'_{qk} \cdot \hat{\mathbf{u}}_j \quad (11-b)$$

where $\hat{u}_{1,2,3} = \hat{x}, \hat{y}, \hat{z}$, $u'_{q1,2,3} = \hat{x}'_q, \hat{y}'_q, \hat{z}'_q$

and where the matrix elements, i.e. the scalar products $\hat{x}'_q \cdot \hat{x}$, $\hat{x}'_q \cdot \hat{y}$, $\hat{y}'_q \cdot \hat{z}$ etc. are given in the appendix in terms of (Θ_{nq}, ϕ_{nq}) , the orientation angles of the surface normal and the incidence angles (Θ_i, ϕ_i) in the global coordinates.

After the fields \mathbf{E}_q^- , \mathbf{H}_q^- have been evaluated in global coordinates, the fields at p are computed in global coordinates using Eqs. (4) and (4-a, b) of [2], which are not changed by our generalization and will not be repeated here. That step, when carried out in a loop over the entire grid of subsurface volume elements, constitutes the “forward path” computation, i.e. that of the illuminating fields within the entire subsurface region of interest.

These illuminating fields at each subsurface point are the incident wavefields for scattering from that point. We now focus attention to the “return path” computation, i.e. that of the received signal voltages due to the superposition of scattered wavefields from all of the subsurface points.

The first phase of that computation is to evaluate the wavefields at the surface point q' due to scattering from p . This is done in global coordinates in our code. The analysis behind the algorithm is based on Eqs. (5) and (6) of [2], which are not changed by our generalization and are not repeated here. The results are $\mathbf{E}_{q'}^-$, $\mathbf{H}_{q'}^-$, the wavefields at the surface point q' immediately below the surface due to scattering from p in the global frame. Again using the tangent plane approximation as in the forward path calculation, we must first find the components of these fields normal to the propagation plane in order to use TE and TM transmission coefficients in evaluating the fields immediately above the surface at q' .

This is accomplished through the 3×3 rotation matrix $A_{q'}$, the inverse of $B_{q'}$ given in (11-a, b) for point q , except that it applies to a different primed coordinate frame, that applying to the point q' , where the local slopes are different than those at q . The unit normal \hat{n} as given by (4), again points into free-space and is in the $z'_{q'}$ direction in this new coordinate system. The $x'_{q'}$ direction is along that of the projection of the propagation vector \mathbf{k}_s (from q' toward the receiver at point 2) onto the $x'_{q'} - y'_{q'}$ plane (Figure 3). The result in vector-matrix notation is

$$[\mathbf{E}_{q'}^-]' = [A_{q'}] [\mathbf{E}_{q'}^-] \quad (12-a)$$

where

$$(A_{q'})_{jk} = \hat{u}'_{q'j} \cdot \hat{u}_k \quad (12-b)$$

and where $[\mathbf{E}_{q'}^-]'$, $[\mathbf{E}_{q'}^-]$, $\hat{u}'_{q'j}$, and \hat{u}_k have the same meanings as in (11-a, b), except that the specific evaluation of the quantities like $\hat{\mathbf{x}}'_{q'}$, $\hat{\mathbf{y}}'$, $\hat{\mathbf{z}}'_{q'}$, $\hat{\mathbf{x}}$, etc. as indicated in the appendix, are now given in terms of (Θ_s, ϕ_s) , the global direction angles of the wave emerging from q' and propagating toward the receiver rather than (Θ_i, ϕ_i) , the global incidence angles of the wave impinging on q from the transmitter.

After $[\mathbf{E}_{q'}^-]$ and $[\mathbf{H}_{q'}^-]$ have been determined, the next computation is of the y' -components of the fields immediately above the boundary at q' , given by

$$E_{y'}^+(\mathbf{r}_{q'}) = T_{10}^{TE} E_{y'}^-(\mathbf{r}_{q'}) , \quad H_{y'}^+(\mathbf{r}_{q'}) = T_{10}^{TM} H_{y'}^-(\mathbf{r}_{q'}) \quad (13)$$

where T_{10}^{TE} and T_{10}^{TM} are the upward transmission coefficients for TE and TM modes respectively. Next, we calculate the fields immediately above the surface at point q' , given by

$$\mathbf{E}_{q'}^+ = \hat{x}'_{q'} \left[Z_0 H_{q'y'q'}^+ \cos \Theta'_t \right] + \hat{y}'_{q'} \left[E_{q'y'q'}^+ \right] \quad (14-a)$$

$$\begin{aligned} & + \hat{z}'_{q'} \left[-Z_0 H_{q'y'q'}^+ \sin \Theta'_t \right] \\ \mathbf{H}_{q'}^+ & = \hat{x}'_{q'} \left[-Y_0 E_{q'y'q'}^+ \cos \Theta'_t \right] + \hat{y}'_{q'} \left[H_{q'y'q'}^+ \right] \\ & + \hat{z}'_{q'} \left[Y_0 E_{q'y'q'}^+ \sin \Theta'_t \right] \end{aligned} \quad (14-b)$$

where Θ'_t is the spherical polar angle (in the primed coordinated frame) of the propagation vector \mathbf{k}_s for the refracted ray propagating from q' to the receiver at point 2, related to Θ'_i , the incidence angle of q' , the wave propagating upward from p to q' in the primed system, through the expressions $\sin \Theta'_t = \nu_1 \sin \Theta'_i$, $\cos \Theta'_t = \sqrt{1 - \epsilon_{1R} \sin^2 \Theta'_i}$,

The next step is to determine the fields given in Eqs. (14-a, b) in the global frame, given by

$$[\mathbf{E}_{q'}^+] = [B_{q'}] [\mathbf{E}_{q'}^+]', \quad [\mathbf{H}_{q'}^+] = [B_{q'}] [\mathbf{H}_{q'}^+]' \quad (15)$$

where the meanings of all quantities in (15) are analogous to those of their counterparts in (11).

The fields $[\mathbf{E}_{q'}^+]$ and $[\mathbf{H}_{q'}^+]$ are now the source fields for another application of Stratton Chu, this time for upward propagation toward the receiver in free space. The equations that describe this (not shown here) are essentially Eqs. (4) and (4-a, b, c) of [2] except that the surface normal is upward rather than downward, resulting in a sign reversal in all terms, the constitutive parameters are those of free space rather than the soil medium and subscripts q and p are replaced by q' and 2 in all quantities containing those subscripts.

When the wave propagated from the upper surface of the patch centered at q' reaches the receiver at point 2, it undergoes a transformation to the coordinates of the antenna aperture and is weighted by the antenna's receptivity pattern and its peak effective aperture area.

An issue not yet covered in this discussion is the integration over the surface of the patch centered at q on the forward path and that over the patch centered at q' on the return path. For the flat horizontal surface where the patches are rectangular and oriented along the global cartesian coordinate axes, the most accurate method is to assume that the integrand is constant over each patch except for the local phase factor $\exp(jk_0[\alpha_{xa}(x - x_q) + \alpha_{ya}(y - y_q)])$ for the forward path and $\exp(jk_0[\alpha_{xb}(x - x_{q'}) + \alpha_{yb}(y - y_{q'})])$ for the return path, where

$$\alpha_{xa} = \beta_{xa} + \gamma_{xa}, \alpha_{ya} = \beta_{ya} + \gamma_{ya}, \alpha_{za} = \beta_{za} + \gamma_{za},$$

$$\alpha_{xb} = \beta_{xb} + \gamma_{xb}, \alpha_{yb} = \beta_{yb} + \gamma_{yb}, \alpha_{zb} = \beta_{zb} + \gamma_{zb}$$

$$\beta_{xa} = \sin \Theta_{q1} \cos \phi_{q1} + \nu_1 \sin \Theta_{qp} \cos \phi_{qp}, \beta_{ya} = \sin \Theta_{q1} \sin \phi_{q1} + \nu_1 \sin \Theta_{qp} \sin \phi_{qp}$$

$$\beta_{xb} = \sin \Theta_{q'2} \cos \phi_{q'2} + \nu_1 \sin \Theta_{q'p} \cos \phi_{q'p}, \beta_{yb} = \sin \Theta_{q'2} \sin \phi_{q'2} + \nu_1 \sin \Theta_{q'p} \sin \phi_{q'p}$$

$$\beta_{za} = \cos \Theta_{q1} + \nu_1 \cos \Theta_{qp}, \beta_{zb} = \cos \Theta_{q'2} + \nu_1 \cos \Theta_{q'p}$$

$$\gamma_{xa} = \beta_{za}(z_x)_q, \gamma_{ya} = \beta_{za}(z_y)_q, \gamma_{xb} = \beta_{zb}(z_x)_{q'}, \gamma_{yb} = \beta_{zb}(z_y)_{q'}$$

If the x and y dimensions of all patches are all the same and denoted by a_x and a_y , and the factor involving the imaginary parts of the complex refractive index can be assumed constant over the patch, then the integral over the patch on the forward path is approximately

$$I_a = a_x a_y \sin c(\pi a_x \text{Re}(\alpha_{xa})/\lambda_0) \sin c(\pi a_y \text{Re}(\alpha_{ya})/\lambda_0) \quad (16)$$

with an analogous expression for the return path where subscript a is replaced by b .

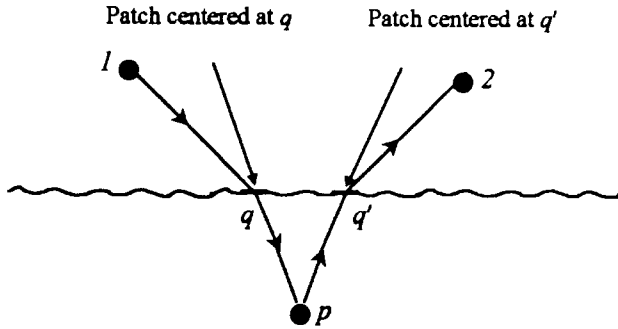
If this integration method is chosen, then in principle the integral must be evaluated for every patch on the surface in order to find the field at any subsurface point on the forward path or from such a point on the return path. That process is prohibitively CPU-time intensive. Alternatively, we can choose an angular region well within the first lobe of the sinc functions and integrate only over the patches within that region. That method saves considerable time but the decision on the limits of the process is a difficult one. The method we finally decided upon was to use the stationary phase ("sp") approximation and integrate only over sp points. For a flat surface, there is only one sp point for each subsurface cell. For a rough surface, there may be many sp points, but it is still faster to find the sp points and perform the integration only at those points than to use the method discussed above. The result for the forward path is

$$I_a = -\frac{j\pi}{k_0} \left[\frac{r_{1q} r_{qp}}{(r_{qp} + \nu_{1R} r_{1q}) |\cos \Theta'_i|} \right] \quad (17)$$

where ν_{1R} is the real part of ν_1 , with an analogous expression for the return path where subscript a is replaced by b .

The result (17) and its counterpart for the return path show reductions in the $1/R$ spreading factors for the illumination of a given subsurface point p and the return from that point, relative to the case where the integral is carried out for only a few patches centered around the angles at the peaks of the sinc functions in (16), as was done to obtain the results shown in [2]. Those spreading factors would be reduced if we integrated over many more patches, but since as indicated earlier that is prohibitively time-intensive, a better approximation should be attainable with stationary phase. The effects of $1/R^2$ and $1/R^3$ terms and longitudinal wavefields, all due to near field interactions, which were a major theme in [2], are still present in results obtained using (17).

(a) Small-scale roughness only



b) Large and small-scale roughness

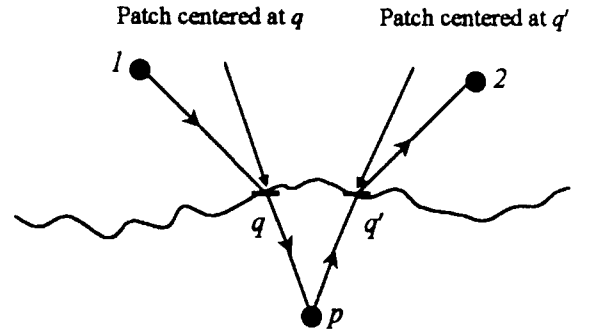


Figure 1: Illustration of Problem Geometry with Rough Surfaces.

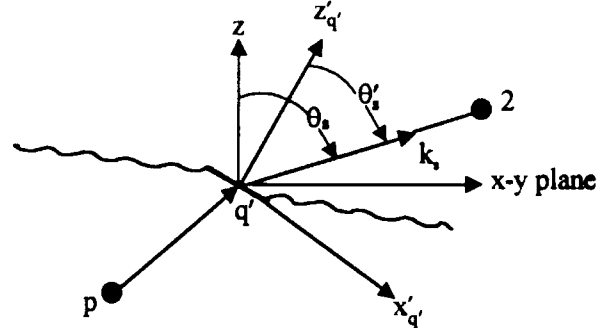
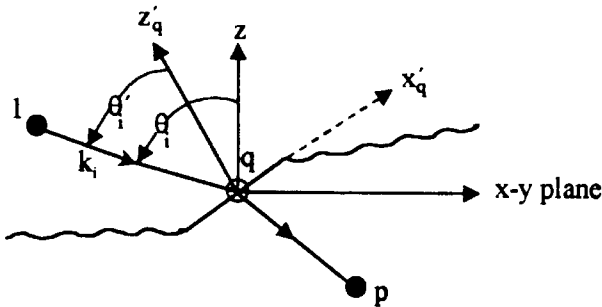


Figure 2: Coordinate System for Forward Path. Figure 3: Coordinate System for Return Path.

3 Results and Conclusions

The simulation of the returns from a subsurface volume containing mines, aggregates of rocks and random variations of complex soil permittivity, was based on the code whose background theory was described in Section 2. As in [1] and [2], the code was run thousands of times in order to simulate the action of the focused array system described in [1] and in more detail in [3]. Referring

back to Section 3 of [1], there are 4 transmitters and 4 receivers, resulting in 16 transmitter-receiver positions. The grid of small volume elements making up the illuminated underground region ($32''$ in crosstrack (“ x ”) direction, $10''$ in alongtrack (“ y ”) direction and $8''$ in depth (“ z ”) consisted of 3072 such elements, each one a cube with x , y and z dimensions of $1''$. Thus the propagation path $1 \rightarrow q \rightarrow p \rightarrow q' \rightarrow 2$ and associated scattering processes described in Section 2, were simulated $16 \times 3072 = 49,152$ times, resulting in 3D image plots as shown in Figures 4 through 9. To obtain these plots we assume a single mine and two rocks in a clutter background due to random soil permittivity inhomogeneties. The burial depth for the mine is $2''$ and those for the rocks are $2.5''$ and $1.75''$. The mine is placed at $x = 0$ and the rocks at $x = 8''$ and $+6''$. The rms deviation of soil permittivity from its mean value is set at 3 percent of the mean. The large-scale surface roughness is modeled through one of the options available in the overall simulation framework [4]. The selected option is based on a two dimensional Fourier series model, where the coefficients can have deterministic and/or random components. To produce these illustrative plots, we have specialized the model to a single sinusoidal function

$$z(x, y) = C \cos \left(2\pi \left[\left(\frac{x}{X_0} \right) + \left(\frac{y}{Y_0} \right) + \Psi \right] \right) \quad (18)$$

and vary the amplitude C from 0 to $0.4''$.

The slopes $(z_x)_q$ and $(z_y)_q$ defined in (4) are:

$$(z_x)_q = \frac{-2\pi}{X_0} C \sin \xi_q, \quad (z_y)_q = -\frac{2\pi}{Y_0} C \sin \xi_q, \quad (19)$$

where

$$\xi_q = 2\pi \left[\frac{x_q}{X_0} + \frac{y_q}{Y_0} + \Psi \right]$$

The slope angles (Θ_{nq}, ϕ_{nq}) can be determined from (5) and (19).

The results presented in this paper are merely illustrative, designed to show that large-scale roughness can have a significant effect on the simulation results under certain conditions. In all of the plots in Figures 4-9, the soil is clay loam (complex permittivity $6.2 - j0.145$ at 1 GHz). The mine is an upright finite circular cylinder of radius $1.5''$ and height $2''$ with permittivity $3 - j0.03$ and the rocks are spheres of radii $.75''$ and $1''$ and permittivity $4.5 - j0.003$ and $4.0 - j0.003$. In Figures 4 and 5, there is no soil permittivity fluctuation clutter. In Figure 4, the air-ground interface is a flat horizontal surface. In Figure 5, the maximum spread of surface height is $0.8''$. There are small differences in the shapes of mine and rock signals on these two plots, but they are not significant. This is because the differences between true and indicated positions are present for a very small number of subsurface scattering points, i.e. those within the volume occupied by the mine and rocks. We conclude that surface height spreads below $0.8''$ have very little effect in the absence of permittivity fluctuation clutter. When we introduce 3% fluctuation clutter, the effect of surface roughness becomes more pronounced, as shown in Figures 6-9. In Figure 6, the interface is horizontal. The mine shows up in the same place as in the zero clutter plots, but the

rock signals are obscured by the clutter. In Figure 7, where the surface height spread is 0.4", the mine is still visible in the clutter and its location is unchanged, but there are many more peaks comparable in intensity with the mine, which could be interpreted as due to discrete objects such as mines or rocks. However, aside from a few peaks that may be due to rocks, nearly all of them would have to be due to permittivity fluctuations that permeate the entire volume. These effects are still more pronounced in Figures 8 and 9, where the clutter level is still 3%, but surface height spreads are 0.6" and 0.8" respectively. In both cases there are so many peaks in various locations that it would be very difficult to infer the presence and location of a mine without further signal processing beyond that inherent in the focused array timing algorithm. An explanation for these effects in the simulation is the possible existence of many stationary phase points for any given subsurface scattering point. Hence some subsurface points would receive (and scatter) much more illumination than predicted by analysis based on a flat horizontal interface and others would receive (and scatter) much less. This should result in more fluctuation in the clutter return than in the absence of surface roughness, as observed in Figures 7, 8 and 9.

4 Continuing Work

Accounting for large-scale surface roughness, the primary focus of this paper, is only one of the generalizations we have implemented in order to improve the accuracy of the modeling. Work is still in progress not only on that issue but also on inclusion of second order scattering effects [5, 6] on the subsurface illuminating fields, but the complete second order theory is prohibitively CPU-time intensive. Even reflections of first-order scattered fields from the underside of the air-ground interface, the simplest of the second order effects to be considered, must be limited to the strongest reflections in order to avoid excessive running time. The decision on which reflections to include is much more difficult with a rough surface, so the current study of this issue is being limited to flat surfaces. The next step is development of a criterion to decide which reflections can be excluded without much of a penalty when the roughness of the interface is accounted for.

Figure 4: no clutter, horizontal interface

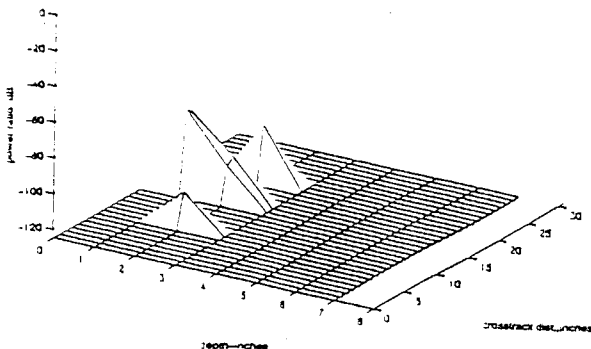


Figure 5: no clutter, interface with .1" spread in height

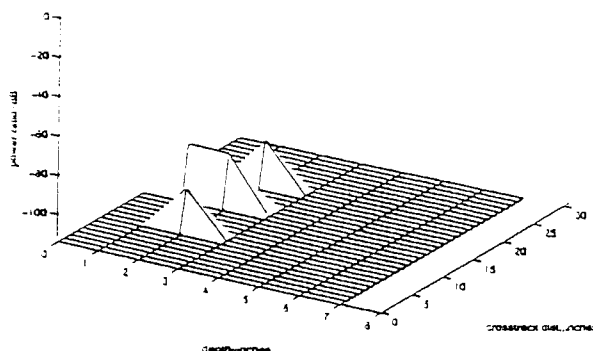


Figure 6: 0% cluster, horizontal interface

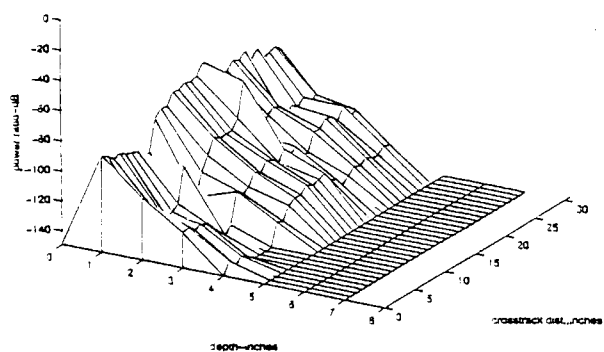


Figure 7: 0% cluster, interface with .4" spread in height

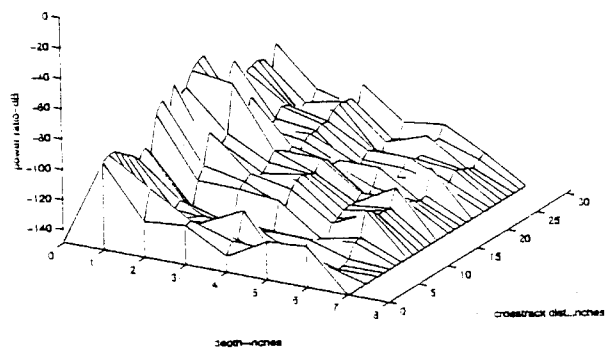


Figure 8: 0% cluster, interface with .6" spread in height

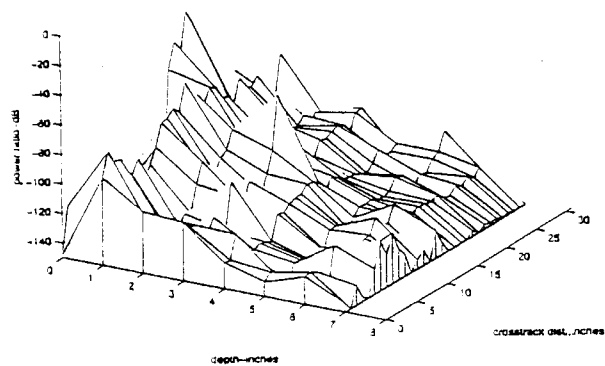
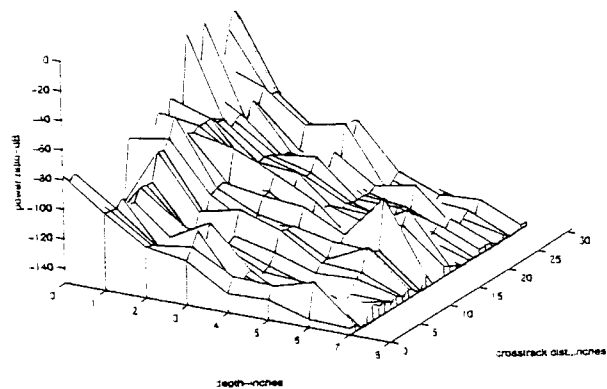


Figure 9: 0% cluster, interface with .8" spread in height



References

- [1] Raemer, H. R., C. M. Rappaport and E. L. Miller, "Frequency Domain Simulation of Focused Array Radar Returns from Buried Mines in Clutter", SPIE Volume 3392, Orlando, FL, April 1998, pp 754-765.
- [2] Raemer, H. R., C. M. Rappaport and E. L. Miller, "Near-Field and Timing Effects in Simulation of Focused Array Radar Signals from a Mine in Subsurface Clutter", SPIE, Volume 3392, Orlando, FL, April 1999.
- [3] Rappaport, C. M. and D. M. Reidy, "Focused Array Radar for Real Time Imaging and Detection", SPIE Volume 2747, Orlando, FL, April 1996, pp 202-213.
- [4] Raemer, H. R., "Analytical Modeling and Computer Simulation of Radar Clutter and Multipath Signals", monograph and project report, Center for Electromagnetics Research, Northeastern University, Boston, MA, Volume 1, May 1989, revised version, 1992, Volume 2, 1994 (chapter contributed by A. K. Bhattacharya), Appendix 3, 1996.
- [5] Ulaby, F. T., R. K. Moore and A. K. Fung, "Microwave Remote Sensing", Volume III, Artech 1986, Chapter 13, pp 1065-1205.
- [6] Tsang, L., J. A. Kong and R. T. Shin, "Theory of Microwave Remote Sensing", Wiley, New York, 1985, Chapter 2, Section 3, pp 32-41 and Chapter 5, Section 2, pp 319-337.

5 Appendix

A_q and B_q matrix elements

$$\begin{aligned}
 (A_q)_{31} &= (B_q)_{13} = s_{tnq} c_{pnq} = a_{31}, \quad (A_q)_{32} = (B_q)_{23} = s_{tnq} s_{pnq} = a_{32} \\
 (A_q)_{33} &= (B_q)_{33} = c_{tnq} = a_{33}, \quad (A_q)_{21} = (B_q)_{12} = K_q (s_{tnq} c_{tiq} s_{pnq} - s_{tiq} c_{tnq} s_{piq}) = a_{21} \\
 (A_q)_{22} &= (B_q)_{22} = K_q (c_{tnq} s_{tiq} c_{piq} - s_{tnq} c_{tiq} c_{pnq}) = a_{22} \\
 (A_q)_{23} &= (B_q)_{32} = K_q s_{tnq} s_{tiq} (s_{piq} c_{pnq} - c_{piq} s_{pnq}) = a_{23}
 \end{aligned}$$

where

$$\begin{aligned}
 s_{tnq} &= \sin \Theta_{nq}, \quad c_{tnq} = \cos \Theta_{nq}, \quad s_{pnq} = \sin \phi_{nq}, \quad c_{pnq} = \cos \phi_{nq} \\
 s_{tiq} &= \sin \Theta_{iq}, \quad c_{tiq} = \cos \Theta_{iq}, \quad s_{piq} = \sin \phi_{iq}, \quad c_{piq} = \cos \phi_{iq} \\
 K_q &= 1 / \sin \Theta'_{iq} \\
 (A_q)_{11} &= (B_q)_{11} = a_{22} a_{33} - a_{23} a_{32}, \quad (A_q)_{12} = (B_q)_{21} = a_{23} a_{31} - a_{21} a_{33} \\
 (A_q)_{13} &= (B_q)_{31} = a_{21} a_{32} - a_{22} a_{31}
 \end{aligned}$$

and where the subscript q on an angle implies that the angle is evaluated at the surface point q . The matrix B_q in (11-b) is evaluated at q , while $A_{q'}$ given in (12-b) is evaluated at q'

GPR Application of the Conical Spiral Antenna Probe

by

Harold Raemer and Carey Rappaport

Department of Electrical and Computer Engineering
Center for Subsurface Sensing and Imaging Systems
Northeastern University, Boston, MA 02115, U.S.A
Harold Raemer: 617-373-3035, raemer@ece.neu.edu
Carey Rappaport: 617-373-2043, rappaport@neu.edu
FAX: 617-373-8627 (for both authors).

Presented at SPIE Aerosense 2001, Orlando, FL, April, 2001.

Abstract

The project motivating this paper is the deployment of a frequency independent antenna on the transceiver of a monostatic ground-penetrating radar used to detect mines. The design goal is that the radiation pattern and input impedance be nearly uniform over a band from 1 GHz to 5GHz if the antenna is partially immersed in a typical soil medium. The contemplated method of deployment is to have the antenna straddle the air-soil interface i.e. partly in free space and partly underground, radiating into the ground. The particular subclass of frequency-independent antenna under investigation for this application is the conical equiangular-spiral antenna, in which thin wires are wound around a conical frame and the radiation is from the apex and reaches its peak in the axial direction. The conical structure, about 50cm long and with a maximum diameter of 12cm, is thrust into the ground apex-first at an angle of about 70° to the vertical.

The subject of this paper is an analysis of this antenna, first solving the integral equation to find the currents on the wire, then determining the radiation pattern at frequencies between 500MHz and 5GHz. The theory of this antenna in infinite free space was developed by K. K. Mei and others in the 1960s (IEEE Transactions, AP-13, 5/65, pp. 374-378 and Yeh and Mei, AP-15, 5, 9/67, pp 634-639 and AP-16, 1, 1/68, pp. 9-21.) The analysis reported in the present paper is based on these sources. An ultimate objective of the reported work is to determine theoretically the degree of frequency uniformity attainable when the antenna geometry is as indicated above. The ideal "frequency independent" property is theoretically possible if the

cone is infinitely long, so the finite length and width of the frustum imposes limits on the attainable bandwidth. In order to begin to determine the feasibility of this antenna in the contemplated application, this paper will be limited to the analysis of the antenna positioned with its axis in the vertical direction and immersed in an infinite lossy dielectric medium (e.g. clay loam soil). The extension to include the effect of the air-soil interface and the partial immersion of the antenna in the soil at an angle other than vertical will be given in a later paper.

KEYWORDS: mine detection, antennas, ground-penetrating radar, frequency-independent antennas, conical spiral antenna.

1 Introduction

The conical spiral antenna, a subclass of "frequency independent antennas," was studied extensively by a number of researchers in the 1960's [1] - [7] and has been discussed in some more recent antenna texts [8] - [10]. Because of the very large bandwidths attainable with these antennas and appropriate form-factor as a probe, we are investigating their use in anti-personnel land-mine detection by ground-penetrating radar. Ideally, in that application, the antenna's bandwidth should be comparable to that of the transmitted pulses used in the many GPR systems whose durations are of the order of nanoseconds. A conical spiral can achieve that kind of bandwidth in free space and might be able to do so in soil media in which mines are typically buried. It also produces radiation patterns that peak in the axial direction, and although not very directive, focus nearly all of the radiated energy into the hemisphere outside the cone. In the contemplated application, the tip of the cone is manually thrust into the ground at selected intervals at an angle of about 70° from the vertical as illustrated in Figure 1. Each thrust of this probe will illuminate a subsurface volume possibly containing a mine. It is expected that such a hand-held probe would supplement other detectors as a confirmatory sensor.

An obvious issue in the use of a GPR probe configured as described above is that of the operation of the conical spiral in a typical soil, especially a soil with significant lossiness. Another issue is the effect of the air-ground interface on the current distribution, input impedance and radiation pattern, particularly if the interface has significant roughness.

The nominal geometry assumes a circular wire of constant radius wound around a cone 50cm long and with a maximum diameter of 12cm, with cone angle $\theta_0 = 13^\circ$. This antenna has two arms, the first of which is governed by the equation $r = r_0 e^{a\phi}$ and the second by $r = r_0 e^{a(\phi-\tau)}$.

This paper reports on the first part of a particular task in this investigation, which is the generation of a computer simulation to determine the current distribution, near-field pattern and input impedance of this antenna when immersed in a soil medium modeled as a lossy dielectric. In the above-mentioned theoretical and experimental work done in the 1960's, the antenna was assumed to be in infinite free space. Although the early results provide precise information about the behavior of the antenna in free space, and qualitative insights into its expected properties in a dielectric medium, its behavior in a lossy soil cannot easily be quantitatively inferred from them. For this reason the analysis must be repeated for those cases, using the enormous improvements in

computer power during the past thirty years to produce a wider range of results that will aid us in the development of the GPR probe.

The ultimate goal in this simulation phase of the project is to include the effects of a possibly rough air-ground interface and a buried mine on a conical spiral with arbitrary orientation. That level of generality is required to completely evaluate the configuration we have in mind. But the matlab simulation code developed to-date is limited to the case where the antenna is immersed in an infinite lossy dielectric, a theoretically simple extension of the work of K. K. Mei in 1965 [1] and Yeh and Mei in 1967 and 1968 [2, 3].

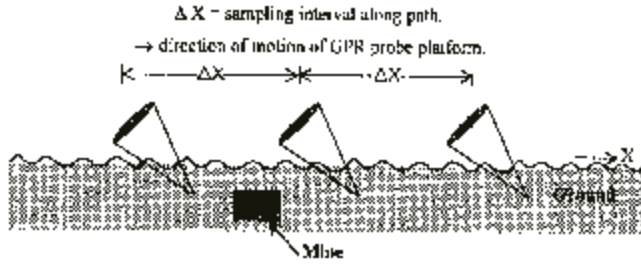


Figure 1: Deployment of conical spiral as GPR probe

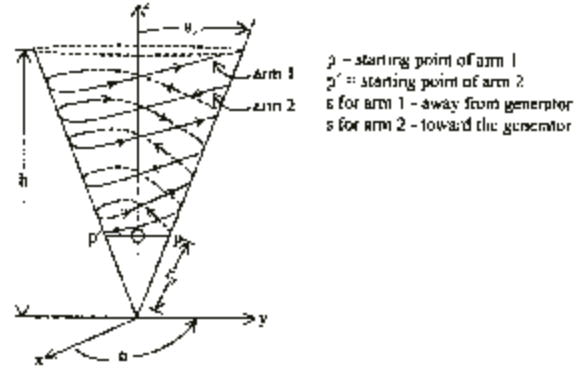


Figure 2: Basic geometry of conical spiral antenna.

2 Theoretical Background

The axis of the cone is along the z -axis of a cartesian coordinate frame whose origin is at the apex, as shown in Figure 2. The spherical coordinates of an infinitesimal current element on the wire wound around the surface of the cone are (r', θ', ϕ') , where

$$r' = r_0 e^{a\phi'}, \theta' = \theta_0 \quad (1)$$

where a is the “spiral constant”, a monotonically increasing function of the pitch angle, θ_0 is the cone angle, assumed constant, and r_0 , is the straight-line distance on the cone surface from the origin to the point p shown in Figure 2. The vector from the origin to the center of the element, is

$$\mathbf{r}' = r_0 e^{a\phi'} [\sin \theta_0 (\hat{\mathbf{x}} \cos \phi' + \hat{\mathbf{y}} \sin \phi') + \hat{\mathbf{z}} \cos \theta_0] \quad (2)$$

The incremental change in \mathbf{r}' as ϕ' varies along the wire is

$$d\mathbf{r}' = [a\mathbf{r}' - r_0 e^{a\phi'} \sin \theta_0 (\hat{\mathbf{x}} \sin \phi' - \hat{\mathbf{y}} \cos \phi')] d\phi' \quad (3)$$

The magnitude of $d\mathbf{r}$ is (with the aid of (1))

$$\begin{aligned}
|d\mathbf{r}'| &= a r_0 e^{a\phi'} d\phi' \sqrt{\sin^2 \theta_0 \left[\left(\cos \phi' - \frac{\sin \phi'}{a} \right)^2 + \left(\sin \phi' + \frac{\cos \phi'}{a} \right)^2 \right] + \cos^2 \theta_0} \\
&= Q dr' , \text{ where } Q = \sqrt{1 + \frac{\sin^2 \theta_0}{a^2}}
\end{aligned} \tag{4}$$

The quantity Q , defined in [2] as the "slowness factor", is a measure of the ratio of the curvilinear distance traversed along the wire to the linear distance r' along the cone surface.

A unit vector $\hat{\mathbf{s}}'$ in the direction of the current element at the point \mathbf{r}' , proportional to $d\mathbf{r}'$ in (3), is

$$\hat{\mathbf{s}}' = \frac{1}{Q} \left\{ \sin \theta_0 \left[\hat{\mathbf{x}} \left(\cos \phi' - \frac{\sin \phi'}{a} \right) + \hat{\mathbf{y}} \left(\sin \phi' + \frac{\cos \phi'}{a} \right) \right] + \hat{\mathbf{z}} \cos \theta_0 \right\} \tag{5}$$

The unprimed version of (5) is corroborated by Eq. (16) of [2]. The distance traversed along the wire, proportional to $|d\mathbf{r}'|$ as given in (4), is

$$s' = Q r' \tag{6}$$

Eqs. (1) - (6) apply to arm 1, which starts at the point p in Figure 2. Arm 2, wound in the same sense as arm 1, starts at point p' , separated from p by 180° . The counterpart of (1) for arm 2 is

$$r'_2 = r_0 e^{a(\phi' - \pi)} \tag{7}$$

If ϕ' is replaced by $\phi' - \pi$ in (2) and the steps from (2) through (6) are repeated, we can obtain the counterparts of (5) and (6) for arm 2, which are given in [2].

To determine the electric field at an arbitrary point designated by a vector \mathbf{r} due to currents along arm 1, we note that the current vector on an element at \mathbf{r}' is denoted by $I(s') \hat{\mathbf{s}}'$. The magnetic vector potential at \mathbf{r} due to this element is

$$\mathbf{A}(\mathbf{r}) = \mu I(s') \hat{\mathbf{s}}' G(\mathbf{r}, \mathbf{r}') ds' \tag{8}$$

where $G(\mathbf{r}, \mathbf{r}')$ is the green's function in the surrounding medium, given by $G(\mathbf{r}, \mathbf{r}') = e^{-jkR}/4\pi R$, and $R = |\mathbf{r} - \mathbf{r}'|$, and where μ and k are the magnetic permeability and propagation constant respectively of that medium. The electric field at \mathbf{r} , as obtained from (8), is

$$\mathbf{E}(\mathbf{r}) = -j\omega \left(\bar{\mathbf{I}} + \frac{1}{k^2} \nabla \nabla \right) \cdot \left[\int_0^L ds' \mu I(s') \hat{\mathbf{s}}' G(\mathbf{r}, \mathbf{r}') \right] \tag{9}$$

where ω is the angular frequency and L is the length of arm 1.

Eq. (9) (or its equivalent in other forms), is the basis of the integral equation from which the currents $I(s')$ are to be evaluated by MOM (Section 2.1) and also the calculation of the field pattern in the surrounding space (Section 2.1). To obtain the field due to both arms, a second term is added

to the right hand side, identical to the expression in (9) but with subscripts 2 on s, \hat{s}', \mathbf{r}' and also on L if arms 1 and 2 have different lengths.

The radius of the wire, denoted by b , is assumed constant in this analysis. That case is discussed on Page 638 of [2]. It is not the theoretically ideal case, in which the antenna is infinitely long and the radius continuously increases along the wire. The prescribed increase in wire radius is feasible in principle but more difficult to implement in practice and obviously a practical antenna must have a finite length. For these reasons, the antenna being investigated here is not truly frequency independent. But it is expected that the constant wire radius version will have a sufficiently large bandwidth for the contemplated application.

2.1 Current Distribution

To obtain the currents $I(s')$, an electric field integral equation (EFIE) is derived from (9) or its equivalent [1]. The point \mathbf{r} is set on arm 1 of the wire. The $r, \mathbf{r}, \hat{\mathbf{s}}$ and s' are given by expressions identical to (1), (2), (5) and (6) respectively except that the primes are removed from all quantities in those expressions. The tangential component of $\mathbf{E}(\mathbf{r})$, equal to $\hat{\mathbf{s}} \cdot \mathbf{E}(\mathbf{r})$, is integrated over the length of arm 1.

The derivation of the EFIE for an arbitrary thin wire antenna is given by Mei in [1] and further specialized to the conical spiral antenna by Yeh and Mei in [2]. It is based on Eq. (11) of [1], which can be derived from (9) in the present paper. The resulting kernel, as given by Eq. (28) of [1] and more specifically by Eq. (12) of [2], is

$$\pi(s, s') = \pi_a(s, s') + \pi_b(s, s') + \pi_c(s, s') \quad (10)$$

where $\pi_a(s, s') = G(s, s') (\hat{\mathbf{s}} \cdot \hat{\mathbf{s}}')$, $G(s, s')$ being the same as $G(\mathbf{r}, \mathbf{r}')$ in (8) but the Green's function is expressed as a function of s and s' ;

$$\pi_b(s, s') = -P \int_0^L dt \left[\frac{\partial G(t, s')}{\partial t} (\hat{\mathbf{t}} \cdot \hat{\mathbf{s}}') + \frac{\partial G(t, s')}{\partial s'} + G(t, s') \left(\frac{d\hat{\mathbf{t}}}{dt} \cdot \hat{\mathbf{s}}' \right) \right] \cos(k(s-t)),$$

P denoting the principal value of the integral on t ;

$$\pi_c(s, s') = -G(s_p, s') \left[(\hat{\mathbf{s}}_p^+ \cdot \hat{\mathbf{s}}') - (\hat{\mathbf{s}}_p^- \cdot \hat{\mathbf{s}}') \right] \cos(k(s-s_p)) u(s-s_p)$$

The last term is due to the step discontinuity at the point p . In this term $\hat{\mathbf{s}}_p^+$ and $\hat{\mathbf{s}}_p^-$ refer to the unit vector $\hat{\mathbf{s}}$ at points just above and just below p respectively and $u(s-s_p)$ is the unit step function, implying that s must exceed s_p .

The counterpart of (10) for arm 2, which is added to (10) to form the complete kernel, is identical to (10) but with s' and $\hat{\mathbf{s}}'$ replaced by $s'_2 = -s'$ and $\hat{\mathbf{s}}'_2$ respectively. The two kernels are additive because $I(s') = I(-s')$, due to the symmetry of the two arms and hence the final form of the EFIE is (Eq. (12) of [2])

$$\int_0^L ds' F(s, s') I(s') = H \cos ks - (jV_0/2Z) \sin ks \quad (11)$$

where $F(s, s') = \pi(s, s') + \pi(s, -s')$, where V_0 and Z are the applied voltage and wave impedance of the medium respectively and where B on the right-hand side is uniquely specified by the requirement that the current must vanish at both ends of the wire, as in the Pocklington and Hallén integral equation ([8], pp. 278-286) for a thin straight wire antenna, of which this antenna is a generalization.

The specific forms of the expressions in (10), i.e. $\hat{s} \cdot \hat{s}'$, $\hat{t} \cdot \hat{s}'$, $\hat{s}_p^\pm \cdot \hat{s}'$, and $\frac{d\hat{t}}{dt} \cdot \hat{s}'$ can be calculated for arm 1 from Eqs. (16) and (17) of [2], where \hat{t} and t replace \hat{s} and s in some of these expressions. Their counterparts for arm 2 can be calculated from (16), (17) and (18) of [2]. These expressions were calculated and incorporated into the matlab code as arrays to form the quantities in (10).

The modeling of the Green's function $G(s, s')$ requires a decision on the inclusion of the integration around the periphery of the circular wire. Referring to literature on the straight circular wire antenna directed along the z-axis [11] in the formulation of the integral equation the current can be assumed to be entirely on the outer surface of the wire and the observation point at the center of the wire. Alternatively, the current can be along the center and the observation point along the surface. These two viewpoints give essentially equivalent results and we have chosen the former. Both viewpoints are based on the assumption that the wire is so thin that the currents flow entirely in the wire direction, i.e. currents transverse to that direction are neglected.

Actually, since the wire is assumed to be a solid perfect conductor, the current flows almost entirely on the outer surface. The true boundary condition is the vanishing of the electric field along the outer wire surface. The choice of both source and observation point on the outer surface in the formulation of the integral equation leads to obvious singularities of the kernel when $s = s'$. Although the effect of these singularities is eliminated in the integrations, as pointed out on Page 637 of [2], they can be entirely circumvented by the approximation indicated above, and for a sufficiently thin wire, should not cause serious errors in the computation.

The modeling of R , the distance between an observation point along the center of the wire and an infinitesimal current element on the outer surface of the wire and parallel to \hat{s}' at the center of the element, can be easily accomplished. The result is

$$R = |\mathbf{r} - (\mathbf{r}' + b\hat{\rho}')| = R(s, s', b, \beta) = \sqrt{A_0 + (u \cos \beta + v \sin \beta)} \quad (12)$$

where β is the local azimuthal angle around the periphery of the wire, and where

$$A_0 = \frac{1}{Q^2} \left\{ s^2 + \left(s'_1 \right)^2 - 2ss'_1 \left(\cos^2 \theta_0 + \sin^2 \theta_0 \cos \Delta \phi \right) + b^2 Q^2 \right\}$$

where b is the wire radius and $\hat{\rho}'$ is a vector normal to \hat{s}' for arm 1 and to \hat{s}'_2 for arm 2. u and v have also been calculated but are not given here.

The Green's function, accounting for integration over the local azimuth angle β , is

$$G(s, s', b) = \frac{b}{4\pi} \int_0^{2\pi} d\beta \frac{e^{-jkR(s, s', b, \beta)}}{R(s, s', b, \beta)} \simeq \frac{be^{-jkR(s, s', b)}}{2R(s, s', b)} \quad (13)$$

where R is approximated as $\sqrt{A_0}$, on the assumption that b is sufficiently small to justify neglect of the terms involving β in (12).

If the terms involving β were comparable in magnitude to A_0 , then the β integration would need to be performed numerically for each current element, which is enormously CPU-time intensive. In the cases to be considered in this study, examination of parameter values indicates that the β terms are sufficiently small compared with A_0 to justify an approximation in the exponential term. It is noted that, as $\theta_0 \rightarrow 0$ and $a \rightarrow \infty$ (i.e. as the spiral approaches a straight wire in the z -direction), $R \rightarrow \sqrt{(z - z')^2 + b^2}$, known as the "thin-wire approximation" for a z -directed linear antenna.

The terms in $\pi_b(s, s')$ involving derivatives of the Green's function can be obtained from (13). In general

$$\frac{\partial G}{\partial s} = \frac{\partial G}{\partial R} \frac{\partial R}{\partial s} = -jkG \left(1 + \frac{1}{jkR} \right) \frac{\partial R}{\partial s} \quad (14)$$

By replacing s with t in (14) and substituting the resulting expressions into $\pi_b(s, s')$ in (10), we can obtain all the quantities needed in the integrand of $\pi_b(s, s')$ for arm 1. With sign changes in these quantities, we can do the same for arm 2.

The next step in obtaining the current on the wire is to expand the current in a set of basis functions $p_n(s')$, i.e.

$$I(s') = \sum_{n=1}^N c_n p_n(s') \quad (15)$$

substitute (15) into the integral equation and apply the method of moments (MOM) to solve for the coefficients c_n . This was actually done by Yeh and Mei in [2], but their technique was not designated by them as MOM because that terminology [12] was not yet in wide use in 1967. They used the most elementary and straightforward form of MOM, that of point matching [13], and a set of basis functions that are nonzero only in the subdomain of s' that contains the matching point, e.g., a set of contiguous rectangular pulse functions. Their pulse functions were multiplied by a set of interpolation functions that were quadratic in s' , which enhanced the rate of convergence relative to flat-top pulse functions. In our work, we have initially used flat-top pulse functions but have studied more sophisticated sets of basis functions [14] with a view toward improving the accuracy of the computations in later work.

2.2 Near-Fields of Antenna

To calculate the near fields within a region around the antenna, we again invoke (9), using $I(s')$ as determined from the MOM solution of the EFIE as discussed in the previous section. To determine the θ , ϕ and r components of the electric field in the same global coordinate system defined earlier, we first find the rectangular components of fields at an arbitrary observation point from a volume distribution of arbitrarily directed infinitesimal current elements from Eqs. (6-108 a, b, c) on pp.

283-284 of [15], which are based on an expression equivalent to (9). This procedure yields a set of electric field components (E_x , E_y , E_z) from a volume integral over current elements J_x , J_y , J_z .

Then the spherical coordinate components are calculated from the relationships

$$\begin{aligned} E_\theta &= \hat{\theta} \cdot \mathbf{E} = \cos \theta (E_x \cos \phi + E_y \sin \phi) - E_z \sin \theta, \\ E_\phi &= \hat{\phi} \cdot \mathbf{E} = -E_x \sin \phi + E_y \cos \phi, \\ E_r &= \hat{r} \cdot \mathbf{E} = \sin \theta (E_x \cos \phi + E_y \sin \phi) + E_z \cos \theta \end{aligned} \quad (16)$$

The volume of integration is then defined as that of the curved wires on arms 1 and 2 of the conical spiral antenna, which (for arm 1) results in

$$E_\theta = \frac{-jZ}{4\pi kQ} \int_0^L ds' I(s') e^{-jkR} \left\{ \begin{array}{ccc} G_1 & [-B] & + \\ " & [A] & " \\ " & [-C] & " \end{array} \begin{array}{ccc} G_2 & [D] & [M_\theta] \\ " & " & [M_r] \\ " & " & [M_\phi] \end{array} \right\} \quad (17)$$

where $D = rA - r_0 e^{a\phi'}$, $M_\theta = r_0 e^{a\phi'} (bb)$, $M_r = r - r_0 e^{a\phi'} (aa)$, $M_\phi = r_0 e^{a\phi'} (cc)$ and where $aa = \cos \theta \cos \theta_0 + \sin \theta \sin \theta_0 \cos \Delta\phi$, $bb = \sin \theta \cos \theta_0 - \cos \theta \sin \theta_0 \cos \Delta\phi$, $cc = \sin \theta_0 \sin \Delta\phi$, $A = aa + \frac{cc}{a} \sin \theta$, $B = bb - \frac{cc}{a} \cos \theta$, $C = cc - \frac{\sin \theta_0 \cos \Delta\phi}{a}$, $G_1 = -[1 + jkR + (jkR)^2]/R^3$, $G_2 = [3 + 3jkR - (jkR)^2]/R^5$, $I(s') = \int_0^{2\pi} d\beta b J(s'\beta)$ and $\Delta\phi = \phi - \phi'$.

Each total field component is the superposition of two terms of the forms in (17), one for each arm. The expression for arm 2 is the same as that for arm 1 except that ϕ' is changed to $\phi' - \pi$ and hence $\Delta\phi$ is changed to $\Delta\phi + \pi$, resulting in sign changes and consequent differences in the values of R , as discussed in subsection 2.1. It is noted that, as required, E_θ and E_ϕ become proportional to $1/r$ and E_r becomes proportional to $1/r^2$ as r approaches infinity.

3 Discussion of Results

Before proceeding to parametric studies of the variation of current and field patterns with frequency and soil properties, we first embarked on an effort to validate the simulation code by comparison with past results. This resulted in the curves shown in Figure 3(a), where our computed currents were compared with those obtained by Yeh and Mei (Figure 7 on Page 638 of [2]). Figure 3(a) shows plots of real and imaginary parts and amplitude of the current in milliamperes against $x = s/\lambda$ as s varies from 1 to 6 wavelengths. For these plots, we have used the same values of input variables as those in Figure 7 of [2], i.e. $a = .035$, $r_0 = .316\lambda$, $b = .0096\lambda$, $\theta_0 = 10^\circ$, and the surrounding medium is infinite free space. The results from [2] are shown in Figure 3(b). Comparing the real and imaginary parts of the current in Figure 3(a) with those in Figure 3(b), we note fairly good agreement in the s -locations of most of the maxima and minima, except for the first peak of the real part, which shows some unstable behavior, and the last maximum and minimum of both real and imaginary parts, whose locations deviate from those of Figure 3(b). Relative amplitudes of some of

the peaks and troughs and the rate of decay of the amplitude are also close to those of Figure 3(b), however, there are enough discrepancies to warrant further analysis of the code, as indicated below. As of this writing, we are not certain of why the two sets of curves are not in exact agreement, but it is probably at least partially due to the fact that our basis functions for the MOM computation were not exactly the same as those used by Yeh and Mei (i.e. we did not use their interpolation functions, as indicated in Section 2.1). Another probable reason is the approximation used in the Green's function (See Eq. (13)) which circumvented the exact integration around the periphery of the wire, which is particularly important in the region in which s is close to s' , where the Green's function reaches very large values. Whatever the reasons for these discrepancies, further work is in progress to improve the accuracy of the kernel of the integral equation, and to try more sophisticated MOM basis functions, both of which are critical to the accuracy of the computed current.

Since all input variables with dimensions of distance are entered in wavelengths in the comparison case discussed above, they are all independent of frequency. They are also independent of constitutive parameters, which are used only to determine wavelength and propagation constant from the inputted frequency and hence are the same for soil media as for free space. Hence they are not useful in the study of the effect of frequency and soil parameters on the current distribution and radiation patterns. To conduct such a study, we enter input variable values in accordance with the model described in Section 1, and those with dimension of distance are entered in meters, not as multiples of a wavelength. Some current distribution computations have been made at various frequencies between 400MHz and 2.4GHz. An example is shown in Figure 4(a) for dry sand as the ambient medium (dielectric constant = 2.55, loss tangent = .008) at a frequency of 1GHz. This result is compared with that in Figure 4(b), where all antenna parameters and the frequency are the same as in Figure 4(a) except that the ambient medium is free space.

In these cases, the total length of the wire is about 200cm. Wavelengths for dry sand and free space are 18.8cm and 30cm respectively, implying wire-lengths of about 10.6 and 6.5 wavelengths respectively at 1GHz. The number of cycles of the real and imaginary parts of the current in Figures 4(a) and 4(b) is obviously the number of wavelengths.

Another example is shown in Figures 5(a) and 5(b), the same as Figures 4(a) and 4(b) except that the frequency is 2GHz. In these cases, the wavelengths are half of those at 1GHz and hence the number of cycles of real and imaginary parts are twice those of Figures 4(a) and 4(b). In all of these cases, the amplitude decays from a high value near the source to a smaller plateau near the termination, but there are significant differences in the shapes of the curves for each medium at the two frequencies and at each frequency for the two media, especially near the source and near the termination. Other sets of results have been obtained for different frequencies within the band of interest and they all show behavior similar to Figures 4(a)-5(b), except those at the low frequencies and, whose amplitudes do not exhibit significant decay with distance.

A general conclusion we can draw from the cases studied to date is that the rates of decay of the current amplitude show an increase with frequency. The frequency above which this occurs is about 950MHz in free space and about 550MHz in dry sand. At frequencies below these values, the amplitude is highly oscillatory and exhibits very little decay with distance along the wire. The wavelength in dry sand is about 63 percent of that in free space for any frequency. This appears

to be the reason for the difference between the two media in the minimum frequency at which significant decay rates begin to occur. For frequencies below that minimum, there is an insufficient number of wavelengths along the wire to realize the typical current distribution of the conical spiral.

As of this writing, we cannot draw conclusions about the effect of these trends on the radiation pattern, which we have not yet studied for the parameter values of our model. Some preliminary radiation pattern computations were made for the validation case, and plots of elevation plane cuts show similar shape but less symmetry than those given in [2]. These results will be shown in the presentation.

Figure 3(a)

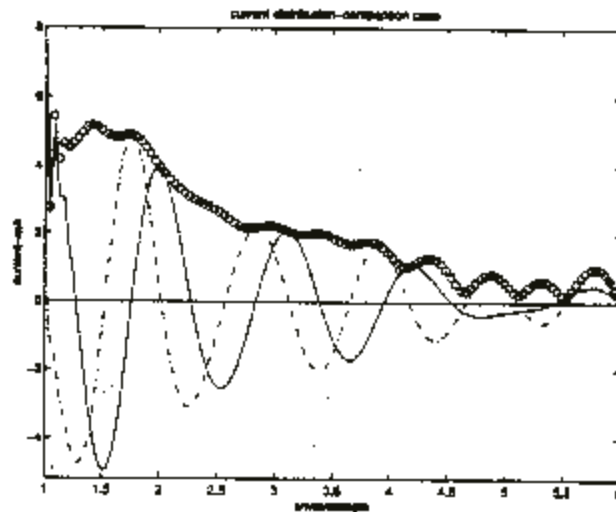


Figure 3(b)

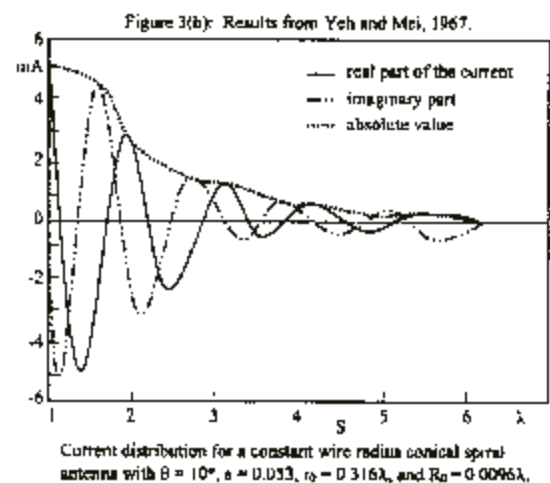


Figure 4(a)

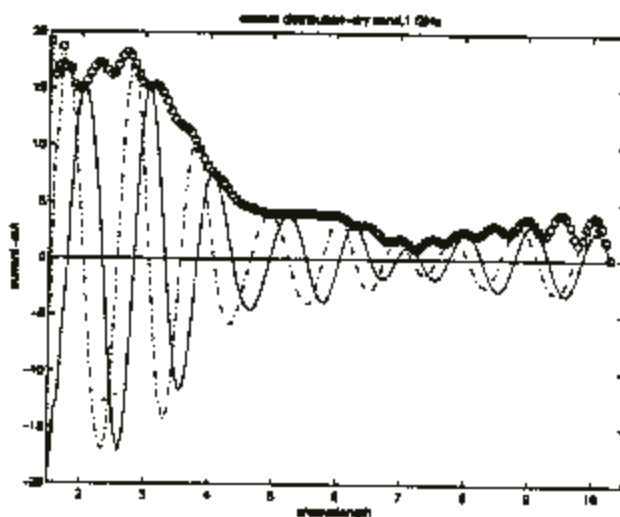


Figure 4(b)

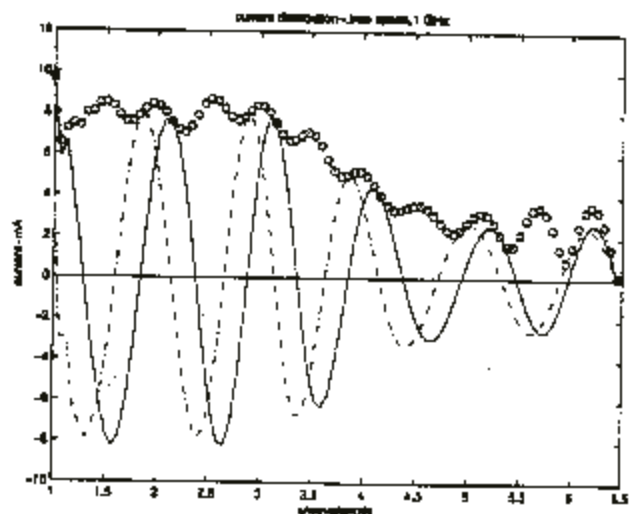


Figure 5(a)

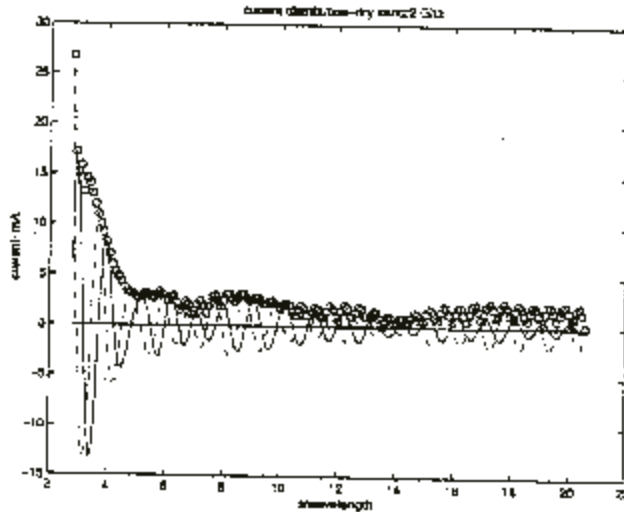
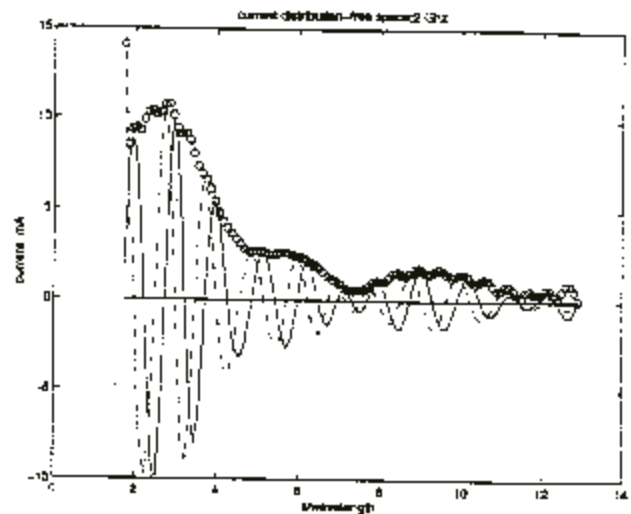


Figure 5(b)



4 Conclusions and Plans for Future Work

The results indicate that the modeling we have done to date will be useful in predicting the behavior of the conical spiral antenna in a soil medium, but more work is required to improve its accuracy for the case of an infinite medium and then to include the effects of the air-soil boundary and a mine and possibly other clutter sources in the near-neighborhood of the antenna. Final validation will require favorable comparisons with results to be obtained with the experimental model.

In the immediate future, before considering these refinements, it is planned to take more accurate account of the integration over the outer periphery of the wire and to try other basis functions and weighting functions in the MOM current computation, in hopes that the results will come closer to those obtained in the earlier literature.

5 Acknowledgments

The authors acknowledge support of this work by the Army Research Office under Demining MURI Grant #DAA 0-55-97-0013.

References

- [1] Mei, K. K., "On the Integral Equation of Thin Wire Antennas," IEEE Trans., AP-13, 3, pp. 374-378, May 1965.
- [2] Yeh, Y. S. and Mei, K. K., "Theory of Conical Equiangular Spiral Antennas, Part 1-Numerical Techniques," IEEE Trans., AP-15, 5, pp. 634-639, September 1967.
- [3] Yeh, Y. S. and Mei, K. K., "Theory of Conical Equiangular Spiral Antennas, Part-2 - Current Distributions and Input Impedances," IEEE Trans., AP-16, 1, pp. 14-21, January 1968.
- [4] Dyson, J. D., "The Characteristics and Design of the Conical Log-Spiral Antenna," IEEE Trans., AP-13, 4, pp. 488-499, July 1965.
- [5] Dyson, J. D., "Equiangular Spiral Antennas," IRE Trans., AP-7, pp. 181-187, April 1959.
- [6] Cheo, B. R. S., Rumsey, V. H. and Welch, W. J., "A Solution to the Frequency-Independent Antenna Problem," IRE, Trans., AP-9, 6, pp. 527-534, November 1961.
- [7] Rumsey, V. H., "Frequency Independent Antennas," Academic Press, NY and London, 1966. Chapter 4.
- [8] Elliott, R. S., "Antenna Theory and Design," Prentice-Hall, Inc., Englewood Cliffs, NJ, 1981.
- [9] Kraus, J. D., "Antennas," second edition, McGraw-Hill, New York, 1988, Chapter 7, pp. 696-703.
- [10] Balanis, C. A., "Antenna Theory," second edition, John Wiley & Sons, New York, 1997, Chapter 11, pp. 542-551.
- [11] Roubine, E. and Bolomey, J. C., "Antennas," (English Translation from French) Hemisphere Publishing Corp., Washington and New York, 1987, Volume 1, pp. 170-177.
- [12] Harrington, R. F., "Matrix Methods for Field Problems," Proc. IEEE, 55, 2, pp. 136-149, February 1967.
- [13] Elliott, Ref. [8], pp. 286-294.
- [14] Sarkar, T. K., Djordjevic, A. R. and Arvas, E., "On the Choice of Expansion and Weighting Functions in the Numerical Solution of Operator Equations," IEEE Trans., AP-33, 9, September 1985, pp. 988-996.
- [15] Balanis, C. A., "Advanced Engineering Electromagnetics," John Wiley & Sons, New York, 1989.

Validation Study of 3 Dimensional Ray-Based GPR Simulation Code

by

Harold Raemer, Carey Rappaport, Eric Miller and Roberta Young

Department of Electrical and Computer Engineering
Center for Electromagnetics Research

Northeastern University, Boston, MA 02115, U.S.A

Harold Raemer: 617-373-3035/racmer@cer.neu.edu

Carey Rappaport: 617-373-2043/rappaport@neu.edu

Eric Miller: 617-373-8386/emiller@cdsp.neu.edu

Roberta Young: 617-373-4874/reyoung@lynx.neu.edu

FAX: 617-373-8627 (for all of the above authors).

Presented at SPIE Acrosense 2000, Orlando, FL, April, 2000.

Abstract

In previous papers (Proc. SPIE, April, 1998, April, 1999 and August, 1999) the authors reported work on a 3D frequency domain simulation of bistatic GPR scenarios involving signals from buried mines and clutter due to random permittivity fluctuations in the soil, roughness of the air-ground interface, and distributions of rocks. The analysis is based on Born approximations. The emphasis in these papers was on simulation of a focused array radar (Rappaport and Reidy, Proc. SPIE 1997), which is a multistatic system and hence its simulation requires a large number of runs. Simulation of a multistatic GPR system places a high premium on speed, which necessitates some loss of accuracy. Work is currently underway on validation of this code through comparison with experimental results and with results obtained with numerical codes that can achieve great accuracy with very long running times. In the work reported in this paper, results obtained with our code, requiring only minutes of running time, are compared with results of a three dimensional FDTD code, which requires many hours of CPU time for the same case.

KEYWORDS

Radar, ground penetrating radar, clutter modeling, electromagnetic scattering, mine detection.

1 INTRODUCTION

In three previous SPIE Conference Proceedings papers [1, 2, 3], the authors reported work on a three-dimensional bistatic frequency domain simulation of ground penetrating radar signals from a subsurface region possibly containing mines. The emphasis in these papers was on application of the code to simulation of the action of the focused array radar system [4] in cases where the subsurface clutter consists of random permittivity inhomogeneities in the soil, and distributions of rocks in the vicinity of the mines.

The focused array system is able to resolve signals from different regions within the soil, using an algorithm based on differences in propagation time delay between returns from different regions. A frequency domain simulation, in order to capture that effect, must contain a feature that separates returns on the basis of time delay. In the work reported in our previous papers, such a feature was present, in the form of a simulation of the focused array radar principle. If the air-ground interface is flat, the multistatic returns with a specific relative delay are superposed and their sum is interpreted as arising from a specified subsurface region.

The work reported in the present paper does not contain a time-resolution feature. It is based on a pure CW computation of the wavefields arising from a static transmitting antenna at a height z_1 above an air-ground interface that has an average height designated as zero, but possibly with a small degree of surface roughness (i.e. σ_z , the rms height of the interface, is small compared with the transmitter height h , and with the burial depth of the top of a cylindrical dielectric mine assumed to be placed directly below the antenna). The horizontal coordinates of the centers of both antenna and mine are $x = 0$, $y = 0$. The wavefields arising from the transmitted wave, reflection of that wave from the air-ground interface, its downward transmission through the interface and scattering from the mine are all superposed coherently, resulting in a distribution of field-strength throughout a rectangular volume including both the above-surface region containing the transmitter and the subsurface region containing the mine. The frequency is 960 MHz.

The purpose of the work is to compare the modeling methodology used in [1, 2, 3] as applied to the geometry described above with a FDFD computation of the same geometry performed by R. Young and C. Rappaport [5]. The latter produces extremely accurate results at an expenditure of hours or days of CPU time. The former can obtain results for the same scenario in minutes of CPU time. This is accomplished through the use of approximations discussed in [1, 2, 3]. This produces a less accurate result but possibly one sufficiently accurate to provide a useful forward model for signal processing schemes on data obtained from large scale GPR systems (such as the focused array radar or other multistatic systems, particularly in cases where the simulated scenarios are dynamic or cover large amounts of space with many buried objects to be detected).

2 Description of Problem Geometry

The geometry is illustrated in Figure 1. In the cartesian coordinate system used in this computation, whose $z = 0$ plane is the (average) air-ground interface, the centroid of the transmitting antenna is

at $x = 0$, $y = 0$, $z = h = 10\text{cm}$. The vertical center of the buried mine, which is modeled as a finite upright circular cylindrical object, is at $x = 0$, $y = 0$, $z = -d_m = -7\text{cm}$. The mine's radius r_m is 5cm and its vertical thickness t_m is 4cm . Thus the top surface of the mine is at $z = -d_m + t_m/2 = -5\text{cm}$ and its bottom surface is at $z = -d_m - t_m/2 = -9\text{cm}$. The mine is characterized as a uniform slightly lossy dielectric with relative permittivity $\epsilon_m = 2.9$ and conductivity $\sigma_m = .004\text{ S/m}$. The soil parameters are $\epsilon_s = 6.2$ and $\sigma_s = .035\text{ S/m}$, which characterizes Puerto Rican clay loam.

The three dimensional region over which we must compute the fields, as illustrated in Figure 1, is rectangular with dimensions 50cm , 50cm , and 40cm in the x , y and z directions respectively. The lower vertical limit of this region is 20cm below the air-ground interface. The upper limit is 20cm above the interface. In the horizontal directions, the coverage region extends from $x = -25\text{cm}$ to $x = +25\text{cm}$ and from $y = -25\text{cm}$ to $y = +25\text{cm}$. In the FDFD work, there are two 3D volume grids, one designated as the course grid, whose number of elements in x , y and z directions are 95, 95, and 77 respectively. The other, designated as the fine grid, has 97, 97, and 85 elements in the x , y and z directions respectively. Wavelength is 31.2cm in air and 12.5cm in the soil.

3 Modeling of Electromagnetic Processes

The theoretical modeling behind the simulation was described in some detail in [1], [2], and [3]. The basic process modeled was illustrated in Figure 1 of [2], repeated here as Figure 2b, which illustrates the transmission of a ray from the transmitter at Point 1, intersecting the air-ground interface at q , traversing the interface and propagating downward in the soil toward point p , scattering from point p toward a point q' on the interface, again traversing the interface into free space, then propagating toward point 2.

The total wavefield at point 2, in addition to the superposition of all wavefields due to processes illustrated in Figure 2-b (i.e. for all air-ground-intersection points q and q' and for all subsurface points p whose permittivity differs from that of the ambient soil medium), but also contains two other contributions that were largely de-emphasized in [1, 2, 3], namely the fields of the direct wave from 1 to 2 and the superposition of rays reflected from the boundary at all surface points s . These contributions, which are in general much larger than the scattered wavefields due to subsurface inhomogeneities and buried objects, are illustrated in Figure 2-a. In [1, 2, 3], where the emphasis was on subsurface scattering with a focused array radar, it was assumed that these contributions could be time-gated out through the signal processing inherent in the action of the focused array and through further signal processing that served to isolate the subsurface returns from the direct and ground-reflected signals. However, in the computations performed in this paper, these contributions cannot be eliminated and the fields arising from them at any point above the surface are the dominant contributions to the fields at that point.

There are a few differences between the analysis performed in [1, 2, 3] and that in this paper. First, the focused array is not a subject of this paper. The transmitting antenna, as in the FDFD work, is a Hertzian dipole oriented in the horizontal direction, or more precisely along the x -axis. The basis of the algorithm for the radiated fields is the very simple and well known theory of a

“short” (compared with wavelength) electric dipole at the origin with uniform current, given in many electromagnetic theory texts, e.g. [6, 7, 8]. From Section 15-5, Eqs. (20), (21) and (23) on Page 747 of [6], the electric and magnetic fields in spherical coordinates r, θ, ϕ are:

$$E_r = 2K_0 \exp(-j\beta r) \cos \theta \left(\frac{1}{cr^2} + \frac{1}{j\omega r^3} \right) \quad (1-a)$$

$$E_\theta = K_0 \exp(-j\beta r) \sin \theta \left(\frac{j\omega}{c^2 r} + \frac{1}{cr^2} + \frac{1}{j\omega r^3} \right) \quad (1-b)$$

$$H_\phi = K_0 \epsilon_0 \exp(-j\beta r) \sin \theta \left(\frac{j\omega}{cr} + \frac{1}{r^2} \right) \quad (1-c)$$

where $K_0 = \frac{I_0 \ell}{4\pi \epsilon_0}$, β is the propagation constant of free space and where the dipole is assumed to be in the z -direction.

If we note that

$$\cos \theta = \frac{z}{r}, \sin \theta = \frac{\rho}{r}, \text{ where } \rho = \sqrt{x^2 + y^2}, r = \sqrt{\rho^2 + z^2} \quad (2)$$

then to redirect the dipole into the x -direction rather than the z direction, we redefine the quantities in (2) as follows

$$\cos \theta = \frac{x}{r}, \sin \theta = \frac{\rho}{r}, \text{ where } \rho = \sqrt{y^2 + (z - h)^2}, r = \sqrt{\rho^2 + x^2} \quad (3)$$

In this new geometry, where $z \rightarrow x, x \rightarrow y, y \rightarrow z$, the cosines and sines of the other angular coordinate ϕ are defined by

$$\cos \phi = \frac{y}{\rho}, \sin \phi = \frac{z - h}{\rho} \quad (4)$$

From the standard expressions for the spherical coordinate unit base vectors, in our new coordinate system

$$\begin{aligned} E_\theta &= \mathbf{E} \cdot \hat{\theta} = \cos \theta (E_y \cos \phi + E_z \sin \phi) - E_x \sin \theta \\ E_r &= \mathbf{E} \cdot \hat{r} = \sin \theta (E_y \cos \phi + E_z \sin \phi) + E_x \cos \theta \\ E_\phi &= \mathbf{E} \cdot \hat{\phi} = -E_y \sin \phi + E_z \cos \phi = 0 \end{aligned} \quad (5)$$

Inversion of (5) yields the rectangular coordinates of the electric field vectors at the observation point due to an x -directed Hertzian dipole at $x = y = 0, z = h$. Expressions analogous to (5) yield H_ϕ in terms of ϕ .

$$E_y = E_\theta \cos \theta \cos \phi + E_r \sin \theta \cos \phi \quad (6-a)$$

$$E_z = E_\theta \cos \theta \sin \phi + E_r \sin \theta \sin \phi \quad (6-b)$$

$$E_x = -E_\theta \sin \theta + E_r \cos \theta \quad (6-c)$$

$$H_y = -H_\phi \sin \phi \quad (7-a)$$

$$H_z = +H_\phi \cos \phi \quad (7-b)$$

$$H_x = 0 \quad (7-c)$$

where E_θ , E_r and H_ϕ are given by (1-b), (1-a) and (1-c) respectively and where, from (3) and (4)

$$\cos \theta = \frac{z}{r}, \sin \theta = \frac{\rho}{r}, \rho = \sqrt{(z-h)^2 + y^2}, r = \sqrt{\rho^2 + x^2}, \cos \phi = \frac{y}{\rho}, \sin \phi = \frac{(z-h)}{\rho}$$

The sequence of computations that lead to the results reported in Figures 3-6 are described in what follows:

3.1 Direct wavefield in above-surface region

The first stage of the computation (the $1 \rightarrow 2$ ray illustrated in Figure 2-a) is to specify the location of the transmitter at $x = y = 0$, $z = 10\text{cm}$, and x , y , z at the center of each element in the above-surface portion of the grid, specify the values of I_0 and ℓ (which will affect the end results only as a scale factor), then calculate E_r and E_θ through (1-a) and (1-b) with the aid of (3) and (4) for the center of every grid element above the air-ground interface, i.e. from $x = -25\text{cm}$ to $+25\text{cm}$, from $y = -25\text{cm}$ to 25cm and z from 0 to 20cm . Then for each of the elements the rectangular components of the electric field are computed from these results, using Eqs. (6-a, b, c).

The resulting three-dimensional arrays of electric field components over the above-ground portion of the grid is the first of the outputs obtained and is designated as the "direct wavefield". It represents the field that would be seen at each point in the absence of the lower-half space occupied by the soil.

3.2 Ground-reflected wavefield in above-surface region

The second contribution to the total field in the region above the air-ground interface is the ground-reflected wavefield (illustrated as the $1 \rightarrow s \rightarrow 2$ ray in Figure 2-a). If the interface is perfectly horizontal, i.e. with no surface roughness, then the major contribution is in the specular direction at each point of intersection of the direct wave with the interface. The method of approach is to begin with the direct wavefields immediately above the ground surface, use the tangent plane approximation for the total field just above the surface and evaluate the Stratton-Chu integral [9] over the portion of the air-ground interface included in our simulation domain (tacitly assuming that the region outside that domain has a negligible influence on the results). If the integration is done

by the stationary phase method, then using simple theory reported in many standard references, the stationary phase ("sp") point on the ground given a receiving point at (x_R, y_R, z_R) and a source point at (x_1, y_1, z_1) is easily shown to be

$$x'_{sp} = \frac{x_R z_1 + x_1 z_R}{z_1 + z_R}, \quad y'_{sp} = 0, \quad z_{sp} = 0 \quad (8)$$

where the primed horizontal coordinates are in a system whose $x' - z$ plane is the vertical plane of propagation of the wave from 1 to s to R . The angle ϕ' between that plane and the $(x - z)$ plane of the unprimed system is defined by

$$\cos \phi' = \frac{x_R - x_1}{\rho'}, \quad \sin \phi' = \frac{y_R - y_1}{\rho'}, \quad (9)$$

where $\rho' = \sqrt{(x_R - x_1)^2 + (y_R - y_1)^2}$. The global coordinates of the sp point are

$$x_{sp} = \frac{x_R z_1 + x_1 z_R}{z_1 + z_R}, \quad y_{sp} = \frac{y_R z_1 + y_1 z_R}{z_1 + z_R}, \quad z_{sp} = 0 \quad (10)$$

The computation of the electric field components through Eqs. (6-a, b, c) includes that of the fields just above the boundary. Eqs. (7-a, b, c) are then used to compute the magnetic field components just above the boundary. The electric field component that is normal to the plane of propagation is that of the transverse electric (TE) incident wave at the sp point and the corresponding magnetic field component is that of the transverse magnetic (TM) incident wave at that point. Then the TE and TM reflection coefficients R^{TE} and R^{TM} are computed at the sp point and used to determine the surface fields under the tangent plane approximation. Those fields are then used to evaluate the integral by the sp method, leading to a factor of the form

$$I_s = \frac{-j\pi}{k_0} \left[\frac{r_{1s} r_{sR}}{r_{1s} + r_{sR}} \right] \frac{1}{|\cos \theta_i|} \quad (11)$$

multiplying the Stratton-Chu integrand at the surface points, where r_{1s} and r_{sR} are the distances between transmitter and s and receiver and s respectively, θ_i is the angle of incidence, and where the TE and TM surface fields are given by $(E_i)_\perp (1 + R^{TE})$ and $(H_i)_\perp (1 + R^{TM})$ where $(E_i)_\perp$ and $(H_i)_\perp$ are the components of the incident fields normal to the plane of propagation. If the physical optics method is used to evaluate the integral for each rectangular surface patch, and there is no surface roughness, then the factor I_s in (11) is replaced by an expression of the form

$$I_s = \Delta S \sin c(\pi a_x) \sin c(\pi a_y) \quad (12)$$

where ΔS is the patch area and a_x and a_y are x and y components of the phase factor on the patch, given by

$$a_x = \sin \theta_i \cos \phi_i + \sin \theta_s \cos \phi_s$$

$$a_y = \sin \theta_i \sin \phi_i + \sin \theta_s \sin \phi_s$$

where subscripts i and s indicate incident and scattered respectively. In our code, both the sp method and the po method are available as options.

3.3 Wavefield transmitted into subsurface medium

The first set of outputs in the subsurface grid is the set of rectangular components of the electric fields of the wave transmitted across the boundary into the soil medium. The background theory for this process (the $1 \rightarrow q \rightarrow p$ process illustrated in Figure 2b) was discussed briefly in [1] (Text from Eq. (1) through the paragraph including Eq. (3)) and in greater detail in [2] (section 2 of [2] “Background Analysis”, including Eqs. (2-a) through (4-c) and the remainder of the paragraph below (4-c)) for a perfectly flat horizontal boundary. In [3] the analysis was extended to account for surface roughness. (The discussion of this process in [3] begins above Eq. (1) and extends one paragraph beyond Eq. (11-b)).

Further analysis in [3] (Eqs. (16) and (17) in [3] and the text surrounding these expressions), relates to the selection of the method of evaluation of the Stratton-Chu integral over the lower side of the air-ground interface. The same tradeoff between the greater accuracy of the physical optics method and the greater speed of the stationary phase methods of integration exists for both the reflection process discussed in Section 3.2 and the transmission process under discussion here. On both processes, the code contains both options.

3.4 Scattered wave from subsurface inhomogeneities

The background analysis for the wave scattered from complex permittivity inhomogeneities in the subsurface medium by the Born approximation was covered in [1] (Eq. (6) and associated text) and [2] (Text beginning above Eq. (5) and extending through Eq. (6)). In our previous work the scattered field from a subsurface region needed to be calculated only for observation points on the underside of the air-ground interface as an interim step in obtaining the fields at a specified above-ground receiver location. The Born scattering code was executed for each subsurface source point and for all observation points on the surface grid. In the current study, for each element within the subsurface volume grid whose complex permittivity differs from that of the ambient medium, Born scattering of a wave incident on that element must be computed for observation points at all other subsurface grid elements. The number of computations of Born scattering from each element is increased from $(N_x N_y)$ in [1, 2, 3] to $(N_x N_y N_z)$ in the present study, where N_x , N_y , and N_z are numbers of volume elements in the subsurface grid in the x , y , and z directions respectively and N_x and N_y are also the numbers of x and y surface patches on the air-ground interface.

3.5 Wavefield due to subsurface scattering into above-surface region

The wavefields due to scattering from subsurface inhomogeneities are superposed on the direct and ground-reflected wavefields described in Sections 3.1 and 3.2 respectively as a contribution to the fields in the above-surface grid. This process, illustrated as the path $p \rightarrow q' \rightarrow 2$ in Figure 2b, is

described for the case of the flat boundary in [1] (the “return path,” Eqs. (4) through (6) and associated text) and [2] (Text beginning above Eq. (6) and extending through Eq. (11-b)). It is extended to include surface roughness in [3] (From three paragraphs above Eq. (12-a) through Eq. (17) and the remaining text in Section 2). The same two options for integrating the Stratton-Chu integral over all surface points q' (the *sp* and *po* methods) are available for the return path wave as for the ground-reflected wave (Section 3.2) and the wave transmitted into the subsurface medium (Section 3.3). The factors multiplying the Stratton-Chu integrand for the return path for the *po* and *sp* methods are given respectively by (16) and (17) in [3] for the general case accounting for surface roughness.

3.6 Additional contributions to wavefields

If there is no subsurface clutter due to permittivity inhomogeneities (as we assume for these computations), then in the absence of a buried object such as a mine, the only contributions to the above surface wavefields are those due to the direct wave (Section 3.1) and the ground reflected wave (Section 3.2) and the only contributions to the subsurface wavefields are those due to the transmitted wave (Section 3.3). That is the first set of cases we are running.

If a plastic mine is present, then the Born scattering processes described in Sections 3.4 and 3.5 are also present and contribute to the total wavefield in the subsurface and above-surface regions respectively. This suffices for the first order Born approximation. It is evident that the contrast between mine and soil permittivity is too large to ensure accuracy with first order Born scattering. Accuracy would probably be significantly enhanced with second or higher order Born, i.e. the incident wavefield at the mine location includes the effect of downward reflection of the first order scattered fields from the underside of the air-ground interface and possibly further back-and-forth interactions between lower order scatterings from the mine and reflections of these scattered fields from the underside of the boundary. Although the code to implement these higher order effects can easily be made available, we have not attempted to include it in this study. The reason is the enormous proliferation of required running time for each higher order process. Since the major advantage of our approach over highly accurate numerical codes such as MOM or FDFD is computation speed, it might be counter productive to go too far in this direction.

4 Results and Conclusions

The results of the computations discussed above are presented in Figures 3 through 6, each of which is a 3 dimensional plot of the amplitude of the electric field vs. a horizontal coordinate (x or y) on one axis and the vertical coordinate (z) on the other axis. The xz plots (Figures 3 and 4) are from $x = -25cm$ to $+25cm$, $z = -20cm$ to $+20cm$ and $y = 0$. The yz plots (Figures 5 and 6) have the same range of y and z with $x = 0$. In each case, the computations were of the three rectangular components of the electric field amplitude, $|E_x|$, $|E_y|$ and $|E_z|$ and the total field amplitude $\sqrt{E \cdot E^*}$. Only the total amplitude is shown on the plots. Figures 3 and 5 are for the case of a flat air-ground

interface without the mine and with the mine respectively. Figures 4 and 6 are also for the cases of no mine and mine respectively. It is evident that there is no perceptible difference between these four plots. First, the mine and no-mine cases cannot be distinguished on the xz plots or on the yz plots. Secondly, the xz plots and the yz plots show nearly perfect symmetry about the yz and xz planes respectively, which is expected since the transmitter is located at the horizontal center of the simulation domain.

Also, the xz and yz plots show only miniscule differences in shape in both sets of plots and the amplitudes appear to be zero in the subsurface region. In the output files, the subsurface fields are not zero but are so small compared to the peak values in the vicinity of the peak that they plot as zero. This is because the amplitude resolution was chosen to include the peaks on the plots. Plots that are made with finer resolution showed small nonzero values in the subsurface grid but truncated the peaks. Small differences between xz and yz plots do appear when amplitudes of field components rather than total field are plotted (Figures 7 and 8). This is expected because the dipole is parallel to the xz plane and therefore the xz and yz planar cuts are different.

The comparisons between these results and FDFD results are not yet completed as of this writing. It is planned that comparison results will be shown at the conference presentation. Current indications are that the general shapes of the two curves are similar for the resolution scales used. However, much better distance resolution is available for the FDFD computations, which are run on a computer with ten times the RAM currently available for the ray-based computations. For this reason, some of the fine detail that shows up on the FDFD plots is suppressed on the ray-based plots. Running times for the FDFD code are between 30 and 40 hours while those currently being done with the ray-based code require about one minute for the no-mine case and about seven minutes for the case with a mine present. For the latter computations, enhancing the distance resolution, using physical optics rather than stationary phase for the integration over the boundary, and using higher order Born approximations for scattering from the mine will certainly increase CPU time, but we have not yet determined the amount of the increase.

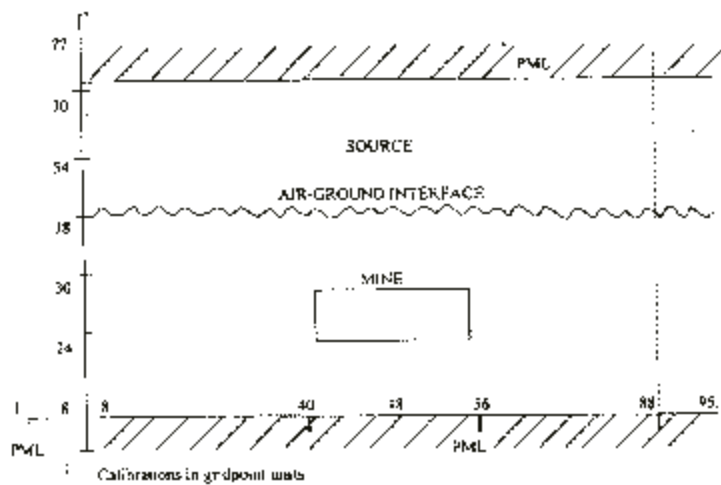


Figure 1: Geometry of volume grids, source and mine

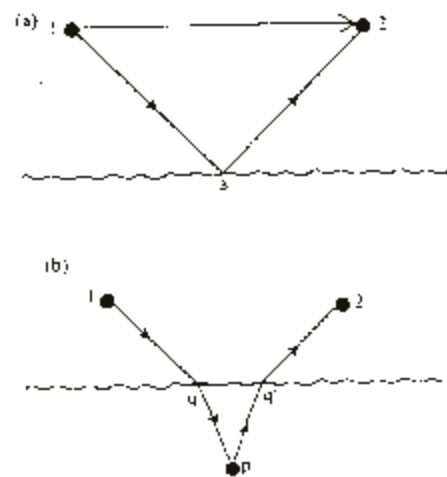


Figure 2: Illustration of Propagation and Scattering Processes

Figure 3: Field amplitude vs. x and y plane

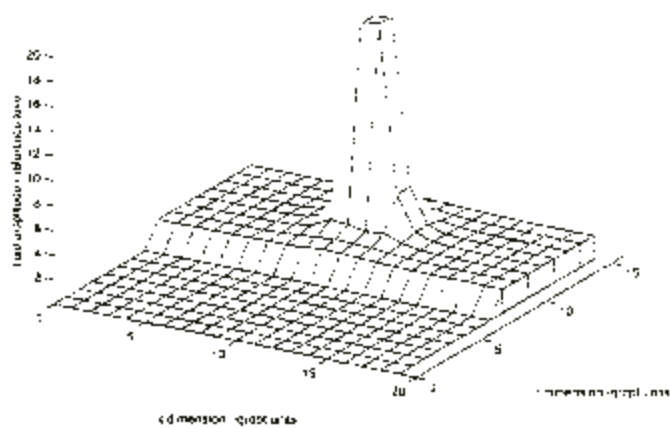


Figure 4: Wave amplitude vs. x and y plane

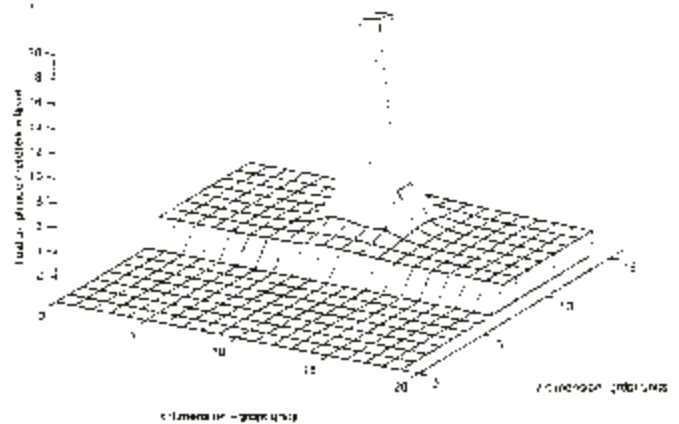


Figure 5, lead amplitude vs y & z, no time

Figure 6, lead amplitude vs y & z, no present

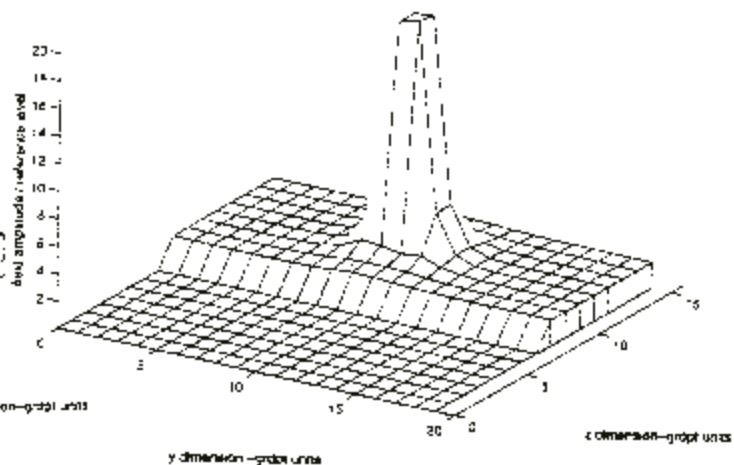
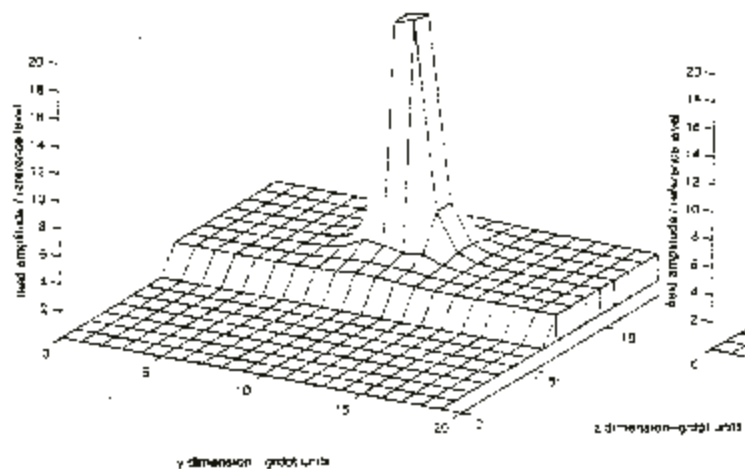
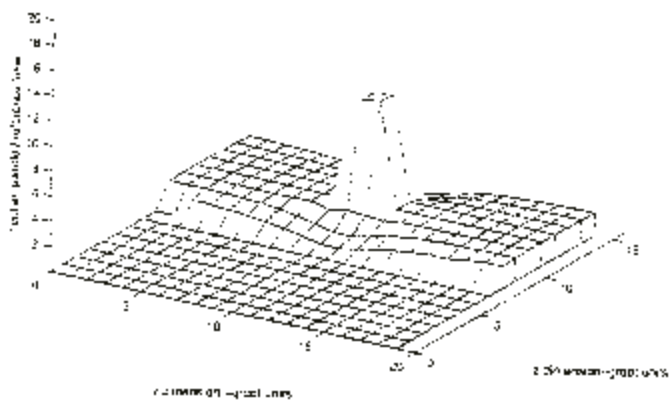
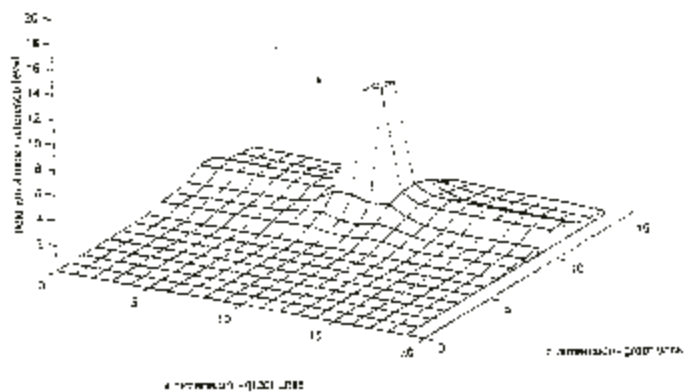


Figure 7, lead component vs x & z

Figure 8, lead component vs x & z



References

- [1] Raemer, H. R., C. M. Rappaport and E. L. Miller, "Frequency Domain Simulation of Focused Array Radar Returns from Buried Mines in Clutter" Proceedings of SPIE, Volume 3392, Orlando, FL, April 1998, pp 754-765.
- [2] Raemer, H. R., C. M. Rappaport and E. L. Miller, "Near-field and Timing Effects in Simulation of Focused Array Radar Signals from a Mine in Subsurface Clutter," Proceedings of SPIE, Volume 3392, Orlando, FL, April 1999.
- [3] Raemer, H. R., C. M. Rappaport and E. L. Miller, "Surface Roughness Effects in Simulation of Mine Signals Using a Focused Array GPR System," Proceedings of SPIE, Volume 3752, Denver, CO, July 1999.
- [4] Rappaport, C. M. and D. M. Reidy, "Focused Array Radar for Real Time Imaging and Detection," Proceedings of SPIE, Volume 2747, Orlando, FL, April 1996, pp 202-213.
- [5] Young, R. and C. M. Rappaport, "Informal Report on FDTD Computation," Northeastern University, November 1999.
- [6] Kraus, J. D., "Electromagnetics," 4th Edition, McGraw-Hill, New York, 1992, p 747.
- [7] Jordan, E. C. and K. Balmain, "Electromagnetic Waves and Radiating Systems," Second Edition, Prentice Hall, Englewood Cliffs, N.J., 1968, pp 321-323.
- [8] Narayana Rao, N., "Basic Electromagnetics with Applications," Prentice Hall, Englewood Cliffs, NJ, 1972, pp 494-495.
- [9] Stratton, J. A., "Electromagnetic Theory," McGraw Hill, New York, 1941, Chapter 8, Section 8.4, pp 464-470.

Suppressing GPR Clutter from Randomly Rough Ground Surfaces to Enhance Nonmetallic Mine Detection

Carey Rappaport, Magda El-Shenawee, and He Zhan
Northeastern University
235 Forsyth Building, Boston, MA 02115
(617) 373-2043 (v), (617) 373-8627 (f)
rappaport@neu.edu

Abstract--This study attempts to quantify the ground penetrating radar rough ground surface clutter by numerical modeling of wave scattering, and establish a strategy to suppress the clutter for given test signals. The goal is to improve the GPR detection statistics for small, buried, low-contrast nonmetallic antipersonnel mines. Using a model of an experimentally measured impulse GPR signal, we simulate the ground surface and buried low-contrast mine target scattered responses. We employ a 2D finite difference time domain (FDTD) method to analyze the pulse shape, delay, and amplitude characteristics of the scattered waves -- with and without buried nonmetallic mine targets -- as a function of roughness parameters. Five hundred Monte Carlo simulations of various test cases of specified ground root mean square height and correlation length were run to generate statistics for the clutter and target signal variations. In addition, the effectiveness of identifying and removing the ground surface clutter signal for detecting subsurface targets is presented. Results indicate that even with moderate roughness, statistics can be generated to enhance the detection of small, shallow, low contrast targets.

1. INTRODUCTION

The problem of detecting buried dielectric targets -- such as nonmetallic antipersonnel mines -- with ground penetrating radar (GPR) is important and challenging. Because the dielectric constant and electrical conductivity of the mine target (2.9 and 0.0001 at 1 GHz) is similar to that of the surrounding soil and its size is comparable to the thickness of soil above it, detection and discrimination are difficult. In addition, the soil dielectric constant may not be well characterized, and the ground surface will usually be rough, often with surface height variations of the order of the target burial depth. While there are many sources of clutter obscuring the mine target signal — including volumetric inhomogeneities (rocks, roots, metal fragments) and surface vegetation — the largest single source of undesirable signal is the ground surface itself. Since the ground has an infinite surface and presents a larger impedance mismatch with the air above it than with the low-contrast, nonmetallic target within it, its contribution to clutter is quite significant. Further, since buried clutter objects can only be inferred by imaging and reconstruction, the ground surface is unobscured, and its effects can be measured directly.

Impulse ground penetrating radar has been used as a robust and relatively inexpensive means of detecting underground objects [1]. By observing the arrival time of a subsurface scattered pulse and eliminating the reflection from the ground surface by time gating, it is possible to detect deeply buried anomalies. However, when the target is small, shallow, and of low contrast, special modeling and processing are required to characterize and separate the ground surface clutter from the target signal. A commonly used procedure of

background averaging to remove the ground clutter signal can be effective for very smooth ground surfaces, but tends to rapidly degrade for moderate roughness. In this study, we simulate the effects of rough ground on the GPR signal using Monte Carlo FDTD modeling of random surface variation. Because of the need for multiple case studies of hundreds of Monte Carlo runs, we used a very fast 2-D TM FDTD code, specifically adapted to frequency dependent, lossy media, with a lossy Perfectly Matched Layer (PML) ABC [2-4]. Clearly, more accurate modeling is afforded with a three-dimensional FDTD model, but each set of 500 runs would require 2000 hours of supercomputer CPU time. Instead, this study presents a range of possible signal delay, attenuation, and distortion characteristics, with associated statistical variation as a function of ground surface roughness.

As a baseline, we model the wave scattering for a bistatic impulse GPR geometry, based on the Geo-Centers EFGPR mine detection system, as shown in Fig. 1, using the measured Geo-Centers TEMR antenna element radiated signal as the excitation, presented as a function of time in Fig. 2 [5]. This pulse is wideband, with frequency response (6 dB roll-off) from 700 MHz to 1.3 GHz. This frequency range is ideal for detecting antitank mines, which are thicker and more deeply buried than antipersonnel mines, but is a little too low for optimal discrimination of objects with scale lengths of the order of 5 cm. The TEMR element radiates a fairly broad beamwidth nearfield antenna pattern, with wideband half-power points at roughly $\pm 60^\circ$ relative to boresight. The FDTD time and space steps used are $\Delta t = 20\text{ps}$ and $\Delta = 1.22\text{cm}$, maintaining a Courant condition $r = 0.5$. Simulations are run for 500 surface realizations -- with and without a

mine target at a typical burial depth of 8.5 cm below the nominal surface level -- for a variety of roughness statistics.

2. ROUGH SURFACE MODEL FORMULATION

Assuming that the random height of the ground surface has a Gaussian distribution with zero mean and standard deviation equal to σ_h , the probability density function of the height z is [6]:

$$p(z) = \frac{1}{\sigma_h \sqrt{2\pi}} \exp\left(-z^2 / 2\sigma_h^2\right) \quad (1)$$

This statistical distribution of the height provides no information about the distances between the hills and the valleys of the surface [6]. An additional function is needed to describe the density of the surface irregularities of the rough surface. This function is the autocorrelation function or its Fourier Transform, the surface profile power spectral density function. The autocorrelation function $R(x_d)$ gives the correlation between the random heights at two different points on the surface, x_1 and x_2 . It is defined by [6, 7]:

$$R(x_d) = \frac{\langle f(x_1)f(x_1 + x_d) \rangle}{\sigma_h^2} \quad (2)$$

where $x_d = x_2 - x_1$. For full correlation, $\lim_{x_d \rightarrow 0} R(x_d) = 1$, and for independence,

$\lim_{x_d \rightarrow \infty} R(x_d) = 0$. Moreover, if the surface profile spectral density function $W(K_x)$ is given,

then the auto-correlation function $R(x_d)$ can be obtained by the inverse Fourier Transform as:

$$R(x_d) = \frac{1}{s_h^2} \int_{-\infty}^{\infty} W(K_x) \exp(iK_x x_d) dK_x \quad (3)$$

Thus, the Gaussian joint probability density function for two heights on the surface, z_1 and z_2 , with zero mean and standard deviation σ_h is given by [6, 8]

$$p(z_1, z_2) = \frac{1}{2\pi s_h^2 \sqrt{1 - R^2(x_d)}} \exp\left(-\frac{z_1^2 - 2R(x_d)z_1z_2 + z_2^2}{2s_h^2(1 - R^2(x_d))}\right) \quad (4)$$

If the surface spectral density $W(K_x)$ is assumed to be Gaussian as [7]

$$W(K_x) = \frac{l_c s_h^2}{2\sqrt{p}} \exp\left(-\frac{K_x^2 l_c^2}{4}\right) \quad (5a)$$

in which K_x is the surface profile wave number. Thus from (3), the auto-correlation function will also be Gaussian given by

$$R(x_d) = \exp\left(-\frac{x_d^2}{l_c^2}\right) \quad (5b)$$

where l_c is the correlation distance for which $R(x_d)$ will drop to the value e^{-1} .

While other statistical variations for randomly rough ground surfaces may be more applicable for some rough ground surfaces, the Gaussian provides a sufficiently rich family of surface realizations to establish a population for analyzing clutter effects.

Although statistical analysis can be done on the received signals themselves, it is more useful to consider the particular effects on these signals caused by scattering by the rough ground. For example, Fig. 3 shows the computed received signals from 100 rough surface realizations of a ground surface with height and correlation length $\sigma_h = 3\text{cm}$ and $l_c = 10\text{cm}$, without (Fig. 3a) and with (Fig. 3b) a buried target. The mean signal and levels

one standard deviation above and below the mean signal is overlaid on each plot. It is clear that there is no discernable difference between the signals for the buried target plot and the target-free plot.

If the correlation length l_c is large, the received signal will retain its pulse shape in time, experiencing primarily a shifting in time and a change in amplitude corresponding to greater or lesser propagation distances from source to ground to receiver. By identifying these parameters and measuring their statistics separately, much added information about the clutter becomes available. In addition, compensating for the time shift and amplitude scaling allows the shape of each individual received signal to be examined, and used in the determination of whether a buried target is present.

To identify the amplitude scaling A_i of the received pulse S_i for a given trial, we compute the square root of the total energy in the signal divided by the energy in a suitable reference signal. We use the computed received scattered signal from an ideal soil half-space with a flat boundary as the reference S_f :

$$A_i = \sqrt{\sum_{n=1}^N |S_i(n)|^2} / \sqrt{\sum_{n=1}^N |S_f(n)|^2} \quad (6)$$

Alternatively, the value of the signal peak could be used for this scaling, but it was determined that energy normalization is superior for rougher surfaces that generate greater pulse distortion.

The time shifting is found by cross-correlating each signal under test with the same reference signal, each normalized to the square root of its energy. The cross-correlation

function indicates the inter-dependence of the values of two different processes at two different times:

$$C_{fi}(m) = \frac{\sum_{n=1}^{N-|m|} S_f(n) S_i(n+m)}{\sqrt{\sum_{n=1}^N |S_f(n)|^2} \sqrt{\sum_{n=1}^N |S_i(n)|^2}}, \quad \text{for } m \geq 0 \quad (7a)$$

$$C_{fi}(m) = C_{if}(-m), \quad \text{for } m < 0 \quad (7b)$$

where $i = 1, 2, 3, \dots, M$ is the rough surface realization index, M is the size of Monte Carlo sample, and N is the total number of time steps. Note that the normalized cross-correlation of the reference signal with itself has a maximum of unity at $m = 0$; that cross-correlation with a shifted copy has a unity maximum at the index corresponding to the shift; and that cross-correlation of dissimilar signals will have a maximum less than unity. The maximum value of this cross-correlation function is a measure of the pulse shape distortion from that of the ideal flat ground response.

3. NUMERICAL RESULTS

Numerical experiments were performed on the well-measured Puerto Rican clay loam with 10% moisture and 1.4 g/cc density [9]. This soil has dielectric constant varying as $6.4 < \epsilon' < 6.1$ and electrical conductivity $0.033 < \sigma < 0.067$ and the wavelength in this soil varies from $17.0 < \lambda < 9.5$ cm over the 700 MHz to 1.3 GHz bandwidth. The dispersive FDTD model employed uses the Z-transform supplemental equation model for frequency dependent conductivity [2]. With this model, the real dielectric constant is

assumed to be constant and the ratio of electric current density and electric field is given by:

$$\mathbf{S}(\mathbf{w}) = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2}}{1 + a_1 Z^{-1}},$$

where $Z = e^{-i\mathbf{w}D}$ for the frequency \mathbf{w} . For the particular type of Puerto Rican clay loam, the parameters have been found to be [3]: $b_0 = 0.916249$, $b_1 = -1.67662$, $b_2 = 0.761072$, $a_1 = -0.88$, and $\epsilon'_{Av} = 4.167$. Note that since the imaginary part of $\sigma(\mathbf{w})$ is non-zero, it will contribute to the real part of the dielectric constant, raising it above the ϵ'_{Av} value to the measured values 6.1 to 6.4 .

For each pair of Gaussian height and correlation parameters, 500 FDTD runs were performed on different surface realizations. In each case, the scaling and shift were determined using (6) and (7). An example of the distribution of these characteristics for the pair $\sigma_h = 3\text{cm}$ and $l_c = 30\text{cm}$, is shown in Fig. 4. The average scaling is 0.995 and the average shift is -55 ps. From these histograms, it is apparent that although the ground heights are normally distributed, the amplitudes and time shifts are not.

With the ultimate goal of target detection, it is important to observe the differences in the received signal when a mine is buried under the rough ground surface. For each surface realization, this difference can be clearly seen by simply subtracting the nominal ground-only signal from the signal with the mine present. In this simulated situation, the clutter would be known *a priori* and numerically removed. In practice, this is not possible, since the clutter signal is not separately available. It is the goal of these Monte Carlo

experiments to reasonably estimate the rough surface clutter effects so that they can be suppressed relative to the mine signal, without the *a priori* knowledge of the ground clutter effect for a given trial.

Using the individual FDTD differences, the responses due to the presence of mines can be compared to the signals from the ground surface alone. The mine position is usually determined by measuring the time delay corresponding to the path to and from the buried target. However, with rough ground, the path from the transmitter to ground to mine and back out to the receiver, changes with the local ground surface height, or the height of the antenna. Since the wave propagation velocity is quite different in air as opposed to soil, the target time delay for a given mine depth varies considerably.

Scatter plots showing the distribution of the shifts of the ground scattered signals $\tau_{\text{gnd}(i)}$ compared to the time delay of the mine scattered signals $\tau_{\text{mine}(i)}$ relative to a nominal perfectly flat ground for different rough surface parameters are shown in Fig. 5. For low contrast targets the dominant aspect of each scattered signal is the ground scattering, so the cross-correlation function gives $\tau_{\text{gnd}(i)}$. To find $\tau_{\text{mine}(i)}$, (7a) is applied to the difference between the signal scattered by the ground and mine and the ground alone (the target signal with *a priori* ground clutter signal removed). The surface root mean square height is $\sigma_h = 3\text{cm}$ in Fig. 5a and Fig. 5b while it is 2cm in Fig. 5c and Fig. 5d, and 1cm in Fig. 5e and Fig. 5f. The correlation length is $l_c = 10\text{cm}$ for the figures on the left and 3cm for those on the right. There is a strong correlation between $\tau_{\text{gnd}(i)}$ and $\tau_{\text{mine}(i)}$. A regression analysis is conducted to fit these simulated data with a straight line. The slope of this line

corresponds to the relative delay between the ground and target scattered signals. The slope is negative, and for very long correlation lengths—corresponding to large, flat raised or lowered portions of ground—would be equal to the $(1-\sqrt{\epsilon'})$. Since a positive shift in τ_{gnd} indicates the presence of a depression, which in turn implies less soil covering the mine, the time delay for the mine signal, τ_{mine} , is reduced by the difference of the wave velocities in air and soil times the depth of the depression. As shown in Fig. 5, the fitting error increases with the root mean square slope $\sigma_s = 1.414 (\sigma_h / l_c)$ [10]. Also, it can be seen that the distribution of points is closer to the mean for small height variation σ_h .

The clutter signal can be suppressed and consequently the target signal can be enhanced using physics-based signal processing. Subtracting the measured or modelled ground surface clutter signal has been a successful strategy to enhance the signal under test [11]. It is possible to improve on this basic algorithm for rougher surfaces using a multiple pass process. First, the average clutter signal is found by shifting each ground-only signal by $-\tau_{\text{gnd}(i)}$, then taking the ensemble average [12]. Note that this is an average over signals, not time, so the resulting signal will be the average pulse shape for the rough ground with delay compensation for the local height variation. Second, this average signal is shifted back by $\tau_{\text{gnd}(i)}$, and scaled by factor A_i in (6) for each mine-in-ground signal and subtracted from these signals. By using the previously modeled (or measured) surface signals to find the ensemble average pulse shape gives the best estimate for the local clutter that corrupts the target signal. Fig. 6 shows the result of this subtraction on each of the 100 signals of Fig. 3. The differences between the ground-only (Fig. 6a) and

mine-present (Fig. 6b) signals are now quite clear. The maximum amplitude of each processed signal is considerably smaller than for each of the original signals in Fig. 3, since much of the clutter has been removed. Since the appropriately delayed average ground clutter signal has been subtracted rather than the *a priori* known clutter signal for each trial, the clutter removal is not perfect. The ensemble averages for the ground-only and mine-in-ground cases in Fig. 6 indicate that for the particular soil type, mine depth, and radar excitation, the presence of signal at a certain time interval and above a given threshold corresponds to a buried target. This provides the basis for a statistical target detection procedure. The results for this example case do not apply for all cases, but the strategy for suppressing ground surface clutter is generally applicable, as long as the mine is not too close to the surface, or the excitation pulse too long. This approach could be used in realistic field measurements, without separately measuring the ground roughness or knowing if a target was present.

If the surface scattered wave were primarily due to a single specular reflection – as would be the case if l_c were large – then this procedure would suppress most of the surface clutter. However, it is possible that the surface scattering occurs at multiple points. In this case, the cross-correlation/shifted ensemble averaging and subtracting procedure is repeated. This process is shown schematically in the flowchart of Fig. 7. The initial box labeled “Raw Signals” refers to either the 500 Monte Carlo FDTD simulated signals or actual measured data, obtained progressively with updated averaging.

The signals produced from these multiple subtractions represent the signals scattered just from the target. The obtained shifts τ_i are amplified by the slope values of the straight lines shown in Fig. 5 and then used align these target-only signals. Fig. 8 shows the result of multiple subtraction and shifting to realign the mine signals. Note that this realignment would have to be performed for every presumed mine depth. This procedure can be performed during actual GPR operation in the field, using several ground-only calibration measurements as the average signal. For hardware focused systems, the ground surface clutter suppression must occur at each receiver, before the signals from multiple receivers are combined.

The average clutter suppressed signals for ground with mine present have been obtained for the same Gaussian roughness parameters of Fig. 5, and shown in Fig. 9. Also shown are curves indicating one standard deviation above and below the mean. As expected, the standard deviation is much smaller for the target signal than for the clutter signals. Also visible in Fig. 9 is the increase of clutter with the surface mean square slope σ_s . Correlating any trial signal for a given statistical ground roughness and the average signal for that roughness with inverse weighting by the standard deviation provides a strong parameter for estimating the presence of a mine at a given position.

Note that the ground signal is detected for each trial signal, so that any given ground roughness variations, or even radar antenna height variations are accounted for. That is if the GPR bounced or sagged at a given sample, the ground signal would still be correlated with the shape of the average ground signal, and the appropriate shift is determined.

4. CONCLUSIONS

Identifying the time shift and amplitude scaling of the ground surface clutter by correlation with the ideal flat ground provides a means for ground surface clutter removal for a given signal. Shifting this clutter-suppressed signal by a time delay roughly proportional to the differential propagation velocity in the effective soil layer (or absence of soil) relative to the nominal soil level realigns the target signal to its expected temporal position. Using this procedure, even fairly shallow buried nonmetallic mines signals can be distinguished from rough ground surface clutter using available (nonideal) impulse GPR sources.

Although this study is based on two-dimensional synthetic modeling, it presents a representative set of randomly cluttered signals, and shows that the physics-based processing suppresses clutter and leads to enhanced target detection. The problem of mine detection is certainly more complicated than detecting anomalous signals generated by subsurface objects. The issue of distinguishing the particular mine target, as opposed to other buried objects: rocks, roots, pockets of water, is not addressed in this study. However, combining the information from multiple views from different radar positions will help characterize the shape of the anomaly. In addition, fusing information from other modalities—such as electromagnetic induction, acoustic, and infrared detection—with GPR will help reduce false alarms, giving some hope in the effort to characterizing buried mines.

ACKNOWLEDGMENTS

This work is supported by the Army Research Office grant No. DAAG55-97-0013 and by CenSSIS, the Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation, award number EEC-9986821.

REFERENCES

- [1] Peters, L., and Young, J., "Applications of subsurface transient radars," I *Time Domain Measurements in Electromagnetics*, Miller, E. Ed., New York: Van Nostrand Reinhold, 1986.
- [2] W. Weedon and C. Rappaport, "A general method for FDTD modeling of wave propagation in arbitrary frequency-dispersive media," *IEEE Trans. Ant. Prop.*, pp. 401-410, March 1997.
- [3] Rappaport, C., Wu, S., and Winton, S., "FDTD Wave Propagation Modeling in Dispersive Soil Using a Single Pole Conductivity Model," *IEEE Trans. Magnetics*, vol. 35, May 1999, pp. 1542--1545
- [4] C. Rappaport and S. Winton, "Using the PML ABC for Air/Soil Wave Interaction Modeling in the Time and Frequency Domains," *International Journal of Subsurface Sensors and Applications*, July 2000.
- [5] A. Sahin, C. Rappaport, and A. Dean, "Design Considerations for Short Time Pulse TEMR Antennas using Finite Difference Time Domain Algorithm" *SPIE Aerosense Conf.*, Orlando, FL, April 1998, pp. 784-793.
- [6] P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, New York, 1963.
- [7] E. Thorsos, "The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum," *J. Acoust. Soc. Am.* 83 (1), 1988.
- [8] A. Mood, F. Graybill, and D. Boes, *Introduction to the Theory of Statistics*, McGraw-Hill, 1974.
- [9] Hipp, J., "Soil Electromagnetic Parameters as Functions of Frequency, Soil Density, and Soil Moisture," *Proc. of the IEEE*, vol. 62, no. 1, January 1974, pp. 98-103.
- [10] F. Bass and I. Fuks, *Wave Scattering from Statistically Rough Surfaces*, Pergamon, New York 1979.
- [11] M. El-Shenawee and C. Rappaport, "Quantifying the Effects of Different Rough Surface Statistics for Mine Detection Using the FDTD Technique," *Proc. SPIE*, April 2000.
- [12] Warrick, A., Azevedo, S., and Mast, J., "Prediction of Buried Mine-like Target Radar Signatures using Wideband Electromagnetic Modeling." *Proc. of SPIE*, vol. 3392, pp. 776-783, 1998.

FIGURE CAPTIONS

Figure 1. Bistatic ground penetrating radar scattering configuration with rough ground surface and buried mine target, based on the Geo-Centers, Inc. EFGPR system

Figure 2. Measured time pulse transmitted by the Geo-Centers, Inc. EFGPR antenna element. Time units are in 20 ps time steps, amplitude is given in arbitrary relative intensity units.

Figure 3. One hundred sample Monte Carlo FDTD calculations of field intensity as a function of time, received after scattering from a half space of Puerto Rican clay loam (10% moisture and 1.4 g/cc density) with randomly rough ground surface (Gaussian parameters: $\sigma_h = 3\text{cm}$ and $l_c = 10\text{cm}$) for; a) soil alone, and b) soil with a 10 cm by 5 cm rectangular region of TNT buried 8.5 cm below the nominal surface. Also shown are the mean signals and curves one standard deviation above and below the means. Note that the differences between the mean signals are too small to distinguish.

Figure 4. Histograms showing distributions of a) Relative amplitude scaling and b) Relative time shift for received signals scattered by 500 realizations of randomly rough ground ($\sigma_h = 3\text{cm}$ and $l_c = 30\text{cm}$).

Figure 5. Scatter plots showing the correspondence of time delays of the signal from the mine target buried 8.5cm below the mean ground height compared to the time shifts due to the ground surface alone for various Gaussian parameters: a) $\sigma_h = 3\text{cm}$ and $l_c = 10\text{cm}$,

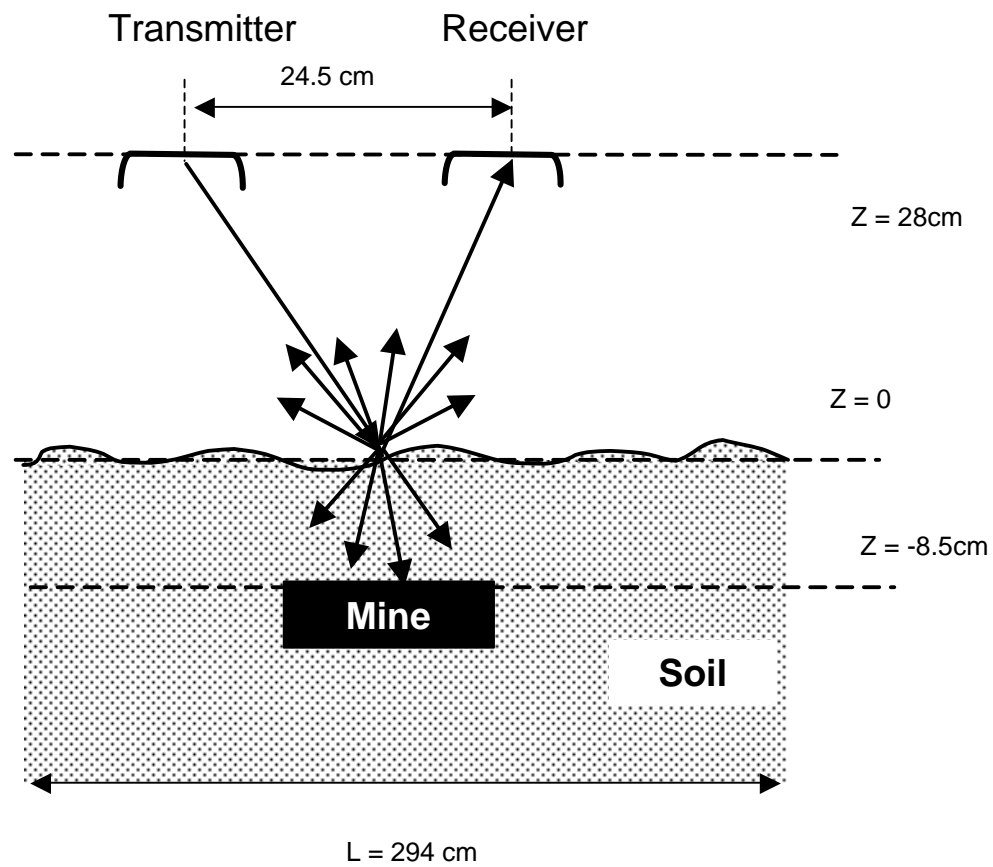
b) $\sigma_h = 3\text{cm}$ and $l_c = 3\text{cm}$, c) $\sigma_h = 2\text{cm}$ and $l_c = 10\text{cm}$, d) $\sigma_h = 2\text{cm}$ and $l_c = 3\text{cm}$, e) $\sigma_h = 1\text{cm}$ and $l_c = 10\text{cm}$, f) $\sigma_h = 1\text{cm}$ and $l_c = 3\text{cm}$.

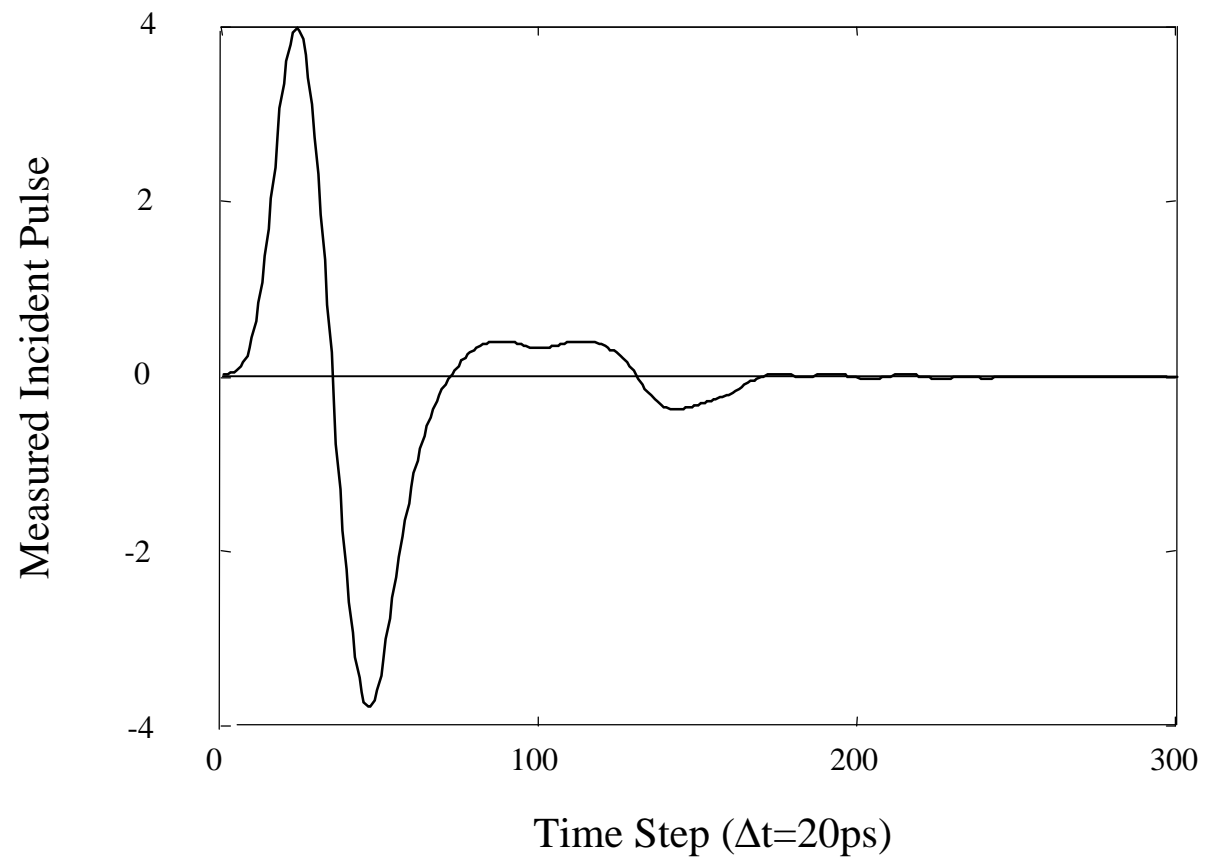
Figure 6. One hundred Monte Carlo calculations of the same physical situation as Figure 3, each processed by subtracting the average ground surface clutter, correlated in time to the sample signal. Note that the amplitude of these processed signals is about one-third of the original signals, and that the mine signal is clearly distinguishable from the mean signal in b) at time step 240, compared to the ground-only mean signal in a).

Figure 7. Flowchart of the ground clutter suppression algorithm.

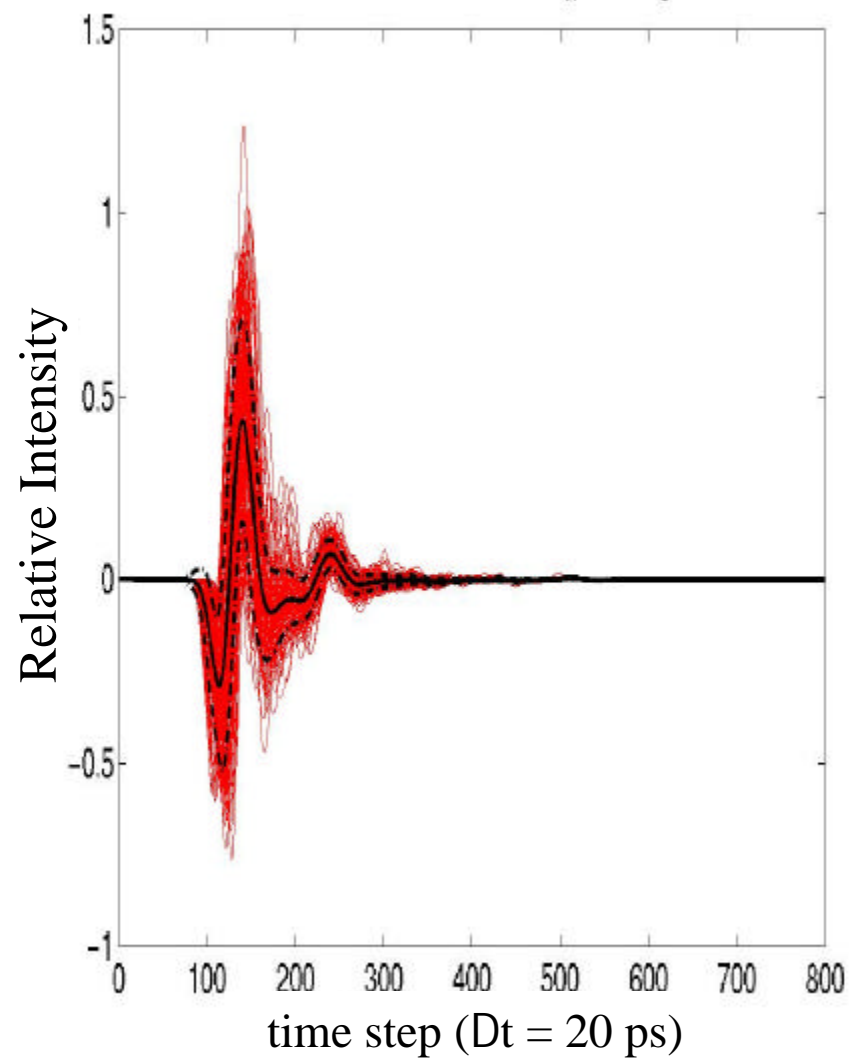
Figure 8. One hundred Monte Carlo calculations of the same physical situation as Figure 3, each processed by multiple subtraction of the average ground surface clutter, correlated in time to the sample signal, as indicated by the flowchart of Figure 7. Almost all the clutter is removed in the mean ground-only signal a), while for the buried mine simulations b), the mine signal is practically the only response. This algorithm is superior if there are sufficient independent views of the ground sample.

Figure 9. Mean and ± 1 standard deviation for 500 Monte Carlo simulations for randomly rough ground with Gaussian parameters a) $\sigma_h = 3\text{cm}$ and $l_c = 10\text{cm}$, b) $\sigma_h = 3\text{cm}$ and $l_c = 3\text{cm}$, c) $\sigma_h = 2\text{cm}$ and $l_c = 10\text{cm}$, d) $\sigma_h = 2\text{cm}$ and $l_c = 3\text{cm}$, e) $\sigma_h = 1\text{cm}$ and $l_c = 10\text{cm}$, f) $\sigma_h = 1\text{cm}$ and $l_c = 3\text{cm}$.

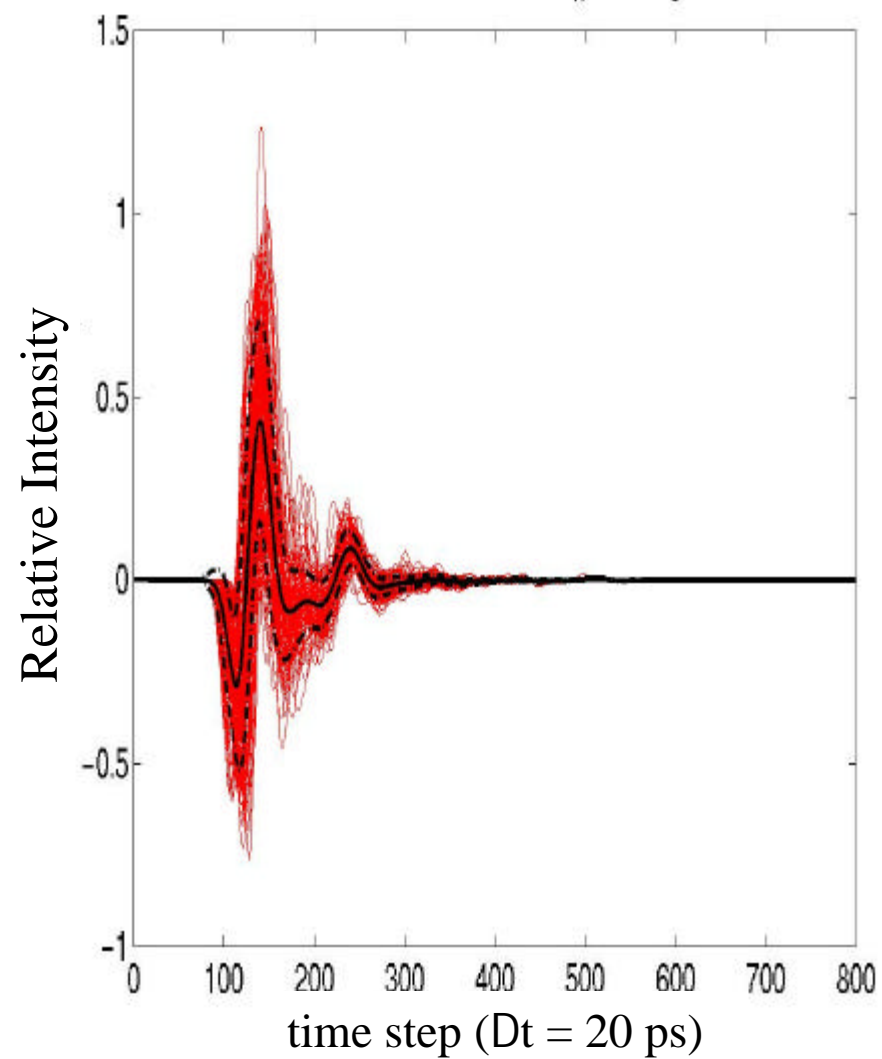


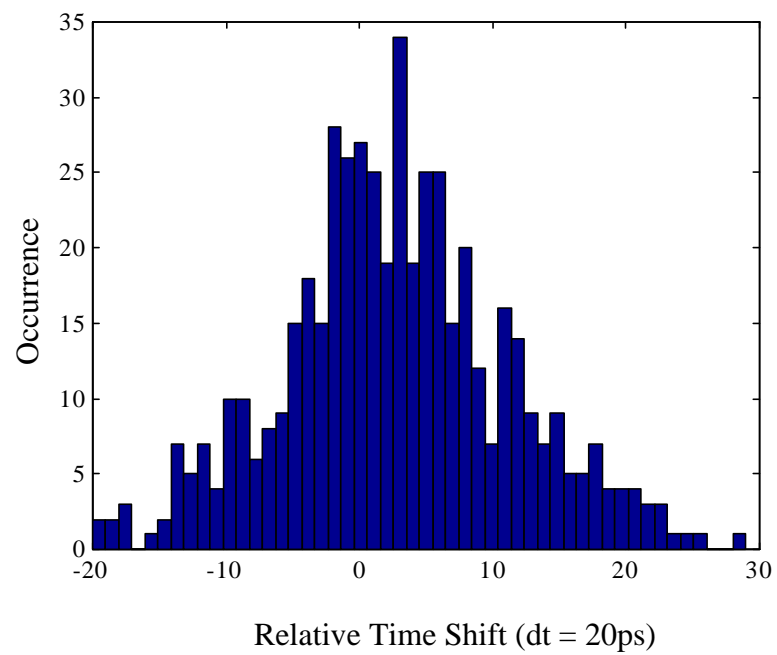
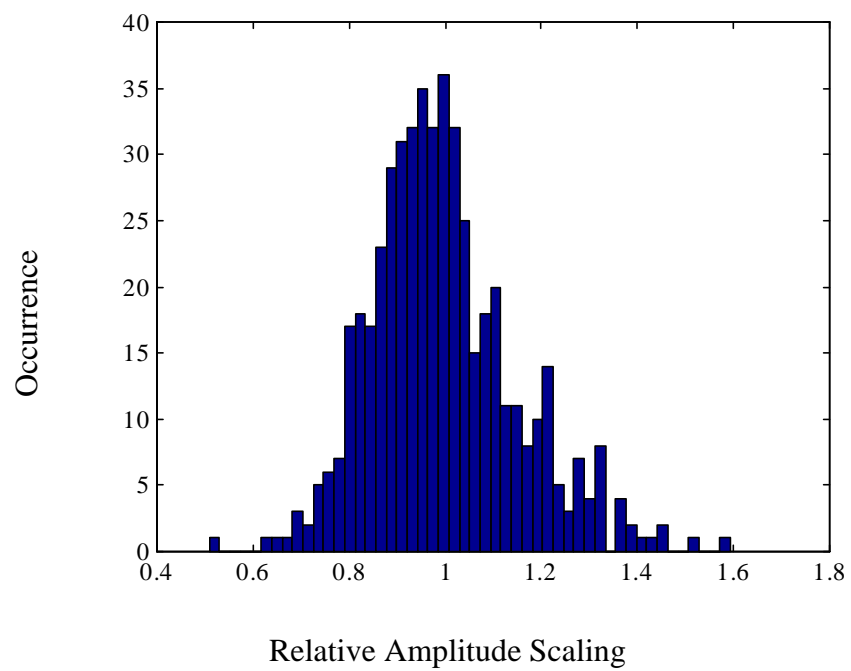


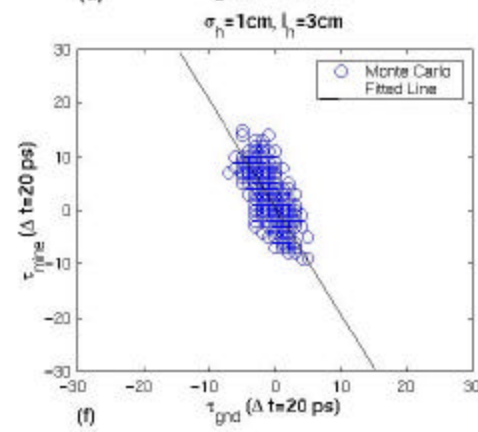
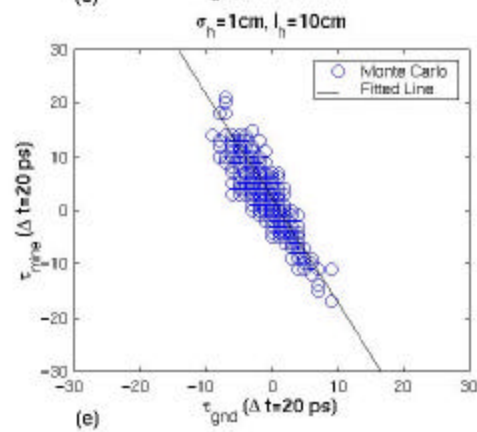
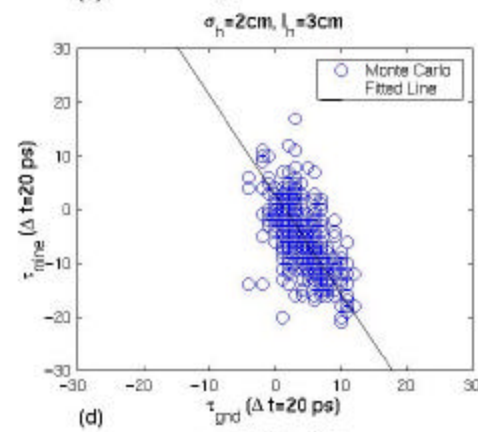
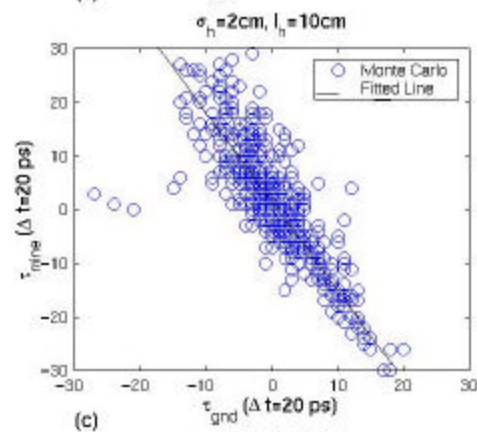
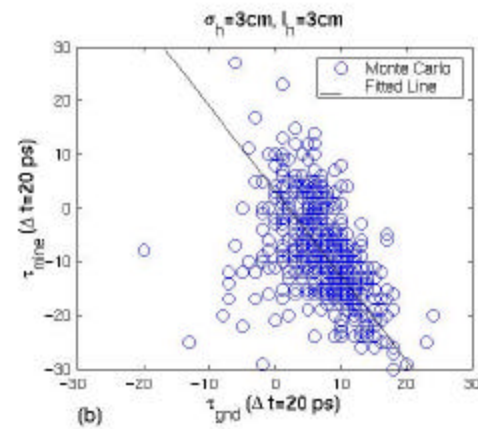
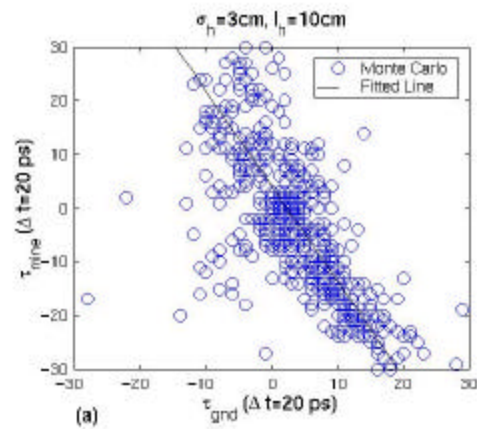
Original Nomine Signals with $\sigma_h = 3\text{cm}$ $l_c = 10\text{cm}$

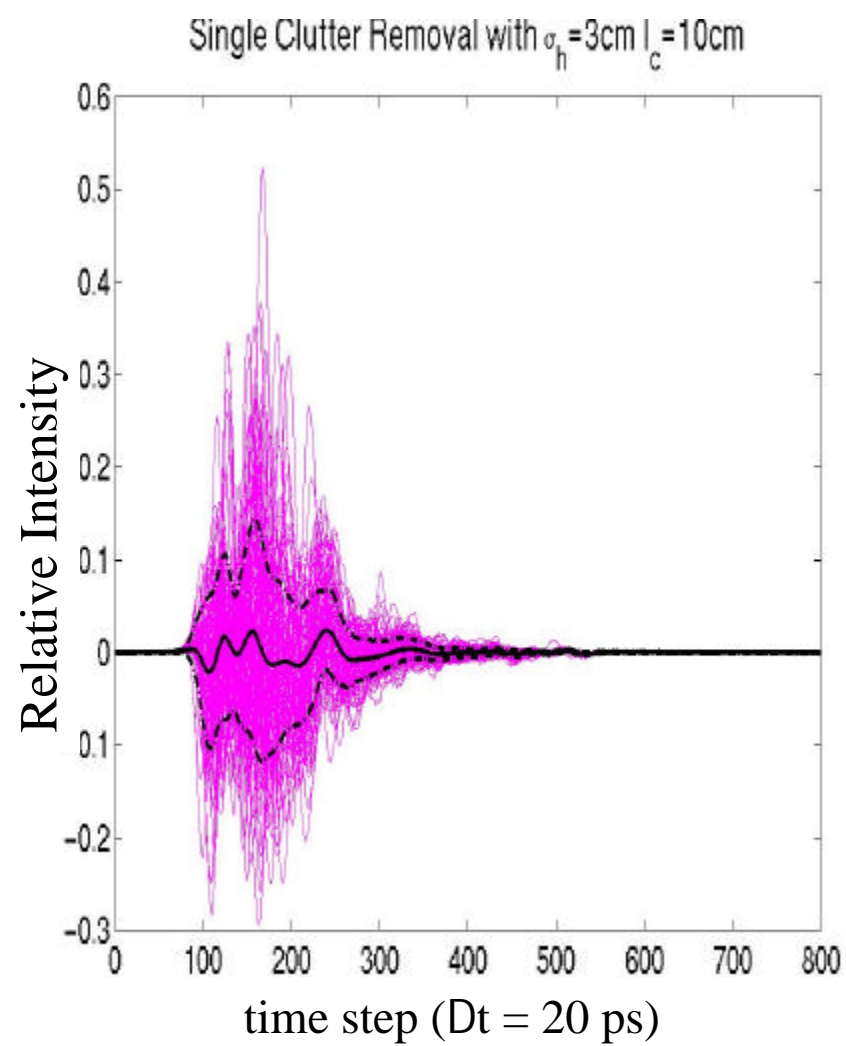
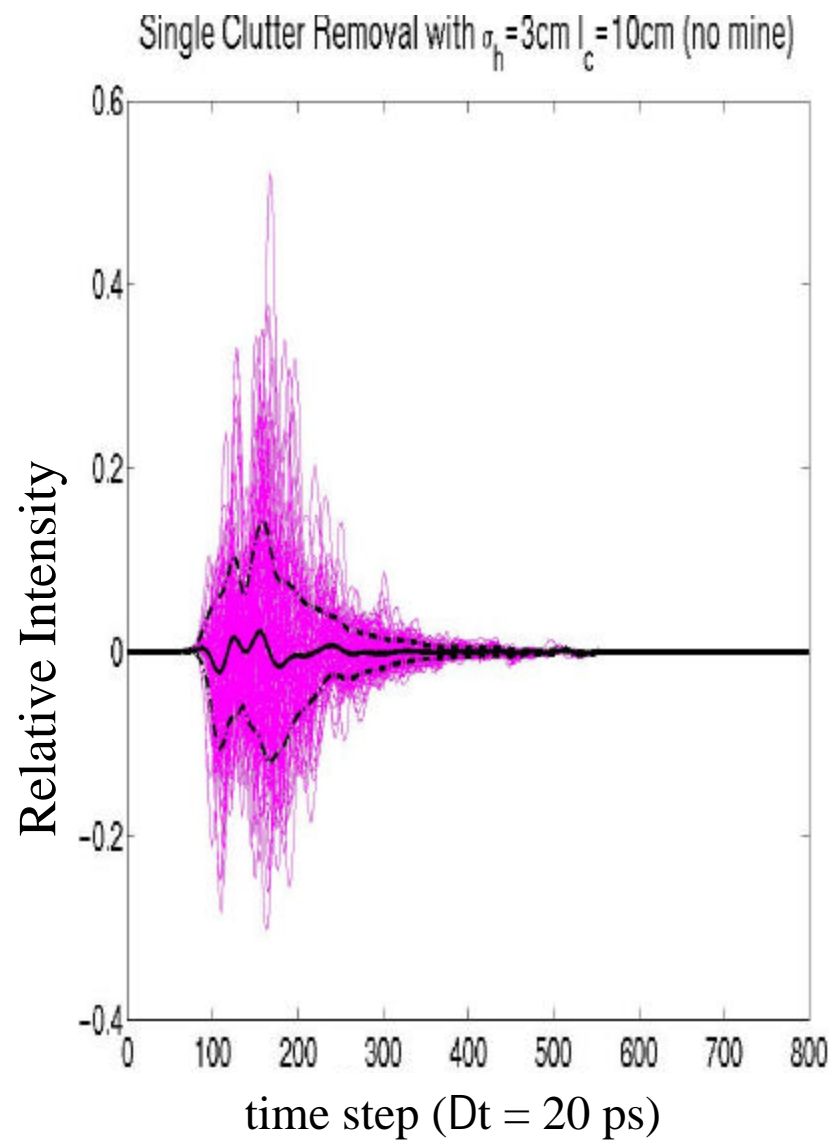


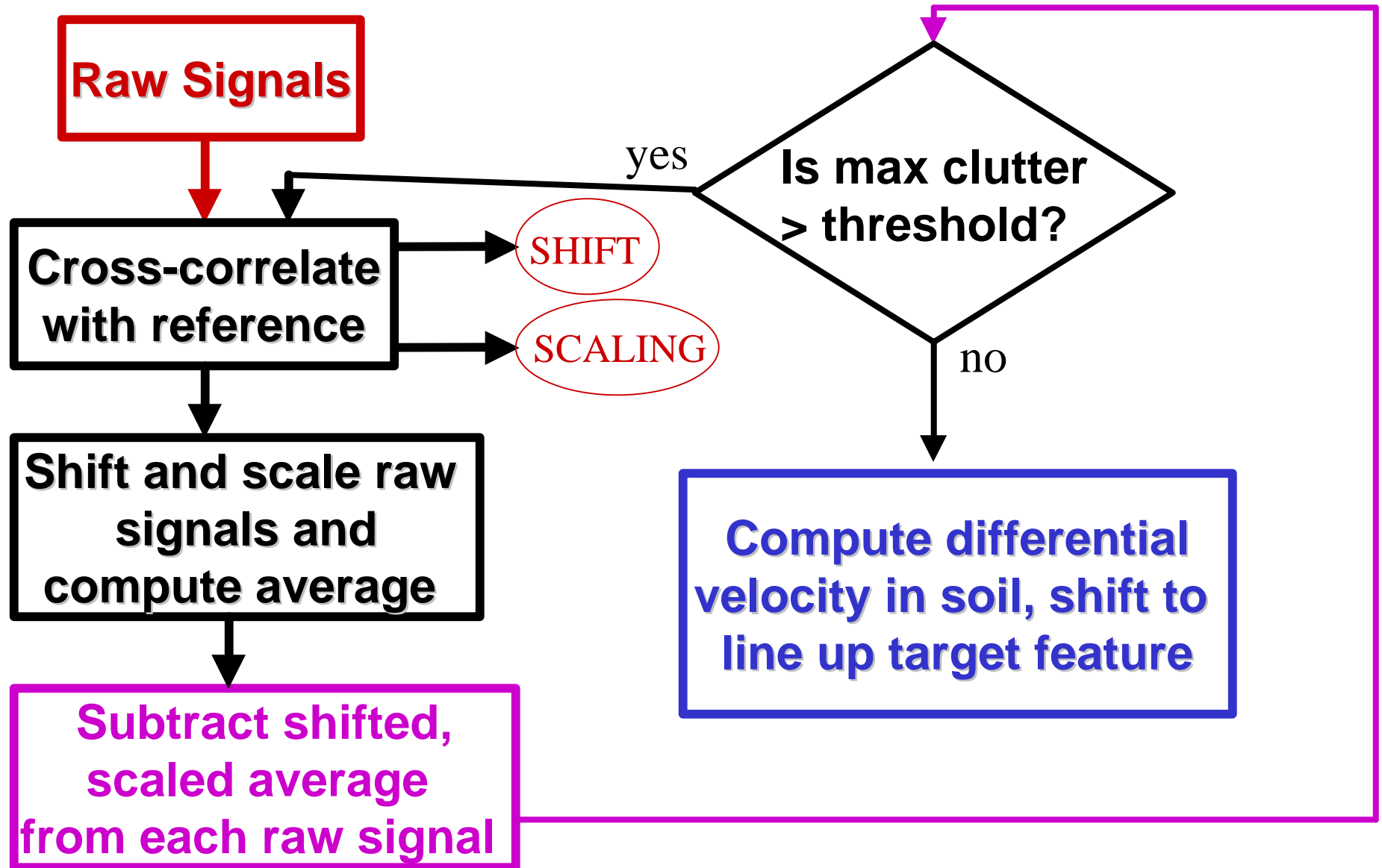
Original Mine Signals with $\sigma_h = 3\text{cm}$ $l_c = 10\text{cm}$

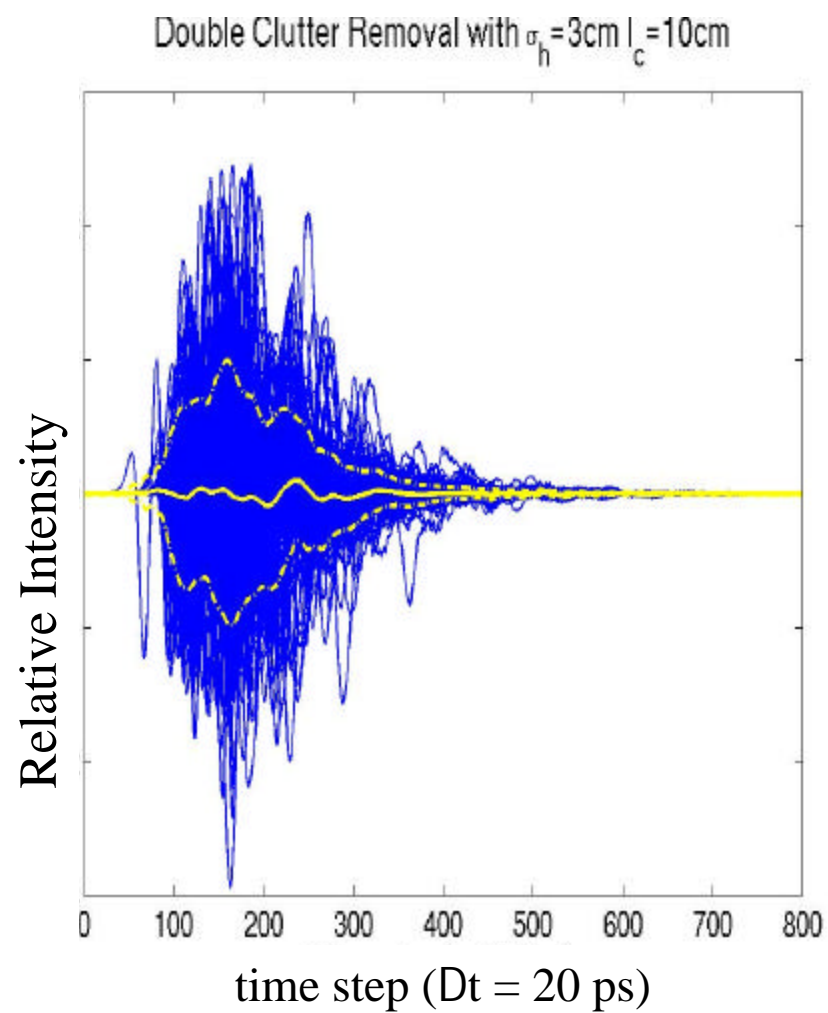
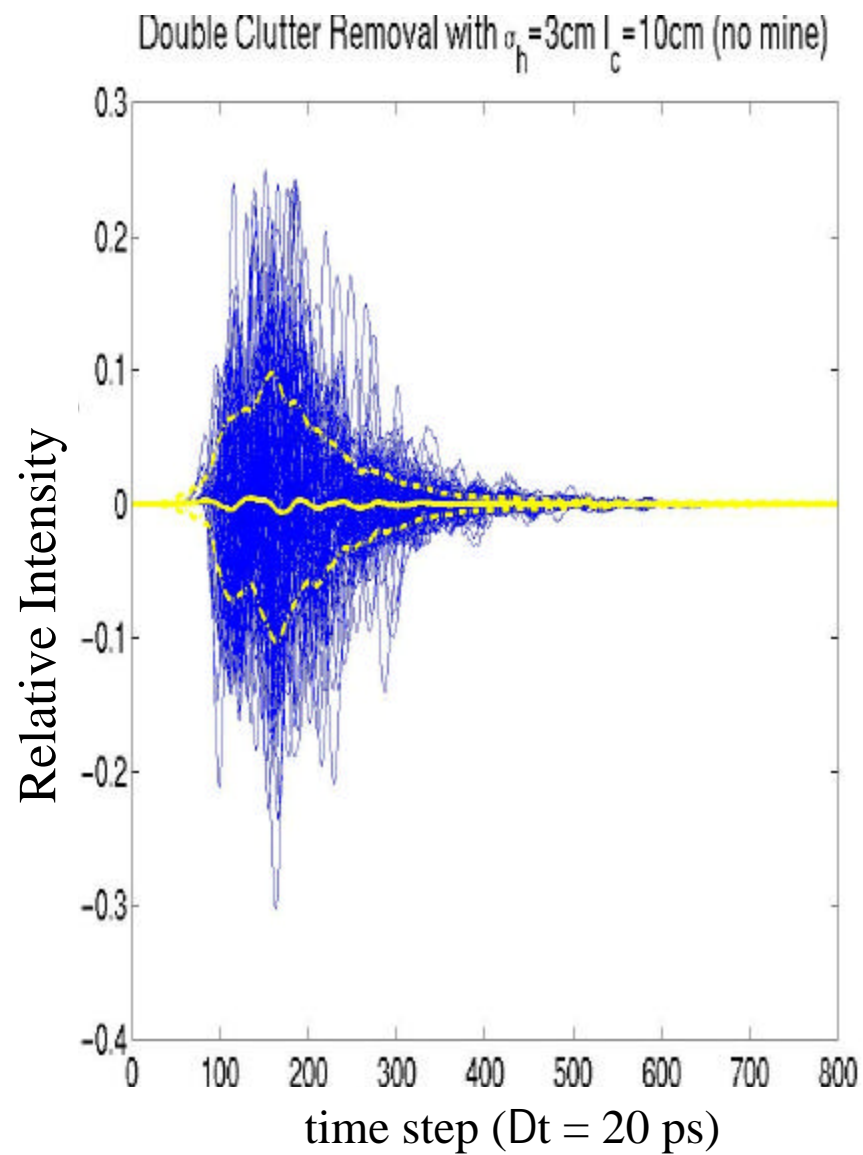


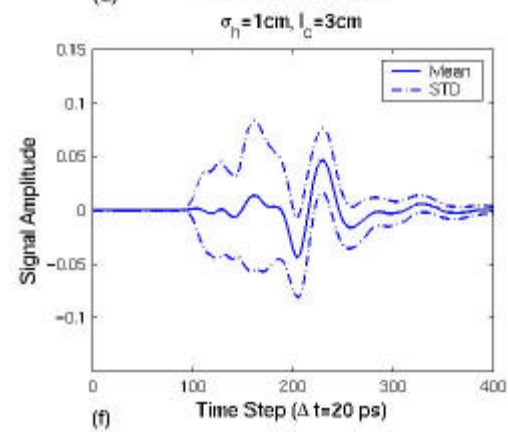
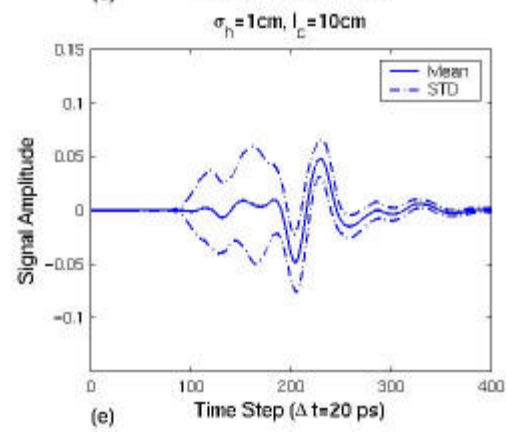
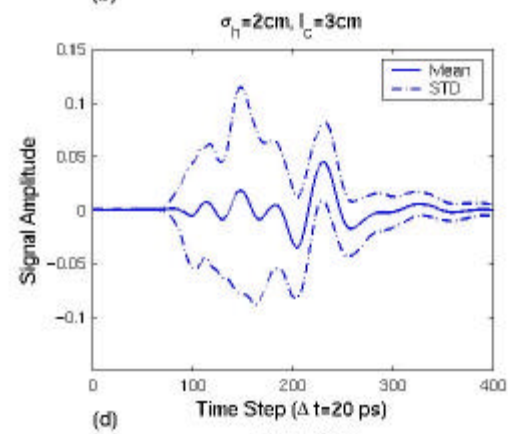
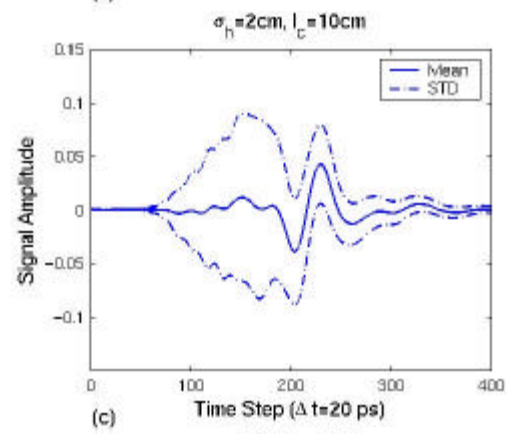
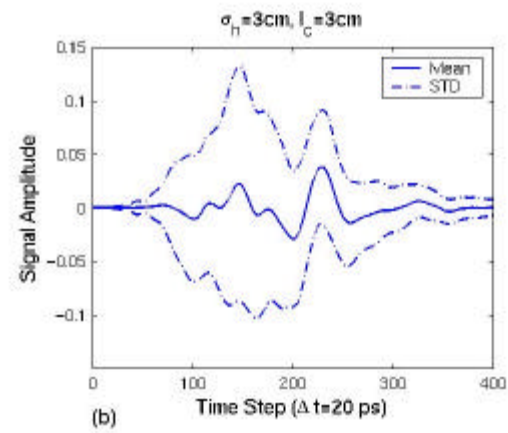
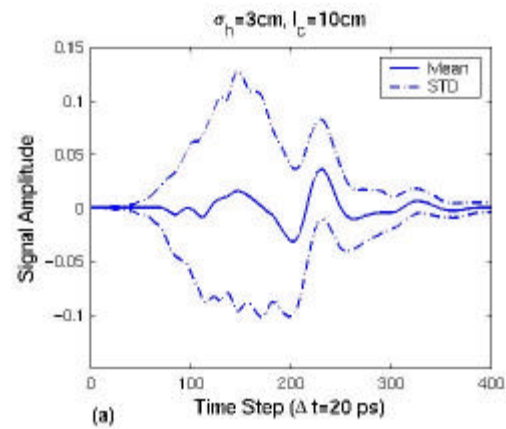












Imaging infrared polarimetry: initial results and potential in detection of scatterable mines and surface disturbances

Herman E. Scott, Stephen H. Jones and Frank Iannarilli
Aerodyne Research, Inc., 45 Manning Road, Billerica, MA 01821-3976¹

ABSTRACT

Over the past year with the support of the Army Humanitarian Demining MURI, Aerodyne has substantially moved forward in developing and demonstrating the value of an affordable and fieldworthy IR polarimetric hyperspectral imager for inclusion in multisensor demining. Such technology promises powerful clutter suppression and enhancement of man made objects, particularly applicable to the reliable detection of scatterable mines, especially plastics, and any UXO that are partially exposed.. We have achieved the first 3 steps of a 4 step, controlled-risk program defined as follows: (1) LWIR (non-imaging) Spectral Polarimeter to demonstrate the effectiveness of combined polarimetric and hyperspectral discrimination capabilities in observations on static scenes; (2) LWIR Uncooled FPA Imaging (broadband) Polarimeter to verify the sensitivity of an affordable Uncooled FPA in a broadband configuration against static scenes; (3) Multispectral Imaging Polarimeter to quantify clutter rejection performance improvements to be realized from multispectral imaging polarimetry; and (4) IR Polarimetric Hyperspectral Imager designed with optimal spatial and spectral resolution and sufficient throughput to achieve the reliable performance required in surface mine and UXO detection applications. We present results for Steps 1 and 2, and initial results for Step 3 from the ongoing demonstrations in simulated surface mine detection.

Keywords: Infrared polarimetry hyperspectral multispectral imaging demining,unexploded ordnance UXO microbolometer

1. INTRODUCTION

1.1. Background

For more than 15 years Aerodyne Research has worked systematically toward the development of passive IR spectral and polarimetric discrimination techniques. In the first 10 years of these activities Aerodyne focussed on the development and validation of a first principles computer model called POLAR^{1,2,3} that computes the spectrally dependent Stokes vector and user-selected in-band IR polarimetric images of targets given the geometry, optical properties⁴, and thermal description of the target. Support for the POLAR model came from Northrop⁵, the NADC/Warminster^{6,7,8,9}, and internal R&D. Then in recent years, emphasis has turned to the systems applications of polarimetric attributes such as non-cooperative target identification^{9,10} and the discrimination of targets in highly cluttered natural backgrounds.

In 1993 - 95 several important steps were accomplished: Aerodyne and the Boeing Defense and Space Group with the support of NAWC/Warminster used the modified IRAMMP sensor, POLIRAMMP, to collect quasi-polarimetric IR imagery of aircraft on the ground to validate the POLAR model.¹¹ Results showed a strong correlation between the analytic model predictions and measured data. Subsequently in 1994 Aerodyne produced a seminal paper¹⁰ on a polarimeter sensor concept (3-Channel Simultaneous Acquisition Polarimeter) specifically designed for the long range Counter-Air Target Identification application. Shortly thereafter, Boeing and Aerodyne applied the Improved POLIRAMMP sensor in the 1994 Navy Long Jump Tests at China Lake to collect surface-to-air polarization imagery of aircraft and thereby further validate the POLAR model.¹² It was at this test site that we conducted our first primitive simulation of surface mine detection using the two-channel POLIRAMMP sensor. Even though this test simulated only surface mine detection¹³, it provided quite convincing evidence of the clutter suppression capability of IR polarimetry.

In 1997 under a Phase I SBIR for the Air Force Research Laboratory (AFRL), Aerodyne invented a IR Polarimetric Hyperspectral Imager concept¹⁴ capable of providing the perfectly co-registered, polarimetric hyperspectral imagery desired for locating polarized objects in a cluttered background from a dynamic moving platform. We are in the process of building the prototype hyperspectral IR polarimeter under Phase II SBIR support from the AFRL. In a parallel project sponsored by

¹ Further author information -

Email: scott@aerodyne.com; www.aerodyne.com; TEL: (978) 663-9500; FAX (978) 663-4918

the Army Research Office (ARO) Humanitarian Demining MURI / Northeastern University (NEU) Team, we are taking incremental steps to demonstrate the value and reliability of LWIR spectral polarimetry in the detection of antipersonnel (AP) mines, especially plastic flush-buried and scatterable types.¹⁵ Our stepwise progress and the LWIR polarimetric spectral imaging proof-of-concept demonstrations in the demining application are the topics of this paper.

1.2. Scope

Many researchers have investigated LWIR and multiband passive infrared sensing techniques for applications in demining;¹⁶ a few investigations have considered passive IR polarimetry and costly hyperspectral imaging methods separately for locating land mines;¹⁷ yet insofar as we have found, no previous investigators have pursued the combination of hyperspectral IR polarimetry and affordable LWIR uncooled focal plane arrays. Effort in this Demining MURI is devoted to deliberate, controlled-risk, proof-of-concept demonstrations that verify our particular innovative combination of LWIR hyperspectral polarimetry and affordable uncooled FPA technology can be reliable and effective in surveying and locating exposed surface mines of any material type. The exploitation of polarimetric attributes that are demonstrated in this project effort were in most cases suggested and supported by our earlier analytic investigations into systems applications of IR polarimetry^{9,10,18} In addition to the proof-of-concept tasks, we are investigating surface optical properties of the relevant classes of materials and coatings for mine fabrication; this is a concurrent laboratory effort to fill the important gaps remaining from prior optical properties measurements.

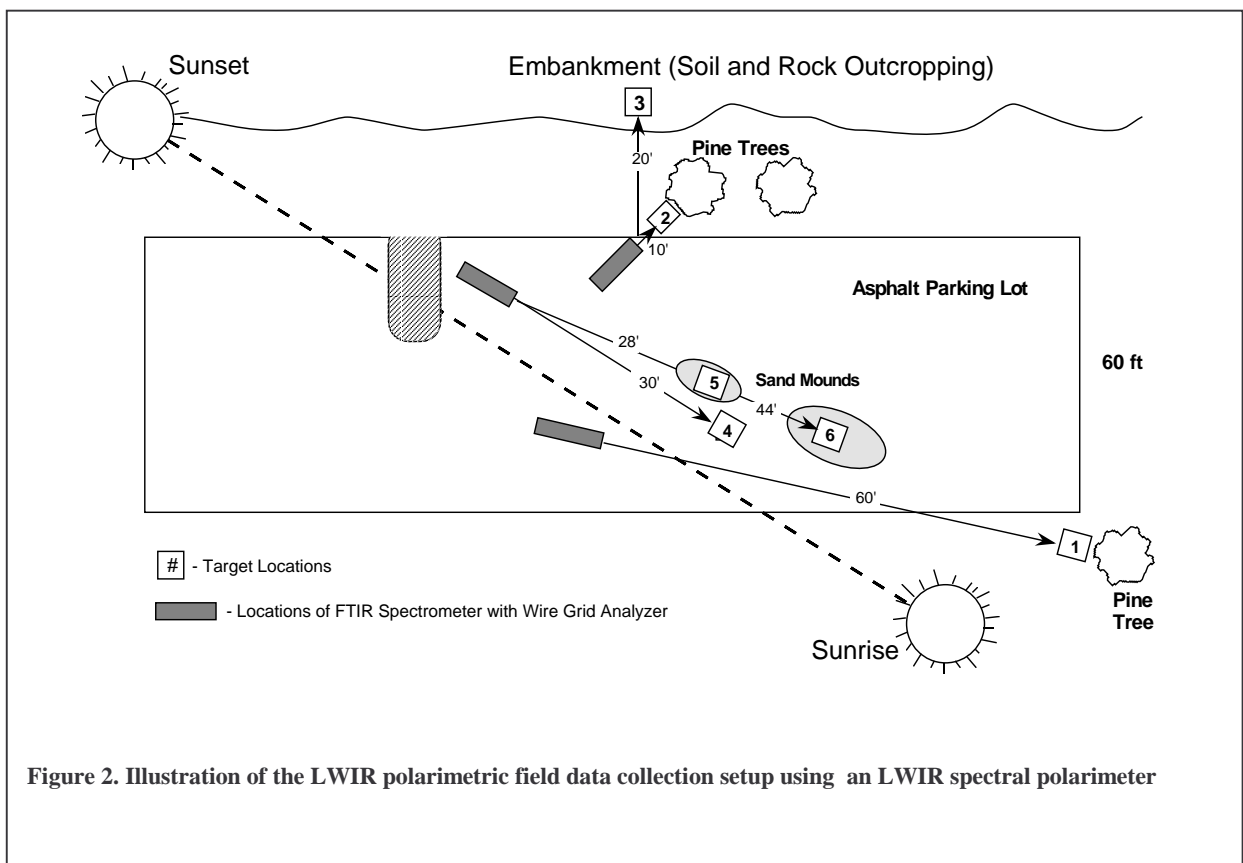
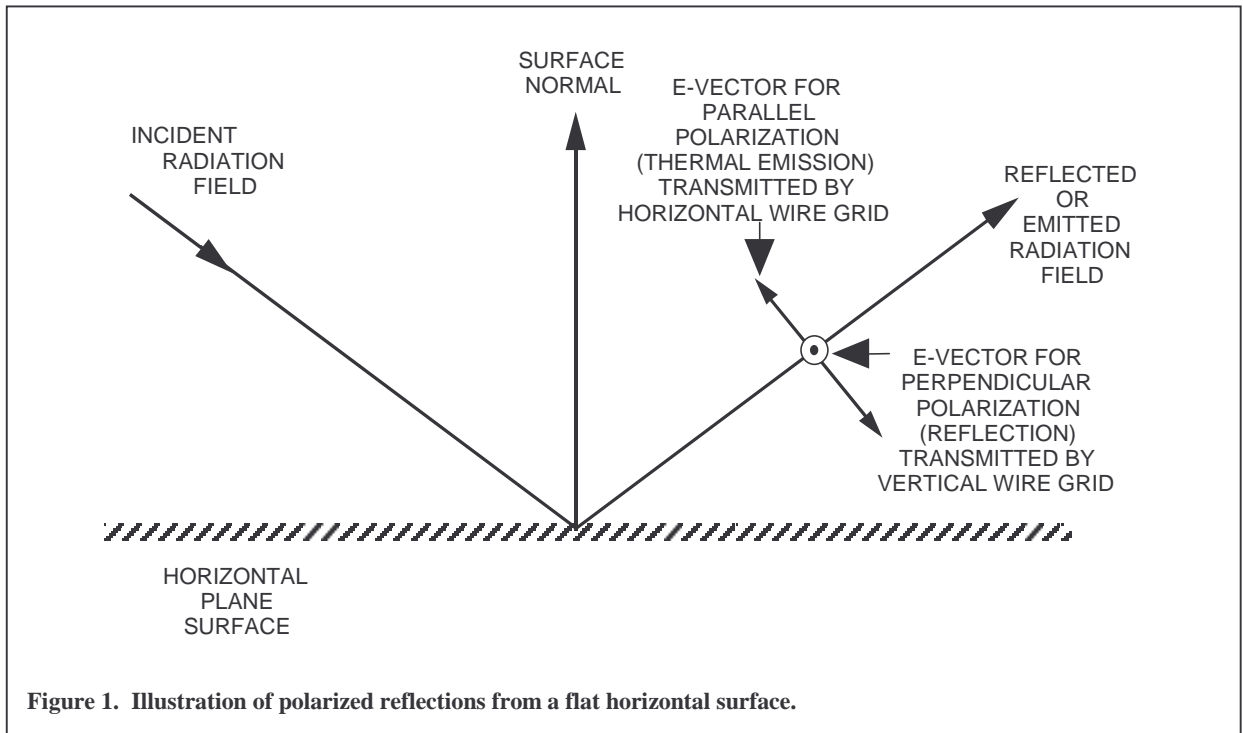
The present effort is focused on antipersonnel mines, many of which are partially exposed on the surface or are buried in a very shallow manner. Therefore, we are placing much greater emphasis on the optical properties of materials, emissivity and reflectance, and less emphasis than usual on the material thermal properties. We make no claims to be dealing with the full complexity of thermal conduction issues associated with buried mines, consisting of a wide range of materials in all manner of soils and terrain. Many investigators^{16,19} have examined the conductance and differential heating/cooling effects for opportunities to exploit in locating buried mines using passive IR techniques. Our premise is simple; IR spectral polarimetry is only one of a large arsenal of affordable and reliable technologies required to deal with the extensive, worldwide demining and unexploded ordnance problems. Our objective is to make IR spectral polarimetry an affordable and reliable technique to locate mines having any significant exposure on the surface, regardless of the material composition and including common types of paints and coatings. If we are successful, our method simultaneously positions us to include and exploit significant thermal effects as well since these effects exhibit themselves in the same LWIR wavelength regime. We must give close attention to the magnitude and impact of the heat transfer effects even as we examine the optical properties of materials and coatings.

Since we are implementing a controlled-risk, step-by-step effort to demonstrate a new combination of technologies including the uncooled focal plane array (UFPA) microbolometer, it is important that we establish the terminology for these steps and the technologies to be demonstrated:

- Step (1) - LWIR Spectral Polarimeter
(8 - 12 μm Spectral Region, Non-Imaging, Static Scenes).
- Step (2) - LWIR Uncooled FPA Imaging Polarimeter
(8 - 12 μm Broadband, No Spectral Resolution, Static Scenes).
- Step (3) - LWIR Uncooled FPA Multispectral Imaging Polarimeter
(Multiple Bands in 8 - 12 μm Spectral Region, Static Scenes).
- Step (4) - Uncooled IR Polarimetric Hyperspectral Imager
(8 - 12 μm Spectral Region, Spectral Resolution $\sim 0.02 \mu\text{m}$, Imaging,
Registration and Simultaneity in Polarimetric Channels for Dynamic Scenes).

2. LWIR SPECTRAL POLARIMETER DEMONSTRATION TEST

The data reported in this first proof-of-concept demonstration were collected for a variety of mine-like materials and coatings, including Army CARC (Chemical Agent Resistant Coating, both tan and green), a plastic frisbee, asphalt, and quartz sand. The backgrounds consisted of sand or sandy asphalt in all cases. The asphalt is a particularly challenging background, having a degree of polarization in the infrared that is comparable to or higher than many paints and plastics. Table 1 summarizes the data collection geometries and simulated mine target areas for each location. It is particularly



important to note the ranges and the grazing angles between the sensor (polarimeter) line-of-sight and the target normal. These geometries closely approximate a reasonable deployment geometry for an LWIR spectral polarimeter mounted on either a land vehicle with boom or a low flying UAV platform; however, the modified FTIR spectrometer used in this proof-of-concept demonstration was in no way related to our concept for a field deployable multi-channel polarimeter.

The LWIR spectral polarimeter consisted of an FTIR spectrometer with a wire grid analyzer adapted to it. For this demonstration we used 4 cm^{-1} resolution at 10 scans/sec and collected 200 scans per data point per analyzer angle. The detector was MCT with $D^*_{\text{peak}} = 4 \times 10^{10}\text{ cm Hz}^{1/2}\text{ W}^{-1}$. The wire grid analyzer was 1200 lines/mm on a ZnSe substrate with antireflection coatings on both sides.

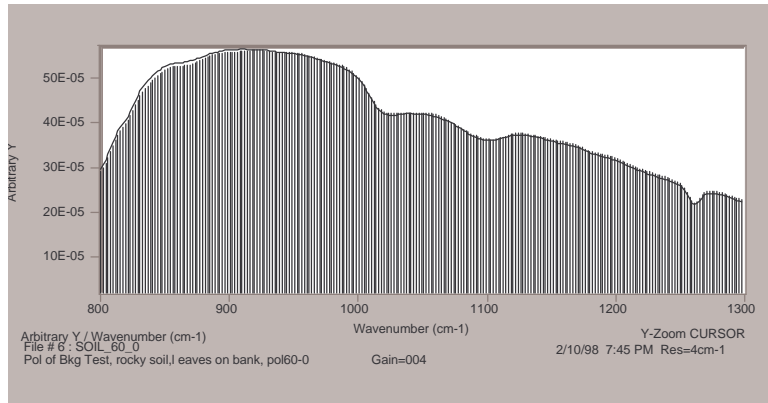
As illustrated in Figure 1, reflections from a flat surface induce polarization perpendicular to the plane of incidence (the plane of the page). Emission is polarized parallel to this plane.²⁰ Wire grid polarizers transmit radiation when the E-vector is perpendicular to the wire and reflect radiation when the E-vector is parallel to the wire. We use the convention of 0-degree polarizer angle to mean the wire grid is oriented vertically. Also, we use the spectroscopic notation $\sigma(\sigma)$ for wavenumber (cm^{-1}), and our plots will appear reversed to many readers. The materials used in this demonstration do not transmit in the spectral region of interest, so the simple relationship between the reflectance and emissivity of the materials applies, $\epsilon(\sigma) = 1 - \rho(\sigma)$.

This proof-of-concept demonstration was set up locally in a parking area (see Figure 2) surrounded by several large white pine trees and on the northwest side by a steep embankment of rock outcroppings, rocky soil and scrubs. The parking surface was asphalt, and in the winter months when these data were collected, the asphalt was lightly covered with a residue of sand. The data of interest were collected at locations 4 and 5 as shown in Figure 2 and consisted of fifty polarimeter spectra for the various combinations of target, background, target angle, wire grid analyzer angle, ambient temperature and cloud cover. Again, the setup geometry details are listed in Table 1. Most of the data for these locations were collected in the late afternoon, both before and after sunset, on two successive days with the ambient temperatures between 0- and 15-C. Each spectrum required 20 sec to collect. Since the data collection time was not a primary driver in this demonstration, we made no effort to minimize the time, but it is quite clear that the SNR achieved would allow the collection time to be less than 2 sec per data point per analyzer angle with no important degradation of the results.

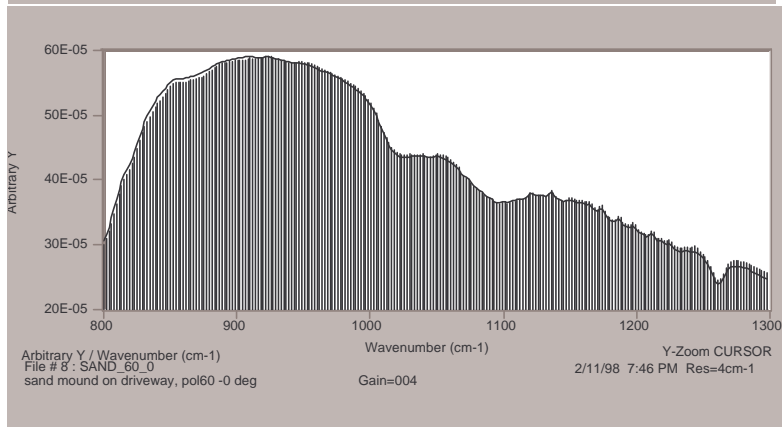
2.1. Interpretation of the LWIR spectral polarimeter data

For a view of the spectral data and an understanding of the analysis procedures, we show several overlays of the spectral data in Figures 3 and 4, each at two angles of the wire grid analyzer. As called for in our analysis procedure discussed below, data were normally collected for wire grid analyzer angles of 0-deg (vertical wire grid) and 60-deg; for verification and redundancy we also collected data intermittently at wire grid orientations of 90-deg and 120-deg. Our convention in this analysis is to use the 0-deg and 60-deg wire grid angles as the orthogonal components to differentiate the reflected and thermal emitted radiation components coming from the target and background sources within the field-of-view. The spectra are plotted with the envelope of the shaded area being the spectrum for 0-deg analyzer angle; this spectrum represents the polarized reflection component in the 'horizontal earth' geometry selected for the demonstration. Relative to this the predominantly emission component, passed by the wire grid analyzer in the 60- or 90-deg position, is plotted as a thin solid line lying close to and just above the envelope of the shaded area. The white empty band area, or the difference, between these two envelopes is a measure of the degree of linear polarization (DoLP) of the spatially unresolved scene in the field-of-view of the spectral polarimeter. Note that the DoLP and hence the polarimetric SNR of a spatially resolved mine target will typically be higher than this average of the entire scene, proportionate to the target/scene fill factor. When the scene radiation is totally unpolarized, the orthogonal polarimetric component spectra are coincident over the full spectral range of sensitivity for the polarimeter. Here the spectral region of interest is $800 - 1250\text{ cm}^{-1}$ (or $8 - 12.5\text{ }\mu\text{m}$). The general procedure for extracting scene DoLP from the spectra is given below, and the results are plotted in Figure 5.

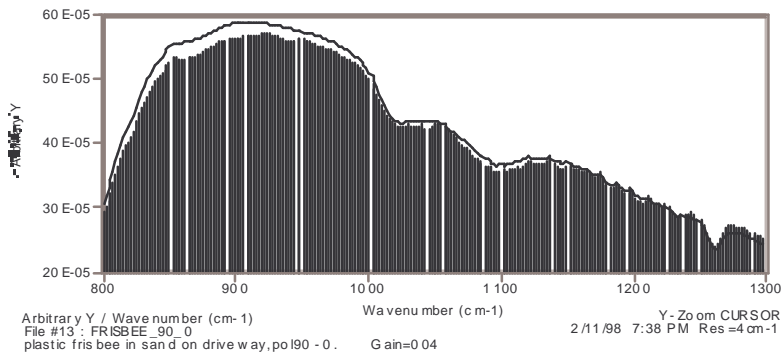
Alternatively, the white empty band in between the shaded area envelope and the line plot above represents the spectrally dependent, available polarimetric signal to be exploited. For any given broadband polarimeter the average area under the 2 orthogonal polarizer orientation spectra such as in Figure 3(a) is a measure of the total signal while the difference between these integrated orthogonal polarizer spectra is a measure of the scene polarization signal, the fraction of scene radiation that is linearly polarized. From the integrated areas measured for the materials in Figures 3(c), 3(d) and 4(a), we find the adjusted fraction of target radiation that is linearly polarized when extrapolated to fill the entire FOV of the sensor to be of the order of 10%.



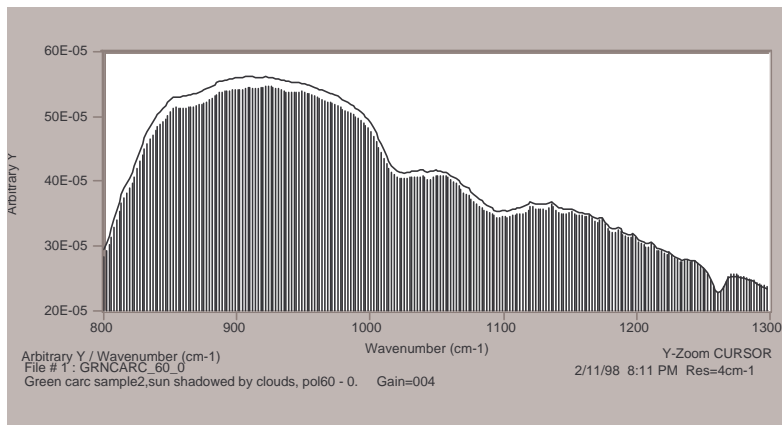
(a) SOIL



(b) SAND



(c) FRISBEE ON SAND



(d) CARC ON SAND

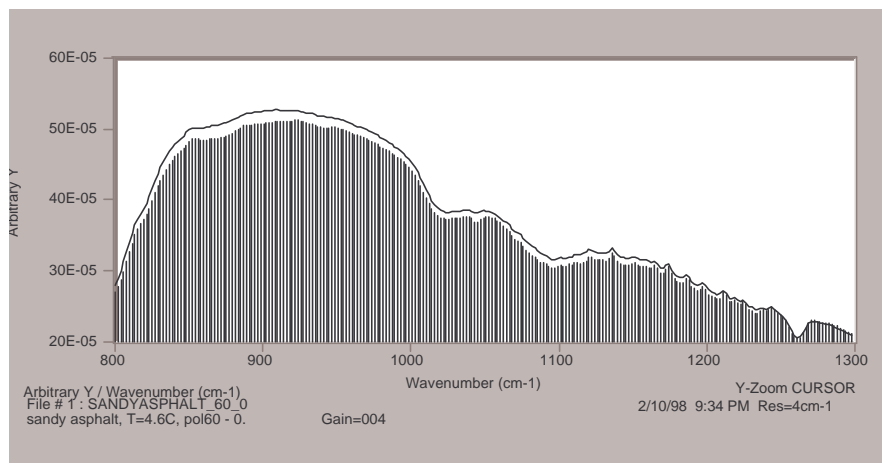
Figure 3. Raw data plots from spectral polarimeter, each at two analyzer angles, for four scenes. The envelope of the shaded area is the spectrum for analyzer angle 0-deg, and the solid line for analyzer angle 60-deg.

Table 1. Summary of the Data Collection Geometries and Results of the LWIR Spectral Polarimeter Measurements for the Target and Background Materials at Locations 4 and 5.

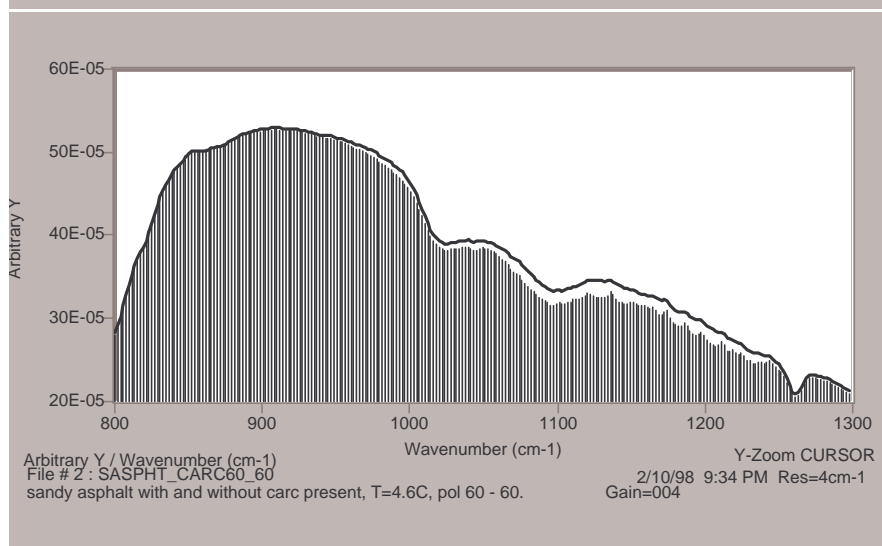
Target Name and Location	Sandy Asphalt Location 4	Tan CARC Location 4	Plastic Frisbee Location 5	Green CARC Location 5
Plane of Target Surface Relative to Vertical (deg)	86	66	85	79
Background Type	Sandy Asphalt	Sandy Asphalt	Sand Mound	Sand Mound
Range (ft)	30	30	28	28
Target Total Area (sq in)	737	64	33 Exposed	56
Angle Between Sensor LOS and Target Surface Normal (deg)	79	59	75	69
Target Area Projected into FTIR FOV (sq. in.)	140	33	9	20
FTIR Weighted FOV Scene Footprint at Range (sq. in.)	140	140	130	130
Fractional Area of Total Weighted FOV Occupied by Projected Area of Target	Do Not Know Frac of Asphalt Covered by Sand	0.24	0.07	0.15
DoLP of <u>Scene</u> , adjusted to remove any DoLP of Sandy Asphalt, for the Spectral Region 800-1250 cm^{-1}	0.02	0.02	0.01	0.02
Percent of <u>Target</u> Radiation that is Linearly Polarized When Extrapolated to Fill the FOV of Sensor	----	8%	14%	13%

Upon closely examining the polarimetric component spectra in Figures 3(a) and 3(b), we find that these 0-deg and 60-deg spectra are coincident over the 800 - 1250 cm^{-1} region for soil and sand near normal incidence, so we may expect the degree of linear polarization (DoLP) for these background materials to be very small or zero. Below, we compute the DoLP spectra and indeed show these background materials are substantially unpolarized.

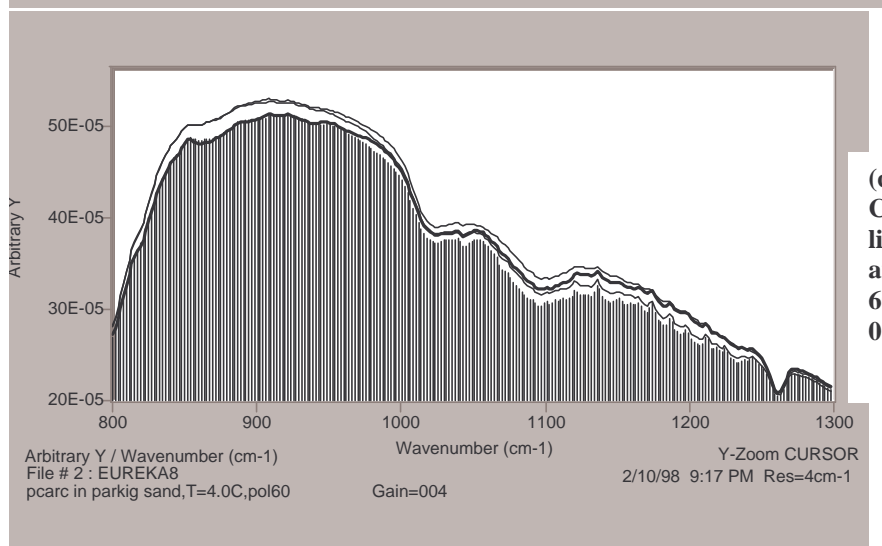
This is not the case, however, when we examine the polarimetric component spectra for the plastic frisbee on sand, CARC on sand, and sandy asphalt by itself in Figures 3(c), 3(d) and 4(a), respectively. In Figures 3(d) and 4(a) we find large separations in the polarimetric component spectra indicating that CARC and asphalt exhibit strong LWIR polarization effects. Furthermore, the separations in CARC and asphalt show an interesting spectral dependence, being very strong near the greybody curve peak at 925 cm^{-1} and gradually narrowing until the gap closes at about 1250 cm^{-1} . While the general shape of the spectra are determined by the greybody temperature and the spectral polarimeter response function (including the optics, analyzer, antireflection coatings and detector), the sharp spectral features in the 1100 - 1200 cm^{-1} region are attributed to quartz sand. Notice these features are not present in the soil spectrum, Figure 3(a). These spectra, rich in features, demonstrate very clearly our interest in the LWIR and our rationale for pursuing hyperspectral polarimetry as a powerful technique to discriminate man made target materials against natural background clutter.



(a) Sandy Asphalt
(Shaded=0-deg,
Line=60-deg)



(b) Sandy Asphalt
with CARC at 60-
deg (line), and
without CARC
(shaded)



(c) Sandy Asphalt with
CARC at 60-deg (top
line), 0-deg (bold line),
and without CARC at
60-deg (mid line) and
0-deg (shaded)

Figure 4. Spectral polarimeter measurements at the noted analyzer angles for sandy asphalt with and without CARC panel present. The relative intensity for the same analyzer angle is always greater with the CARC panel present in the FOV.

2.2. Analysis of polarimetric measurements

The degree and angle of linear polarization (DoLP and AoLP) of a *partially linearly polarized* wave can be obtained from measurements made at three different analyzer angles. The intensity measured at the output of the analyzer oriented at angle α is given by^{10, 20}:

$$I(\alpha) = 1/2 I_u + I_p \cos^2(\theta_p - \alpha) \quad (1)$$

where

$$\begin{aligned} I_p &= \text{Polarized light from target} \\ I_{u-tgt} &= \text{Unpolarized light from target} \\ I_{u-bkg} &= \text{Unpolarized light from background (here assumed unpolarized)} \\ I_u &= I_{u-tgt} + I_{u-bkg} \\ \text{DoLP} &= I_p / (I_{u-tgt} + I_p) = \text{Degree of linear polarization of target} \\ \theta_p &= \text{Angle of linear polarization.} \end{aligned}$$

As is frequently the case with simultaneous trigonometric equations, general solutions in terms of well known functions are difficult to find; a numerical solution can be obtained by searching for it using Newton's method based on minimizing an objective function such as the sum of the squared differences between the measurements and the theoretical value (i.e. left hand side minus right hand side of Eq (1)). The field measurements obtained at Aerodyne were performed such that the surface normal vector of the targets was in the plane of incidence, thereby constraining the emitted light to a 90-degree analyzer angle and the reflected light to a 0-degree analyzer angle. Careful attention was paid to orienting the spectral polarimeter such that cold sky was reflected. Under these circumstances, it is reasonable to make the simplifying assumption that $\theta_p = 90$ degrees (i.e. the reflected component of I_p is negligible). Equation (1) then becomes

$$I(\alpha) = 1/2 I_u + I_p \sin^2(\alpha) \quad (2)$$

By choosing an analysis angle of 0-degrees, the unpolarized intensity is immediately obtained. The polarized intensity can then be obtained by using any other angle (albeit within the limitations of SNR). In the present case of 90-degree angle of polarization, the greatest signal to noise ratio is obtained by using 90-degrees for the second analysis angle. A 1.25 dB loss in SNR results from the use of 60-degrees instead of 90 in this case.

2.3. Results

Using the simplified analysis approach represented in Equation (2), we extracted a quick-look approximate "residual scene DoLP spectrum" for each of the materials observed in this demonstration, presented in Figure 5. Since the quartz sand background was present in varying yet nearly equal abundances within the field of view for 4 of these data, a constant intensity spectrum of sand (corresponding to 100% fill of the field of view) was subtracted in each case. Each residual scene DoLP spectrum thus is an approximate unnormalized (for target fill factor) target material DoLP spectrum. Other features in these DoLP spectra deserving comment are:

1. Soil is effectively unpolarized and provides a point of reference for the other materials plotted in Figure 5.
2. CARC and sandy asphalt exhibit essentially the same DoLP, indicating they will not be easily distinguished using a broadband polarimeter. Refer ahead to view the CARC panel on sandy asphalt in Figure 9.
3. In the 1000 – 1100 cm⁻¹ region the trace effect of ozone remains observable in the spectrum.
4. Fringes apparent in the spectra are attributed to a polarization sensitive beam splitter effect in the FTIR spectrometer.

2.4. The special case of CARC target on sandy asphalt background

The combination of the CARC target imbedded in a sandy asphalt background is a key piece of data from our proof-of-concept demonstration. The reason for this can be seen in Figure 5. Over the spectral interval 800 - 1250 cm⁻¹ the measured DoLPs are the same for CARC on sandy asphalt and sandy asphalt alone; CARC and sandy asphalt cannot be differentiated solely by the broadband linear polarization attribute, i.e. by a single broadband imaging polarimeter. In Section 4 we show image data to confirm this. However, this case provides an excellent example of the powerful tool obtained when the

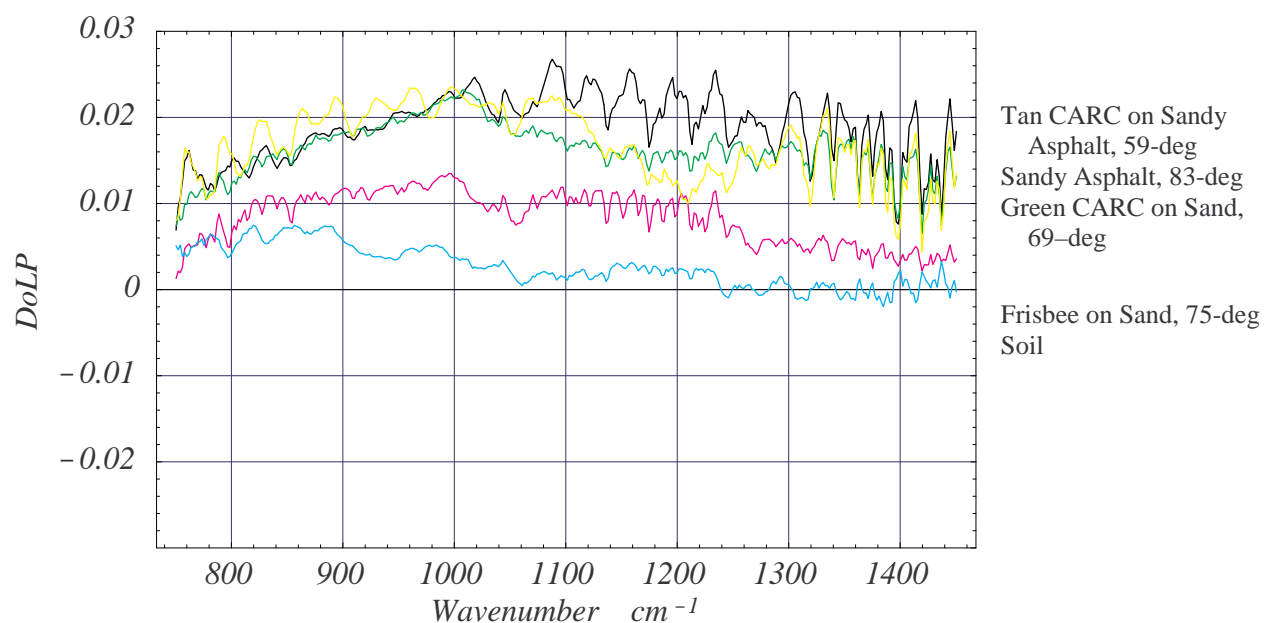
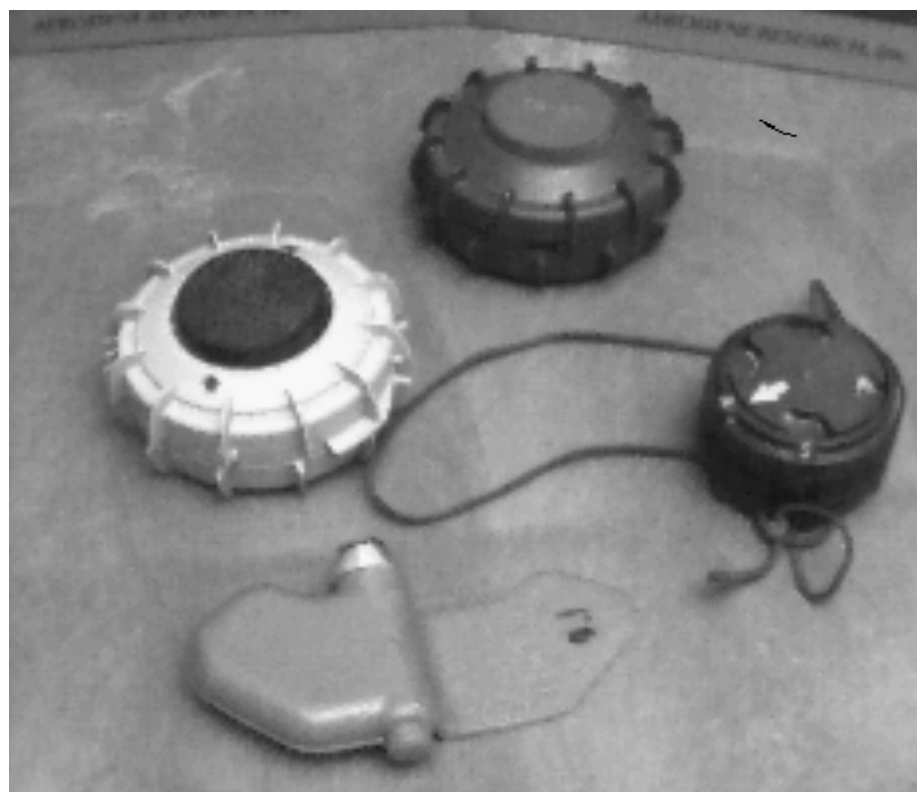


Figure 5. A plot of the residual scene degree of linear polarization (DoLP) for 4 scenes, and soil-only. A constant sand background intensity, with varying abundances in the FOV for the four scene measurements, has been subtracted as a quick-look means of isolating the approximate target material spectrum. The angle noted is between the sensor LOS and the material surface normal.



TS-50

VS-50

M-14

PFM-1

Figure 6. Inert antipersonnel mines used in grazing angle reflectance measurements and in imaging polarimeter field demonstration

polarimetric and spectral capabilities are combined in an LWIR spectral polarimeter. One set of the data for tan CARC on sandy asphalt collected with our basic combination of FTIR spectrometer and wire grid polarizer are shown on a magnified scale in Figures 4(a), 4(b) and 4(c). Figure 4(c) is an overlay of four plots - two for the 'CARC present' and two for 'only sandy asphalt'. By inspection of their spectral dependence we can easily distinguish the presence of CARC in the sandy asphalt background. At 900 cm^{-1} the 0-deg and the 60-deg analyzer plots are identical with and without the CARC present; no differentiation here. However, as we move along the wave number axis toward 1200 cm^{-1} , two effects are evident. First, the separation by analyzer angle for each pair of curves (with and without the CARC present) is becoming less and washes out completely by 1250 cm^{-1} (see Figure 4(c) for this to be apparent in the 'CARC present' case). Secondly, the 'CARC present' traces are decidedly higher than 'without CARC' plots over the spectral region $1050 - 1250\text{ cm}^{-1}$. This effect is shown clearly in Figure 4(c). Therefore, in the spectral dimension (at $1050 - 1250\text{ cm}^{-1}$, not 900 cm^{-1}) we have a very strong differentiation of the CARC present in the sandy asphalt background even though the DoLP for the combined CARC and sandy asphalt background afforded no differentiation. This is true even though separately both the CARC and the sandy asphalt exhibit strong polarization effects. These data make a clear case for the importance of adding the spectral dimension to the LWIR polarimeter. Our goal in the demining application is to demonstrate an affordable and effective polarimeter having both spectral and imaging capabilities.

3. LABORATORY MEASUREMENT OF GRAZING ANGLE IR SPECULAR REFLECTANCE OF ANTIPERSONNEL MINE CASINGS

In order to exploit unique spectral features (particularly low emissivity regions) of land mine casings and similar man made objects (typical plastics such as ABS and blended polymers) we must know the spectral features of both the man made materials and the natural backgrounds in these spectral regions. Our laboratory measurement of grazing angle specular (80-deg) reflectance of four AP mine casings and several related, simulant materials of interest confirm the availability of many LWIR spectral features to be exploited in multi- or hyper-spectral polarimetry. These reflectance spectra exhibit the characteristic dispersion curve shape near the centers of absorption features in the materials.

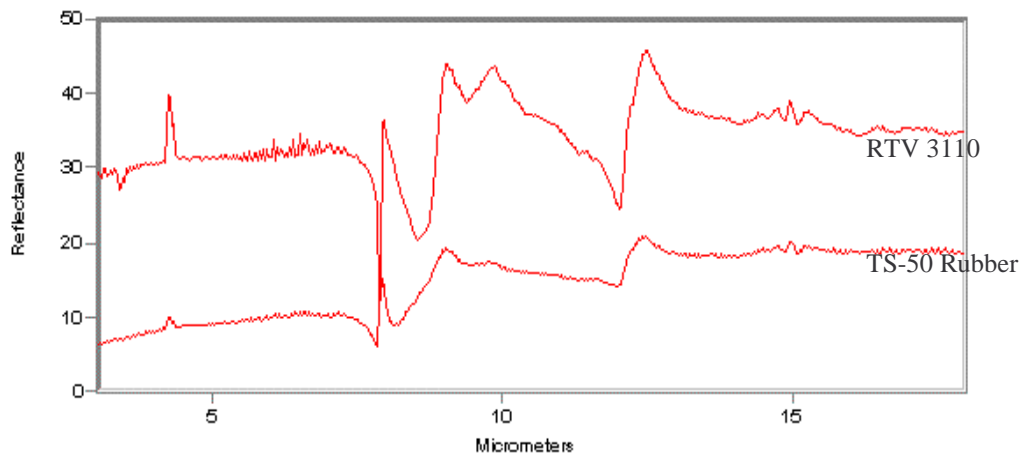
Our laboratory grazing angle reflectance spectra were collected with an FTIR spectrometer at a resolution of 4 cm^{-1} and generally covering the range of 3 - 15 microns ($666 - 3333\text{ cm}^{-1}$). In each case the sample spectrum has been ratioed to the spectrum of a front surface gold mirror that has a reflectance greater than 0.99 over the entire spectral range covered here. The inerted AP mine casings that were available for study are listed in Table 2 and pictured in Figure 6. In addition, Table 2 includes several related or simulant materials of interest for which we collected reflectance spectra, but not all of these will be discussed here.

Sample grazing angle reflectance spectra are plotted in Figures 7(a), (b) and (c). To make the examination of these spectra easy it is important first to eliminate from further consideration all spectral features attributed to the incomplete removal of water vapor and carbon dioxide from the optical path in the laboratory spectrometer; even though the path was constantly under purge with dry nitrogen, this did not completely remove all atmospheric gases from the sample compartment of the instrument. The spectral features to be ignored are water vapor bands at 2.7 and 6.3 microns and carbon dioxide bands at 4.3 and 15 microns. Then it is important to remember that the dispersion features in the reflectance spectra mark the locations of absorption features for the material. Since the material does not transmit radiation at these characteristic absorption wavelengths, the grazing angle reflectance, like the index of refraction, shows a strong change across the width of an absorption feature.

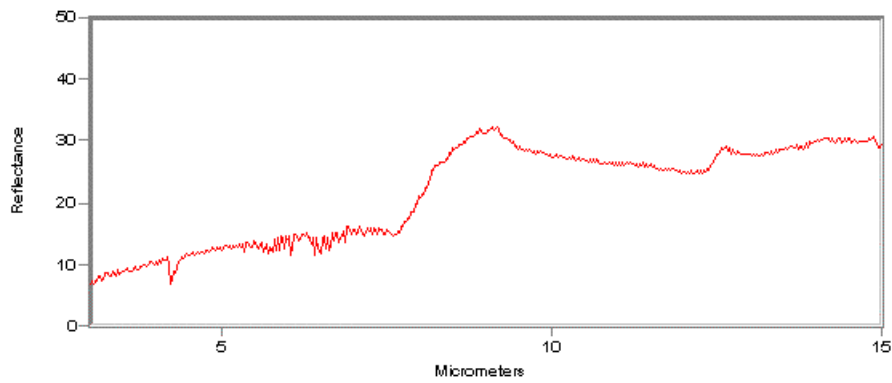
Our hyperspectral polarimetry approach to developing a reliable detection method for partially exposed scatterable land mines exploits differences in the emissivity and

Table 2. List of Anti-Personnel Mine Casings and Related Materials Included in Laboratory Measurement of Grazing Angle IR Reflectance Spectra.

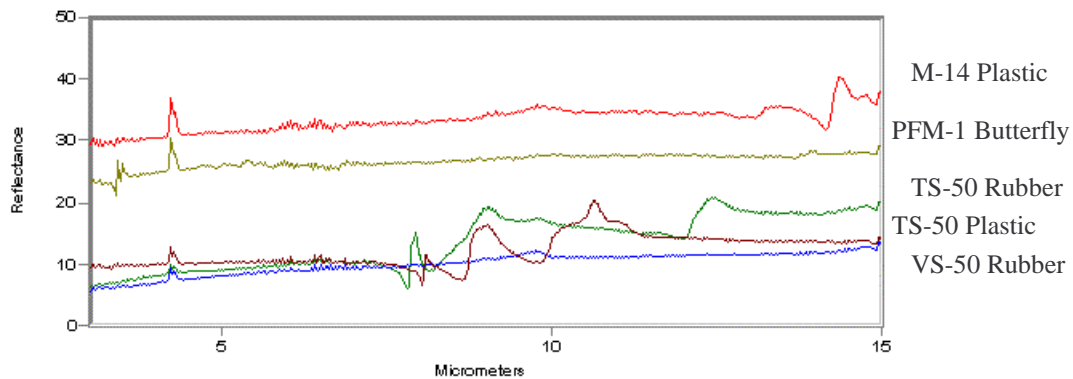
Anti-Personnel Mine Component	Related Materials or Simulants
M-14 Plastic	CARC (Chemical Agent Resistant Coating), Both Green and Tan
PFM-1 (Butterfly)	
TS-50 Rubber Center	RTV 3110 Silicone Rubber
TS-50 and VS-50 Plastic	Floppy Diskette Plastic Shell
VS-50 Rubber Center	Rubber Hockey Puck
	Asphalt
	Sand
	Soil



(a) Comparison of grazing angle reflectance of RTV 3110 and TS-50 mine casing (rubber center)



(b) Grazing angle reflectance of Tan CARC



(c) Comparison of grazing angle reflectance of antipersonnel mine casings

Figure 7. Laboratory measurements of grazing angle specular reflectance for several antipersonnel mine casings and related materials of interest.

polarimetric attributes of mine materials and coatings relative to these same properties for natural backgrounds. Typically, we will examine the spectral reflectance, and thereby the spectral emittance, of the mine materials for unique spectral features and high grazing angle reflectance compared to the backgrounds. With this in mind the measured reflectance spectra presented in Figure 7 have a number of features of interest. First, examining the spectra in Figure 7(c) for similarities, we find the M-14 and Butterfly (PFM-1) mine casings very similar in having very little spectral structure and a difference in overall reflectance level that is accounted for simply by the difference in surface roughness. Although not shown here, reflectance at 14.3 microns is similar between the M-14 mine and the floppy diskette shell, making the latter a very sensible simulant for M-14 surface properties. Also in Figure 7(c) the plastic shells of the TS-50 and VS-50 (not over plotted to avoid confusion) have identical spectral features at 8.1, 8.9 and 10.7 microns and are apparently the same polymer blend. We believe this to be an ABS (acrylonitrile butadiene styrene) plastic, but we have more work to do in matching spectra to polymer blends before claiming to have identified the materials.

Second, examining the spectra of the rubber materials in Figures 7(a) and (c), we find that the TS-50 rubber center and RTV 3110 have the same spectral features at 7.87, 8.96, and 12.2 microns. We conclude the TS-50 center element is actually an RTV silicone rubber. Comparing these same features to the reflectance spectrum for the VS-50 center element in Figure 7(c), we find them to be wholly different and learn by comparison to another spectral database that the VS-50 center element is natural rubber.

In Figure 7(b), we examine the spectrum of an Army camouflage paint of interest, green and tan CARC, because it is a low reflectance coating purposefully formulated to have the optical properties of the natural background. The grazing angle spectral reflectance features measured for tan CARC are located at 9 and 12.5 microns, with the 9 micron feature being very robust. The infrared spectral features of CARC have been studied extensively in the Joint Multispectral Program (JMSP)²¹ for the purpose of identifying spectral bands that provide a dependable two-color discriminant for CARC in a variety of backgrounds. Using apparent spectral contrast data collected over a wide range of backgrounds and local meteorological conditions, this study concluded that the 9 micron feature provides a persistent two band spectral contrast discriminant for CARC (8.73 / 9.23 microns). This is a successful illustration of our effort to use narrow band spectral features for discrimination of mines within clutter when broadband polarimetry alone cannot eliminate all of the clutter.

In general, our laboratory measurements of grazing angle specular reflectance for the materials listed in Table 2 compare well to the more comprehensive and dedicated measurement efforts of Salisbury and D'Aria²² on the spectral reflectance of terrain background materials, of Johnson and coworkers²³ on the spectral emissivity variations in disturbed soils, and of Cederquist and coworkers^{24,25} on the spectral reflectance of target materials and coatings. For insight into the spectral features of blended polymers we found the book by Garton²⁶ to be very useful. Our purpose in making comparisons to these dedicated studies is to confirm our approach and to put a quantitative foundation under our efforts to apply spectral polarimetry to mine detection.

Given that reliable mine detection has proven to be a difficult and frustrating challenge for many qualified investigators, we stand firmly by our stepwise, controlled risk approach as outlined above, and note here again that our laboratory measurement of mine materials spectral reflectance features is a key element in our approach. These data tell us if and where we can expect to apply spectral polarimetry to antipersonnel mine detection successfully. Overall we are encouraged by the availability and robustness of spectral features for plastic mine casing materials found in the LWIR region of interest.

4. LWIR IMAGING POLARIMETER DEMONSTRATION MEASUREMENTS

A series of demonstration measurements were performed on a sampling of inerted non-metallic antipersonnel (AP) mines in various natural backgrounds. The images were collected using an uncooled microbolometer imager operating in the 8–12 μm band. A manually rotated wire grid polarizer (1200 lines/mm) was mounted to the camera in front of a 25 mm focal length f/1.0 germanium lens. Although complete characterization of the polarization state requires four independent measurements (three if partial linear polarization is assumed), only two measurements were made in this proof-of-concept demonstration. The justifications for this are twofold. First, it can be assumed that the elliptical polarization is small. Second, the geometry of the scene was chosen such that the angle of polarization for the objects of interest was either vertical or horizontal, in which case the maximum image contrast will be seen in the difference between images made with the polarizer in these two orientations. A sequence of 10 frames was co-added for each polarization angle before subtraction.

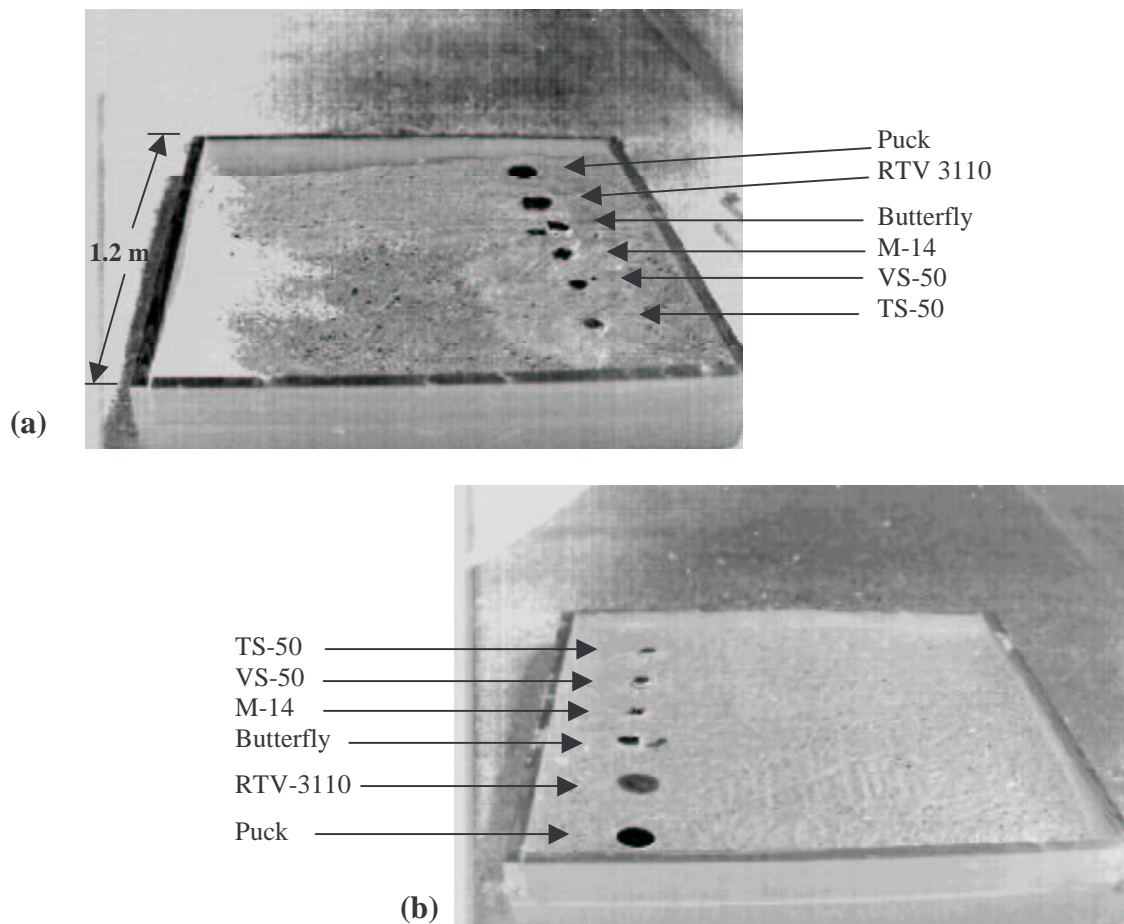


Figure 8. Polarization difference image (0 - 90 degrees) of landmines in sand. (a) Note the raked area at top and disturbed area around lower three mines; (b) Note the sand has been raked and graded to a uniform depth.

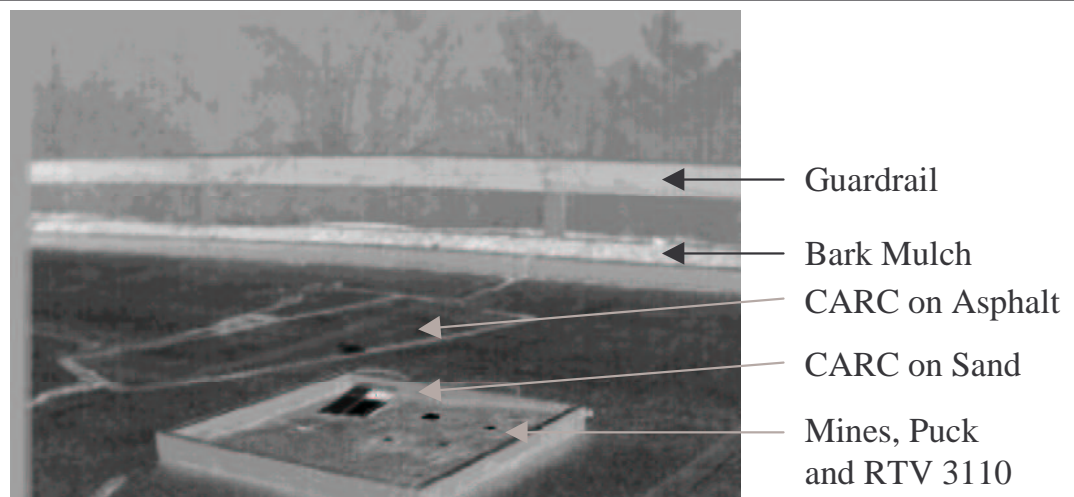


Figure 9. Polarization difference image (0 - 90 degrees) of landmines in sand. Note the polarized emission from the vertical faces of the guardrail and asphalt curbing and the polarized reflection from the horizontal strip of bark mulch beneath the railing.

The Fresnel reflectance relations indicate that the angle of polarization of reflected radiation is perpendicular to the plane defined by the incident and reflected beams while that of emitted radiation is parallel to this plane. Wire grid polarizers transmit radiation when the E-vector is perpendicular to the wire and reflect radiation when the E-vector is parallel to the wire. In this paper we use the convention of 0-degree polarizer angle to mean the wire grid is oriented vertically. Consequently, in the intensity difference images (0 - 90 degrees) discussed below, net reflected (emitted) polarized regions appear brighter (darker) than unpolarized regions for horizontal surfaces, with the converse for vertical surfaces.

4.1. Discussion of the LWIR imaging polarimeter data

Figure 8(a) shows the polarization difference image for four mines, a disk of RTV 3110, and a rubber hockey puck buried flush to the surface in sand. The mines, from front to back, are as follows: TS50, VS50, M14, and PFM-1 (Butterfly), followed by the RTV disk and the puck. It is clear from the image that these materials are easily distinguished from this relatively unpolarized and clutter free background. The polarization appears due to emission. It is noteworthy that the top surfaces of the sandbox frame appear significantly polarized, wood being a material used in the construction of certain land mines. A more subtle, yet equally important feature of this image is the contrast evident between the disturbed and undisturbed regions of the sand. A horizontal (i.e., left to right) strip at the rear of the sandbox, extending to the butterfly mine was purposefully raked with a spring steel rake, and the areas surrounding the mines were disturbed to bury them. These areas are clearly distinguishable in the original images. This effect is attributed to a combination of contrast reduction caused by fine-grained particle coatings on the disturbed soil and the increased moisture content of the surface sand in the disturbed regions.^{22,23} In these measurements the undisturbed sand had received two days of sun since the last rainfall, so the surface was very dry. The disturbed areas brought noticeably moist, but not wet, sand to the surface.

Figure 8(b) shows the sandbox rotated 180° so the objects are in reversed order. In this case, the entire surface was well raked and graded to a uniform depth. The uniformity in the image is evident.

Figure 10 shows the same objects in a more natural background consisting largely of undisturbed aged bark mulch. Other elements in the scene include a rock (about 15 cm diameter), a bush, and some small weeds. The two objects labeled “CARC panel” are metal panels painted with CARC. In this image, the absolute value of the polarization difference is taken so net emission and reflection polarization energy both appear as bright. The bush, a small central portion of the rock, and the rear CARC panel were deliberately underexposed while the front CARC panel and the left edge of the mulch were overexposed. This was done to achieve maximum sensitivity in the region containing the mines and is a limitation of the 8-bit digitizer in the frame grabber rather than the imaging polarimeter itself. Clearly, the mines are evident in Figure 10, yet there is much less contrast against the bark mulch, which itself yields partially polarized radiation, compared to the sand in Figures 8(a) and 8(b). Although the contrast is low, most of the clutter in the thermal image has been suppressed. By contrast, Figure 11, showing only the 0-degree image component of Figure 10, exhibits the wide range of thermal variations present in the scene. Notice that in polarization the two CARC panels appear very similar on the gray scale in Figure 10 while in the intensity image (Figure 11) the sunlit CARC panel appears quite hot in comparison to the shaded panel near the bush.

Figure 9 shows the sandbox using the same camera and lens but now at a greater range (about 4 m). The mines and soil disturbances are still evident, but the scene also includes a large wood guardrail, a strip of aged bark mulch below it, and a vertical asphalt curbing. For two materials in the scene, wood and asphalt, we have both horizontal and vertical surfaces of the same materials in view at the same time. It is useful to observe how the contrast flips for the same material as the orientation and therefore the plane of incidence changes between horizontal and vertical. The vertical surfaces of the guardrail, sandbox wall, and curbing are bright in this difference image (0 - 90 degrees) because the plane of incidence is now horizontal.

There are several other basic polarization phenomena to be observed in this scene (Figure 9). The railing is painted dark brown and consequently heats up considerably during the day. The left end of the railing is seen to yield net polarized emission as the view angle increases off-normal. The mulch exhibits polarized reflection when viewed at this grazing angle. The odd-shaped reflective lines near the sandbox are tar-filled cracks in the asphalt. One of the CARC panels is located on the asphalt just beyond the top left corner of the sandbox. Even though the CARC can be distinguished from the asphalt when we know where to look for it, the two exhibit nearly the same degree of linear polarization as seen here in the gray scale image. This confirms the quantitative DoLP results (Figure 5) obtained earlier in the LWIR spectral polarimeter data for CARC on an asphalt background. The top horizontal edges of the wood sandbox are dark, indicating polarized emission from the wood. Since the

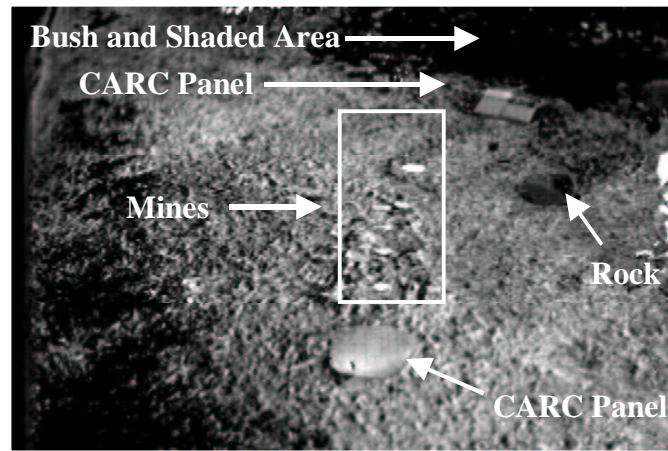


Figure 10. Polarization absolute value difference image, Abs(0-90 degrees), of landmines in bark mulch background with various natural debris and small plants.

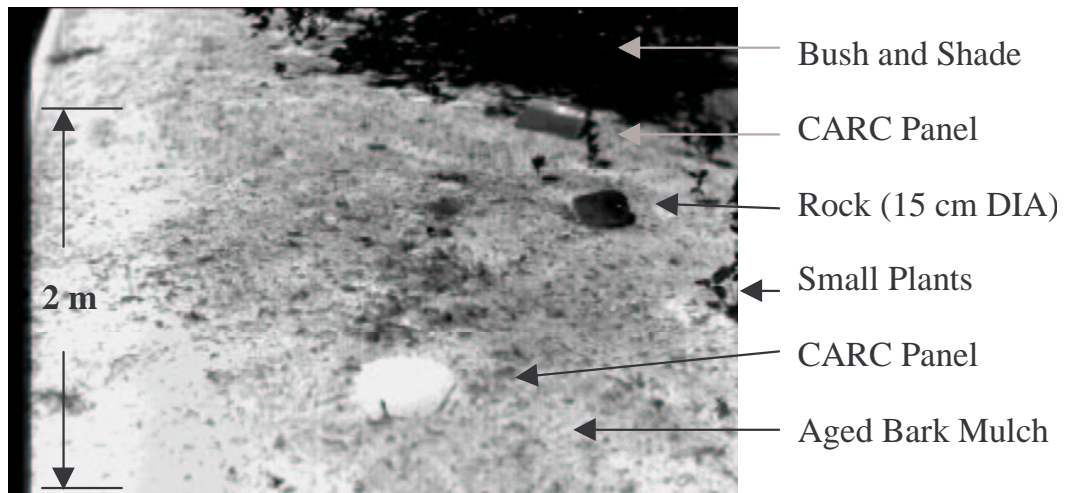


Figure 11. Intensity image (at 0-degree analyzer angle) of landmines in bark mulch background with various natural debris and small plants.

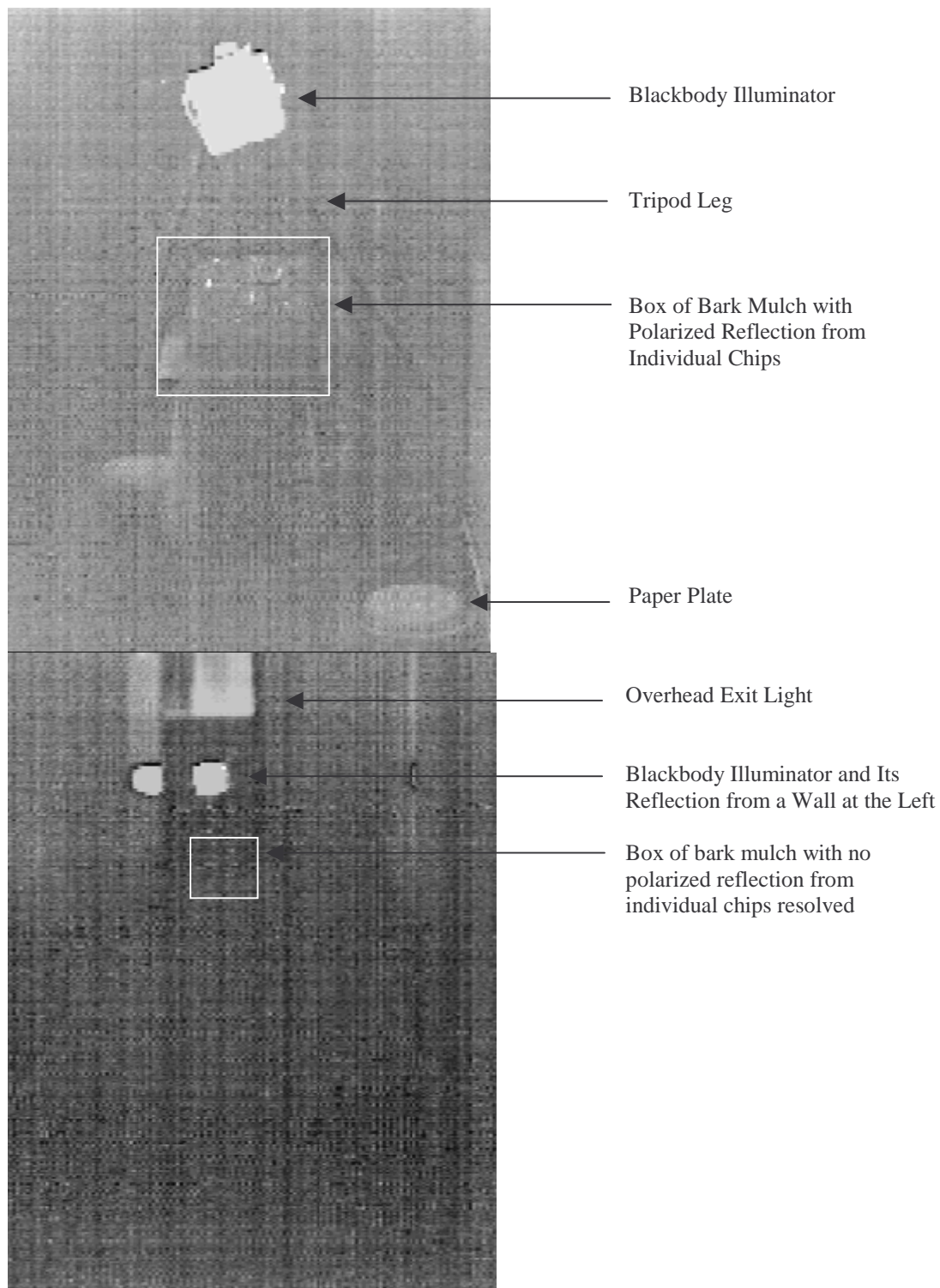


Figure 12. Polarization difference image (0-90 degrees) of bark mulch (a) at a range of 12 ft. Note the polarized reflection from individual bark mulch chips at this range where detector IFOV footprint is about 0.5 inches; (b) at range of 45 ft. Note the bark mulch, directly below the black body illuminator, is not discernible from the background at this range where detector IFOV footprint is about 2 inches.



Figure 13. Bark Mulch Spatial Scale

wood is in contact with the hot sand and asphalt, it can be expected to be warm; however, when wood is in contact with cool moist soil, this may not be the case. Finally, notice the rings of disturbed soil around the mines in the sandbox. Our consistency of detecting disturbed soil encourages us to refine our approach by using narrow band filters (multispectral) that are optimized for the characteristic spectral features of various soil types.^{22, 23}

4.2. LWIR imaging polarimeter - importance of spatial scale

These imaging polarimeter data yield an empirical understanding of the spatial scale requirements for a polarization sensor. The polarimetric signal manifested at the detector will be a superposition of the vectorial (Stokes) polarization contributions from the various elemental surfaces of the target, as "ensquared" within a sensor's instantaneous field-of-view IFOV. If the IFOV is too large for the spatial scale of the target features in the scene, the distinctive polarization attributes of the target can be obliterated. Even though it does not provide a direct example of a wrong choice of spatial resolution, the degree of polarization exhibited by the bark mulch, that is the clutter, in Figure 10 illustrates the importance of spatial scale. Here the bark mulch and the flush-buried mines are similar in dimensions. As a result the bark mulch is an effective polarimetric background clutter for mines in this instance where the only attribute being measured is the degree of polarization with a broadband polarimeter. It is correct to infer from this that an investigator who chooses to apply broadband polarimetry alone to scenes of arbitrary spatial scale will fail.

For a more direct illustration of the importance of spatial scale, Figures 12(a) and (b) are polarization difference images, recorded under controlled conditions, of the box of bark mulch shown in the photograph, Figure 13. For Figures 12(a) and (b) the angles between the illuminator, the normal to the bark mulch box and the polarimeter were held constant while the range between the bark mulch and the polarimeter increased from 12 ft in Figure 12(a) to 45 ft in Figure 12(b). The IFOV footprint at the bark mulch is about 1/2 inch in Figure 12(a) and 2 inches in Figure 12(b), compared to a mean dimension of 3/4 inches for the bark mulch chips. As expected the polarized reflectance from individually resolved bark mulch chips are visible in Figure 12(a) but obliterated by the coarser spatial resolution in Figure 12(b).

Clearly, in the design of an hyperspectral polarimeter, the spatial resolution is an enabling design parameter²⁷, making the spectral and polarization attributes of the target accessible to be used as additional discriminants. At the same time, no amount of effort in tailoring the spatial resolution will help in the event the spatial dimension and polarimetric attributes of the clutter are nearly the same as that of the target. In this situation the broad spectral band imaging polarimeter will not be successful, and the need to invoke hyperspectral polarimetry, thereby adding the spectral attributes for discrimination purposes, is clear.

4.3. LWIR imaging polarimeter - performance

The microbolometer FPA demonstrated strong performance when applied in the broad spectral band imaging polarimeter mode with the 1200 line/mm polarizer mounted in the optical train. In addition to the imaging polarimeter that we have assembled around an uncooled FPA being affordable, we can report the following performance on static scenes:

1. With f/1 optics and modest additional effort to hold the temperature of the optics constant when operating in the field, the polarimeter delivered a NEDT $< 0.1^{\circ}\text{C}$.
2. The polarimeter consistently detected disturbances in sand, soil and bark mulch created by burying objects and bringing fresh sample to the surface.
3. The system detected flush-buried antipersonnel mine casings and simulants both night and day, with and without cloud cover.

5. LWIR UNCOOLED FPA MULTISPECTRAL IMAGING POLARIMETER

Even though we are quite pleased with the performance of the uncooled FPA microbolometer in the mode of a broadband imaging polarimeter, it is necessary to next demonstrate its performance as a multispectral imaging polarimeter. There are two goals here. The first is to combine spectral, polarimetric, and imaging feature spaces together in order to demonstrate the increased discrimination gain, relative to earlier abbreviated combinations (i.e. spectral/polarimetric, spatial/polarimetric). The second is to experimentally ascertain the radiometric sensitivity performance when using an uncooled FPA with spectral filters. We recognize that the NEDT of the current uncooled FPA technology is an order of magnitude higher than for cooled FPAs, and thus we anticipate the division of the available polarization signature into say 10 (multispectral) or 100 (hyperspectral) bins implies the uncooled FPA performance as a multi- or hyper-spectral polarimeter may be marginal. We

nevertheless are of course motivated to see how far the uncooled FPA technology can be pushed, given its nearly two orders of magnitude affordability over cooled FPA technology. Our experience to date with the uncooled FPA indicates that its actual NEDT performance can be as much as a factor of two better than the nominal specification.

The LWIR (broadband) imaging polarimeter, implemented using a microbolometer uncooled FPA as described in Section 4, was recently modified to add a 6-slot filter wheel. Two of these bandpass filters were chosen to spectrally discriminate CARC (when using 2-band processing). The desired bandpasses were 8.1–9.2 and 9.2–10.2 μm , but compromising in order to employ inexpensive off-the-shelf filters yielded 8.1–9.3 and 9.3–10.2 μm . Another filter was chosen as an LWIR bandpass of 6.5–14 μm , since the unfiltered microbolometer detector response due to the much broader Germanium optics passband is roughly 3–14 μm , which includes (at times undesired) solar scatter. Two additional filters divide the LWIR into 7.5–10.2 and 10.0–13.6 μm passbands. Finally, the 6th slot was open, allowing the aforementioned full 3–14 μm response. The following sequence of measured imagery constitute early-look results that indicate the discriminative power of polarimetric multispectral imaging.

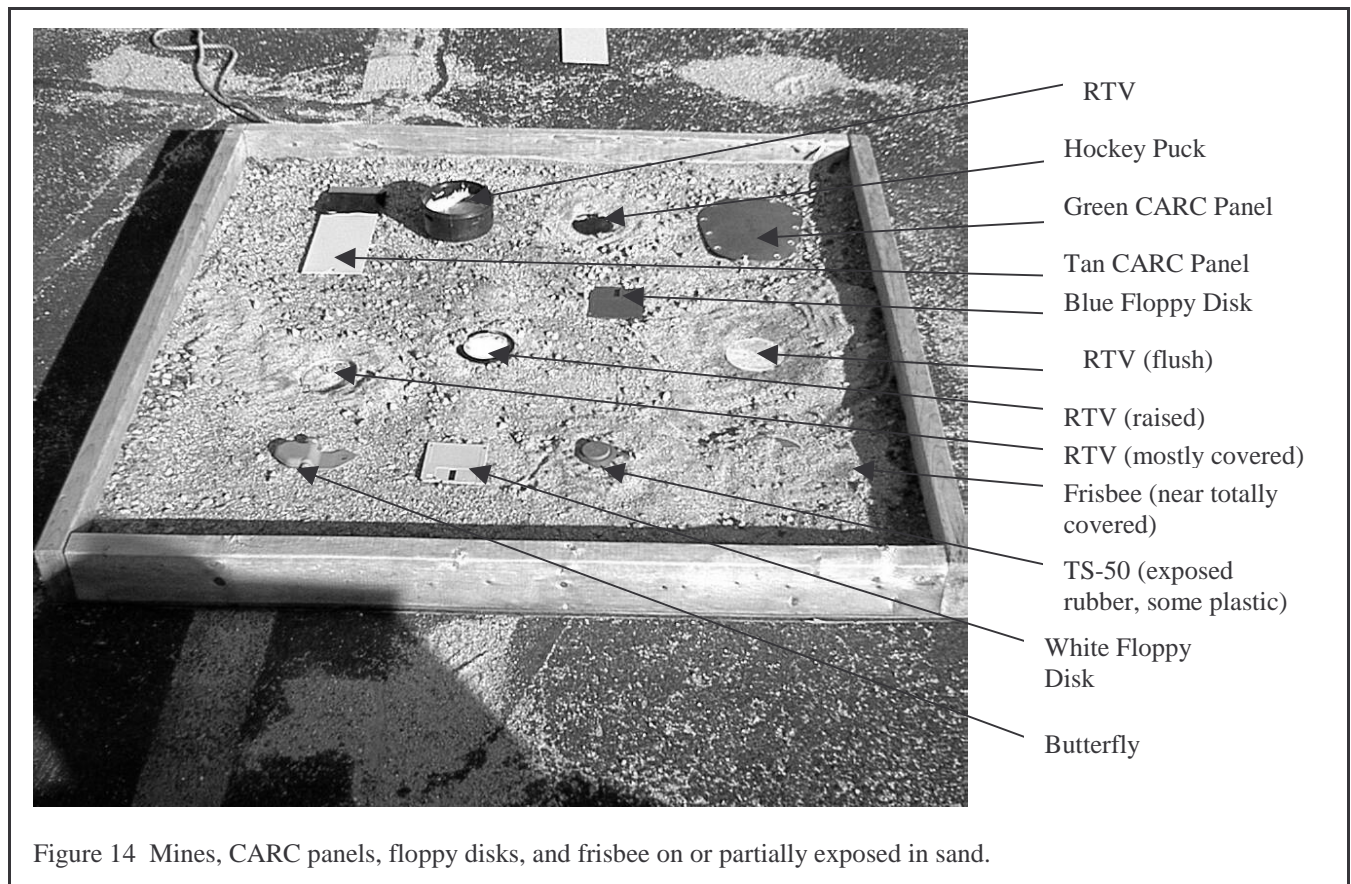


Figure 14 shows the measurement setup of the various target materials on or partially exposed in sand. The RTV silicone rubber is spectrally identical to the rubber center of the TS-50 AP mine (per discussion in Section 3). In order of increasing aggregation of feature spaces (and discrimination processing sophistication) are Figures 15–18. Figure 15 is a broadband LWIR (6.5–14 μm) conventional intensity image. Figure 16 is the polarization difference image (0–90) through the same broadband LWIR filter (6.5–14 μm). Figure 17 is a multispectral PCA (principal components analysis) image showing intensity based on projection of the data comprised of the 5 previously stated LWIR passband filters onto the 2nd most energetic eigenvector. Figure 18 is a polarimetric multispectral PCA image, this time projecting 12-vector imagery consisting of all 6 bands, with 0-deg and 90-deg polarizer orientations for each, onto its 2nd most energetic eigenvector. Initial eigenanalysis on subset polarimetric multispectral data supports the visual conclusions: for the 2nd eigenvector is predominantly synthesizing the polarization difference yet includes some multispectral projection, while the 3rd eigenvector is predominantly synthesizing a spectral filter. The 2nd eigenvalue is about 3x the energy of the 3rd eigenvalue.

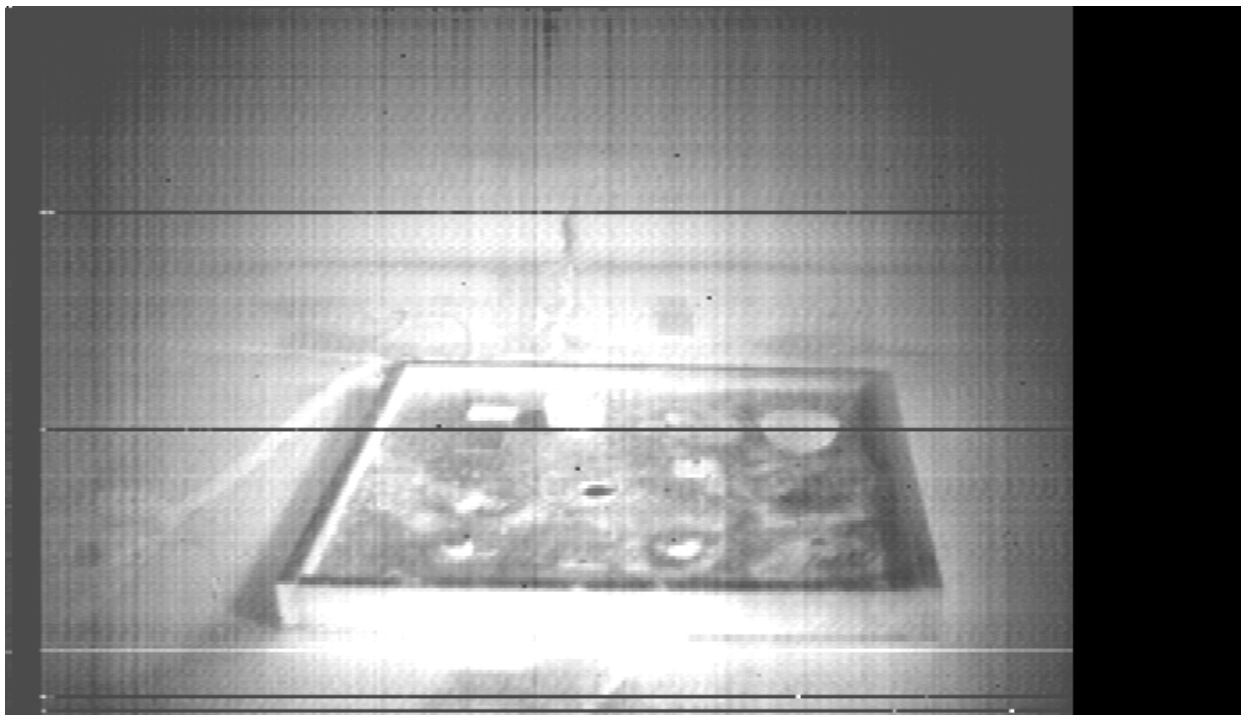


Figure 15 Broadband LWIR (6.5–14 micrometers) intensity image

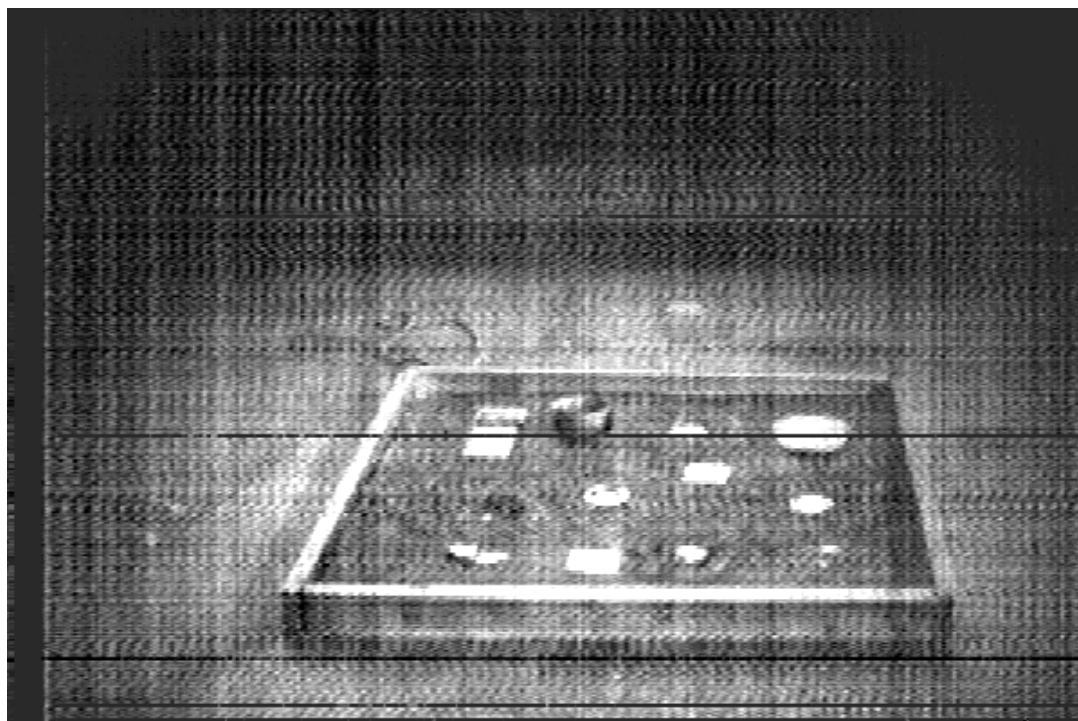


Figure 16 Polarization difference image for broadband LWIR (6.5–14 micrometers).

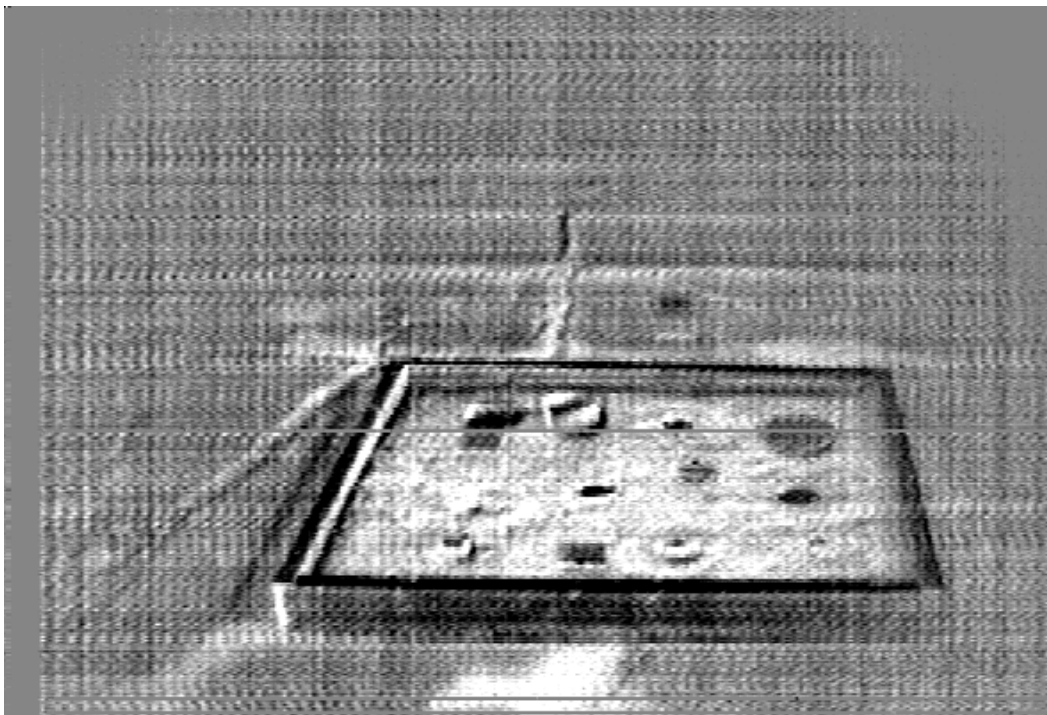


Figure 17 Multispectral PCA image comprising the data of 5 LWIR spectral bands projected onto its 2nd eigenvector.

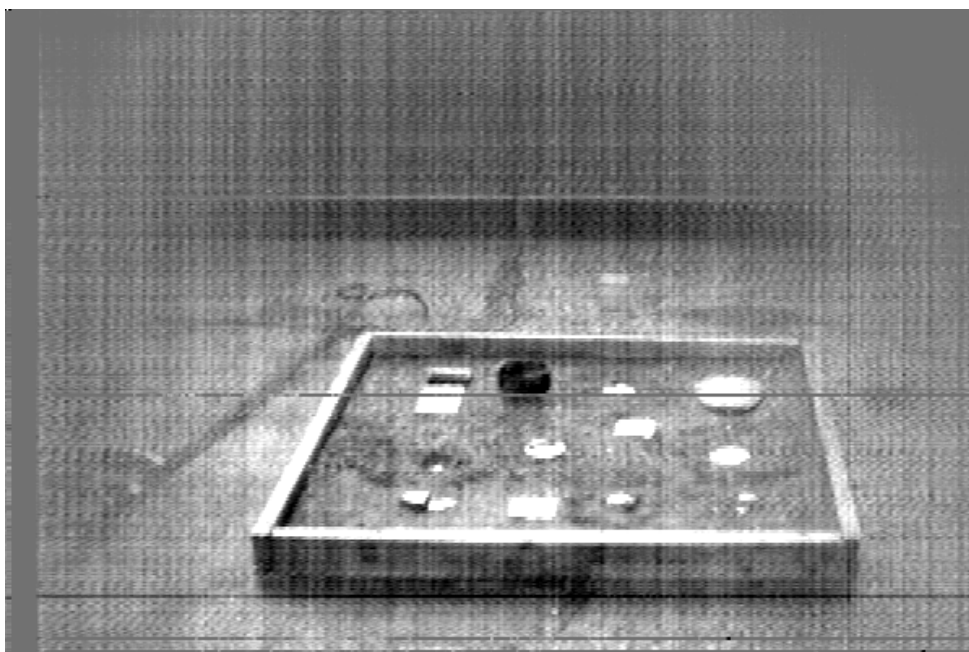


Figure 18 Polarimetric multispectral PCA image comprising the data of 12 image planes (6 bands x 2 polarizer orientations) projected onto its 2nd eigenvector.

6. CONCLUSIONS

6.1. Rationale for combining hyperspectral and polarimetric IR imaging

We have presented several pieces of evidence to motivate the marriage of hyperspectral and polarimetric IR imaging. Basically we recognize that polarimetric and hyperspectral attributes supplement one another, and that all available independent supplemental pieces of information can be used to advantage for discriminating target objects from background clutter. Polarimetry appears particularly suited to segregating partially exposed man made objects from natural background based on the typical smoothness of the former, and typical roughness of the latter. While we expect to exploit the broadband attributes of man made objects, knowledge of the characteristic spectral features for a large number of both man made and natural background materials (for example, quartz sand and the mine casing spectra presented above) enables an additional hyperspectral polarization discrimination capability that comes in the form of the narrow spectral features.

6.2. Polarimeter design process and some useful insights

Our current understanding of the best design process for an imaging multi- or hyper-spectral polarimeter application^{10,13,17,18,28,29,30} includes the following points:

1. The spatial resolution will be determined and even constrained by the spatial scale of the target and primary clutter sources in the scenes and scenarios of interest. A polarimeter geared toward mine detection requires a spatial scale consistent with the mine dimensions and respective of the spatial scale of primary clutter sources. Recall the bark mulch images in Figures 12(a) and (b).
2. Given this spatial scale constraint the spectral range, resolution, and choice of detector are a compromise between the need for signal (broadband operation) and the availability of distinguishing "narrow" spectral features to exploit (multi- or hyper-spectral operation).
3. Prerequisite to this process is a very clear realization¹⁰ that at least two, and preferably three, independent polarization measurements must be acquired at near perfect registration, which for moving platforms means simultaneously. Otherwise the misregistration of polarimetric detection channels will masquerade as serious to overwhelming polarimetric clutter, induced particularly by time-dependent effects such as sensor platform motion, substantial changes in the downwelling radiation incident upon the target, and even small temperature and sensitivity drifts in the detector.
4. Multispectral or hyperspectral polarimetry offers essential advantages over broadband polarimetry in applications such as demining where spectral techniques allow discrimination against similarly polarized backgrounds such as asphalt, and potentially when tuned to discriminate surface spectro-polarimetric texture disturbances indicative of sub-surface mines.

6.3. Overall status of IR polarimetric hyperspectral imaging effort

More important than any one of the above proof-of-concept demonstrations, each of which represents a successful step in the risk-reduction sense, is to put all these elements together into the same sensor, an IR Polarimetric Hyperspectral Imager; this is our goal and represents our most significant effort to go beyond state-of-the-art. If we are successful, we will have captured the spectral, spatial and polarimetric attributes in one sensor, all simultaneous in time and perfectly registered spatially.

Going beyond the current proof-of-concept sensor configuration intended for static scenes, we now have an innovative sensor design in hand and are building the prototype IR Polarimetric Hyperspectral Imager. We have verified in a laboratory visible-band implementation that our design allows us to extract any angle or degree of polarization in the scene at a frame rate determined by the multiplexer and signal processor. For initial application purposes we are using a cooled FPA in the IR prototype, yet we anticipate the opportunity to test uncooled FPAs for the demining/UXO applications. Having all of the spectral, spatial and polarimetric attributes available to us at the same time provides a powerful discrimination and clutter suppression capability that is quite unique. All indications from analyzing our sequentially collected data are that these attributes taken together will comprise a reliable sensor for the detection of surface and flush-buried land mines, as well as for disturbed soil and shallow-buried mines over diurnal cycles or with microwave enhance heating.

Laboratory measurements or other knowledge of the spectral reflectance and polarimetric properties of the targets in each application are necessary to support hyperspectral polarimetry. Our laboratory data help directly to (a) determine the locations and widths of spectral features needed to infer the value added of hyperspectral over multispectral polarimetry and thereby to (b) determine in potential applications whether broadband, multispectral or hyperspectral polarimetry applies.

In summary, using somewhat sparse, but very reproducible, data from field and laboratory measurements to date, we have presented a case for LWIR polarimetric hyperspectral imaging in applications requiring the detection of plastic, as well as metal, flush-buried and scattered surface mines. This approach applies equally well to partially exposed unexploded ordnance (UXO) and to shallow-buried mines subjected to microwave enhanced heating and diurnal cycles. Our rationale for combining hyperspectral and polarimetric imaging was presented.

7. ACKNOWLEDGEMENTS

This work was supported by the Northeastern University Humanitarian Demining Multi-Disciplinary Research Initiative (MURI) under contract to the US Army Research Office / Dr. Jim Harvey, Program Director.

REFERENCES

1. M.A. LeCompte, J.A. Conant, and D. Dvornik, "Study of the phenomenology of polarization in aircraft infrared signatures," *Proceedings of the IRIS Specialty Group on Targets Backgrounds and Discrimination*, Monterey, 1988.
2. M.A. LeCompte and J.A. Conant, "POLAR version 3.0 technical manual," Aerodyne Research, Inc., Report No. ARI-RR-1009, 1993.
3. J.A. Conant and M.A. LeCompte, "POLAR version 3.0 users manual," Aerodyne Research, Inc., Report No. ARI-RR-1136, 1995.
4. J.R. Maxwell and S.F. Weiner, "Polarized emittance Vol 1: Polarized bidirectional reflectance with Lambertian or non-Lambertian diffuse components", US Army Ballistic Research Lab Report BRL-CR-154 (1974).
5. R. Tooley, "Man-made target detection using infrared polarization", SPIE, 1166, pp. 52-58, 1989.
6. M.A. LeCompte and J.A. Conant, "Target identification using passive infrared signature polarization," *Proceedings of the Meeting of the IRIS Specialty Group on Targets, Backgrounds and Discrimination*, Naval Postgraduate School, Monterey, 1990.
7. M.A. LeCompte and J.A. Conant, "Modeling and analysis of aircraft infrared signature polarization," *Proceedings of the IRIS Specialty Group on Targets Backgrounds and Discrimination*, Monterey, 1990.
8. M.A. LeCompte, "Aircraft infrared signature polarization model and measurement comparison," *Proceedings of the IRIS Specialty Group on Targets Backgrounds and Discrimination*, Naval Training Center, Orlando, 1992.
9. M.A. LeCompte, and F. Iannarilli, "Infrared signature polarization for passive non-cooperative target recognition," *Proceedings of the IRIS Specialty Group on Targets, Backgrounds and Discrimination*, Naval Training Center, Orlando, 1992.
10. F.J. Iannarilli and M.A. LeCompte, "Predicted performance of counter-air target ID using IR polarimetry" *Proceedings of the IRIS Specialty Group on Targets, Backgrounds and Discrimination*, 1994.
11. M.R. Hess, G.A. Freund, D.L. McMaster, D.B. Nichols, M.A. LeCompte, F.J. Iannarilli and J.E. Rice, "Measurement and analysis of aircraft infrared signature polarization", *IRIA-IRIS Proceedings, 1994 Meeting of the IRIS Specialty Group on Targets, Backgrounds and Discrimination*, IV, 207-226, 1994.
12. M.R. Hess, M.A. LeCompte, D.B. Nicholls, R.R. Keever, W.B. Shepard, and H.R. Woodman, "Aircraft IR signature polarization measurements at Long Jump '94," *Proceedings of the IRIS Specialty Group on Targets Backgrounds and Discrimination*, Orlando, 1995.
13. M.A. LeCompte, F.J. Iannarilli, D.B. Nichols, and R.R. Keever, "Multispectral IR signature polarimetry for detection of mines and unexploded ordnance (UXO)," *Proc. SPIE* 2496 (1995).
14. F.J. Iannarilli, S.H. Jones, H.E. Scott, and P. Kebabian, "Polarimetric Spectral Intensity Modulation (P-SIM): Enabling simultaneous hyperspectral and polarimetric imaging," *Proc. SPIE* 3698 (1999).
15. H.E. Scott, S.H. Jones, F. Iannarilli, and K. Annen, "Hyperspectral IR polarimetry with applications in demining and unexploded ordnance detection," *Proc. SPIE* 3534 (1998).
16. J.R. Simard, "Experimental Evaluation of the Apparent Temperature Contrast Created by Buried Mines as Seen by An IR Imager," Defense Research Establishment Suffield, Canada, Report Number SR 607 (November 1994).
17. Li-Jen Cheng and G. Reyes, "AOTF Polarimetric Hyperspectral Imaging for Mine Detection," Jet Propulsion Laboratory, *Proceedings SPIE Vol. 2496*, 305 (1997).

-
18. S. Jones, F.J. Iannarilli, H.E. Scott, and K. Annen, "Data Management for Imaging Spectrometers," Air Force Research Laboratory, Space Vehicles Directorate Report AFRL-VS-HA-TR-98-0005. Hanscom AFB (January 1998).
 19. P. Li and A. Maad, "Infrared Imaging of Buried Objects by Thermal Step-Function Excitations," *Applied Optics* **34**, 5809 (1995).
 20. M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, Oxford, 1970.
 21. A.D. Stocker, et al, "Analysis of infrared hyperspectral measurements by the Joint Multispectral Program," *SPIE Conference on Targets and Backgrounds: Characterization and Representation*, Orlando, SPIE 2469, pp.587-602, 1995.
 22. J.W. Salisbury and D.M. D'Aria, "Emissivity of terrestrial materials in the 8-14 micron atmospheric window," *Remote Sens. Environ.* **42**, pp. 83-106, 1992.
 23. J.R. Johnson, P.G. Lucey, K.A. Horton, and E.M. Winter, "Infrared measurements of pristine and disturbed soils. 1. Spectral contrast differences between field and laboratory data," *Remote Sens. Environ.* **64**, pp 34-46, 1998.
 24. J.N. Cederquist, et al, "Multispectral infrared target detection: phenomenology and modeling," *SPIE Proceedings*, Vol. 1954, 1993.
 25. T.J. Rogne, F.G. Smith, J.E. Rice, "Passive target detection using polarized components of infrared signatures," Orlando, SPIE, 1317, pp. 242-251, 1990.
 26. A. Garton, *Infrared Spectroscopy of Polymer Blends, Composites and Surfaces*, Hanser Publishers, Munich, 1992.
 27. F.J. Iannarilli and J.A. Conant, "Quantifying key trade-off between IR polarimetric discriminability versus pixel resolution against complex targets," *Proc. SPIE* 3699 (1999).
 28. B.A. Barbour, M.W. Jones, H.B. Barnes and C.P. Lewis, "Passive IR polarization sensors: a new technology for mine detection", *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets III*, Orlando, SPIE, 3392, pp. 96-103, 1998.
 29. M. Larive, L. Collot, S. Breugnot, H. Botma, P. Ross, "Laid and flush-buried mines detection using 8-12 micron polarimetric imager", *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets III*, Orlando, SPIE, 3392, pp. 115-120, 1998.
 30. C.A. DiMarzio, T. Vo-Dinh, H.E. Scott, "Some approaches to infrared spectroscopy for detection of buried objects", *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets III*, Orlando, SPIE, 3392, pp. 158-166, 1998.

Polarimetric-Spectral Intensity Modulation (P-SIM): Enabling simultaneous hyperspectral and polarimetric imaging

Frank J. Iannarilli, Jr., Stephen H. Jones, Herman E. Scott, and Paul Kebabian

Aerodyne Research, Inc., 45 Manning Rd., Billerica, MA USA 01821

ABSTRACT

We introduce the method of Polarimetric-Spectral Intensity Modulation (P-SIM) and discuss how it enables a new robust class of hyperspectral polarimetric imaging sensors. P-SIM was invented by one of us (P. Kebabian) and has been submitted for patent. We are presently building a sensor, dubbed the IR Polarimetric HyperSpectral Imager (IRPHSI) which implements the P-SIM concept. P-SIM employs a novel and robust optical multiplexing scheme that enables simultaneous measurement of spectral and full elliptical polarimetric image content, employing a single focal plane detector and conventional optics, and eliminating moving parts and difficult alignment issues. The technique is equally viable across the visible through long-wave infrared bands. The P-SIM concept constitutes a breakthrough for the inclusion of polarimetry in optical hyperspectral imaging. To date, even single-band polarimeter designs for remote sensing are compromised due to their lack of spatiotemporal measurement registration, inapplicability to marginally resolved scene elements, costly optical configurations, or polarimetric ambiguity from too few “channels”. Polarimetric imaging from a moving platform (i.e., moving field of regard) or against moving targets rules out the standard methods of time-sequential polarimetry via rotation of a polarizer or waveplate. P-SIM eliminates these limitations while additionally extending polarimetry into the spectral imaging domain.

Keywords: Optical infrared polarimetry hyperspectral discrimination imaging sensing

1. MOTIVATION FOR POLARIMETRIC AND SPECTRO-POLARIMETRIC SENSING

Sensor technology has progressed to the point where it is practical to detect and measure the spatial, spectral, and polarimetric distribution of optical radiation of sample objects or remote scenes. Such in-situ and particularly remotely sensed radiation characteristics reveal a substantial amount of information about the objects or phenomena under study. For instance, numerous air- and spaceborne optical sensors produce spectrally resolved imagery used for military surveillance as well as geophysical and earth resource mapping. The polarimetric state of received radiation also contains valuable information about source object surface roughness and orientation. Optical polarimetric sensing has been advocated for a number of years for various military and remote sensing applications. Because natural backgrounds are typically unpolarized[†], relative to other modalities, polarimetric sensing can offer superior clutter suppression in the discrimination of typically polarized manmade objects. For image recognition applications, the predictable Fresnel behavior of polarimetric observables may enable passive inference of surface orientation and object shape¹. Imaging polarimetric and spectro-polarimetric sensors provide a significant increment in our capability to respond to existing requirements, as well as to our insatiable need for more information in remote sensing applications.

2. FUNDAMENTAL POLARIMETRIC SENSOR REQUIREMENTS FOR DYNAMIC PLATFORMS

2.1. The requirement for spatio-temporally registered full-channel polarimetry

The prior art contains numerous responses to this demand for polarimetric and spectropolarimetric sensors. Polarimeter designs can be classed into 2 general categories: (1) time-simultaneous or (2) time-sequential techniques. Time-simultaneity invariably involves a division of power among sensing “channels”. One approach for time-sequential polarimetry involves the rotation of retarders as in the case of the infrared spectropolarimeter by Goldstein et al, US patent 5,045,701.

[†] Depending on pixel footprint, various background features may exhibit polarization.

In recent years, prototype polarimeters have been introduced into experimental field measurements to validate observability predictions. These time-sequential polarimetric designs, when employed in static sensing configurations, have been adequate to validate predicted polarization effects. Unfortunately for many applications of interest involving a dynamically moving sensor platform and/or object of interest, time-sequential polarimetric techniques are precluded due to the practically uncorrectable spatiotemporal mis-registration of the imaged scene (due to relative scene motion) over the duration of the measurement sequence. Such mis-registration creates artifacts that masquerade as polarization signals. The limits of allowable relative motion are quite restrictive since many polarization signals of interest represent polarization levels on the order of only several percent. Thus, spatiotemporal misregistration must typically be controlled to below 1%.

Specifically, for a temporal sampling sequence duration of dt , and a spatial resolution element of linear angular subtense Δq , in order to employ time-sequential polarimetry the relative sensor pointing slew rate $\dot{\theta}$ must satisfy the following constraint: $\dot{\theta} \cdot dt \leq 0.1 \cdot \Delta q$. This is certainly not the case for most military and remote sensing field applications. For example, at a modest $1^\circ/\text{sec}$ platform rotation rate (e.g., due to buffeting), the footprint of an IFOV of $100\mu\text{rad}$ would shift nearly an entire IFOV subtense during a frame period (τ_{int}) of a mere 5msec! Note that for such applications the employment of time-simultaneous polarimetry might not necessitate sampling extremely rapidly (i.e., to attempt near “freeze-frame” sampling), but rather that exactly the same scene content be dwelled upon in the same spatiotemporal manner across all the sensor’s polarimetric channels for the given spatial resolution element.

To date, time-simultaneous polarimeter designs for remote sensing are compromised due to their inapplicability to marginally resolved scene elements, costly and difficult optical configurations, or polarimetric ambiguity from too few “channels”. Time-simultaneous imaging polarimetric channel acquisition is possible at the expense of important drawbacks by employing “microscale channelization”. This technique inserts a pixelated polarizing filter mask of alternating analyzer orientations at an intermediate image plane within the optical train, at a sacrifice of 2-4 in spatial resolution relative to the underlying focal plane array detector spacing/resolution. A more serious drawback is that microscale channelization designs are suitable only when restricted to applications imaging radiometrically and polarimetrically uniform extended scene elements (i.e., elements with radiometric and polarimetric (auto)correlation lengths roughly at least several pixels). Otherwise, due to the lack of spatial registration (coincidence) of channels, a microscale design attempting polarimetric imaging of marginally resolved scene elements will yield unacceptable phantom polarization artifacts due to the unknown and varying amounts of light incident upon each of the channel pixels as the scene dwell is varied.

Time-simultaneity can also be obtained by employing “macroscale channelization” by means of a “quadfocal” optic (e.g., via a segmented mirror), wherein offset images are projected through oriented analyzers and onto a singular focal plane. This preserves spatial resolution and achieves spatial channel coincidence at the expense of a reduced field of regard, the need for custom optics and precise alignment. Alternately, multiple focal planes can be employed at greatly increased expense and heightened alignment difficulties.

Even designs that embrace the time-simultaneity and spatial channel coincidence requirements fall short when too few “channels” are measured. For instance, a design which employs merely “vertical” and “horizontal” linear polarizers (2 channels) in an attempt to deduce the “degree of polarization” will fail for most encountered angles of polarization, particularly yielding a null signal at a 45 degree polarization orientation.

2.2. Constraints on extension of spatio-temporally registered polarimetry to spectral sensing

The extension of spatio-temporally registered polarimetry designs to spectropolarimetry must of course by definition not employ time-sequential spectrum acquisition methods. An example of the latter is the device of US patent 5,131,742 to Schaff (1992) wherein the ordinary and extraordinary diffracted beams propagated through an acousto-optic tunable filter (AOTF) are intercepted for purposes of polarimetry, but wherein the passband of the AOTF must be swept in time to acquire the spectrum.

Compliance with time-simultaneity can typically be accomplished by insertion of a dispersive diffraction grating or prism into the sample beam. For imaging spectropolarimetry, this typically implies the use of an area focal plane detector wherein a line image is spectrally dispersed along the remaining dimension of the detector array. Nevertheless, this approach may be fraught with practical difficulties, as illustrated in the device described in US patent 5,337,146 to Azzam (1994). To accomplish spectropolarimetry it is suggested therein to insert polarizing means between the diffraction grating and the area detectors for two or more of the intercepted diffraction orders. In practice, this configuration would require precise alignment

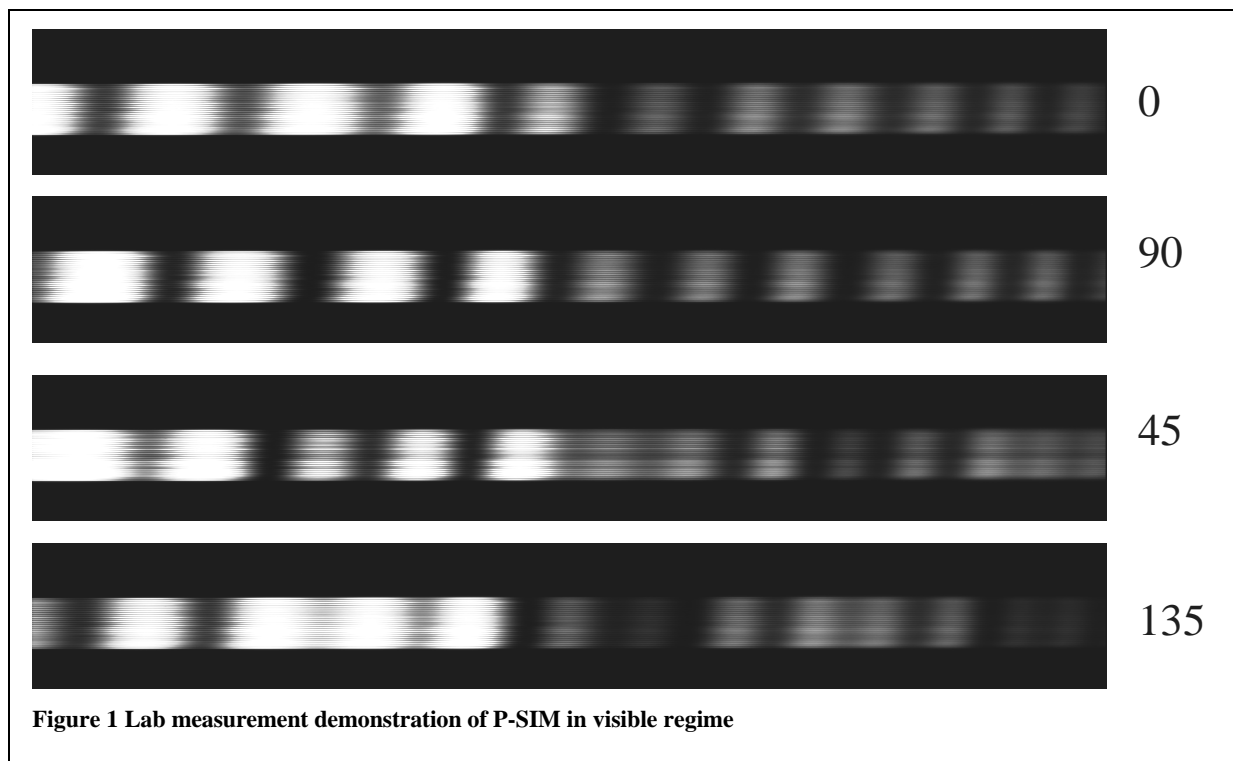


Figure 1 Lab measurement demonstration of P-SIM in visible regime

and registration of the existent beams as imaged onto one or more focal plane array detectors, potentially precluding compact, rugged, or inexpensive embodiments.

3. POLARIMETRIC SPECTRAL INTENSITY MODULATION (P-SIM)

The most telegraphic means of conveying the P-SIM (Polarimetric Spectral Intensity Modulation) concept is to show its employment within a conventional hyperspectral line (pushbroom) imager. We are presently building a prototype of just such a design for the LWIR dubbed “IRPHSI”, which we will describe herein. It is also possible to insert the elemental P-SIM modulator assembly into other spectral imaging designs to overlay polarimetric capability, and we will briefly suggest an example pertaining to a 2D chromo-tomographic imaging spectrometer (CTIS).

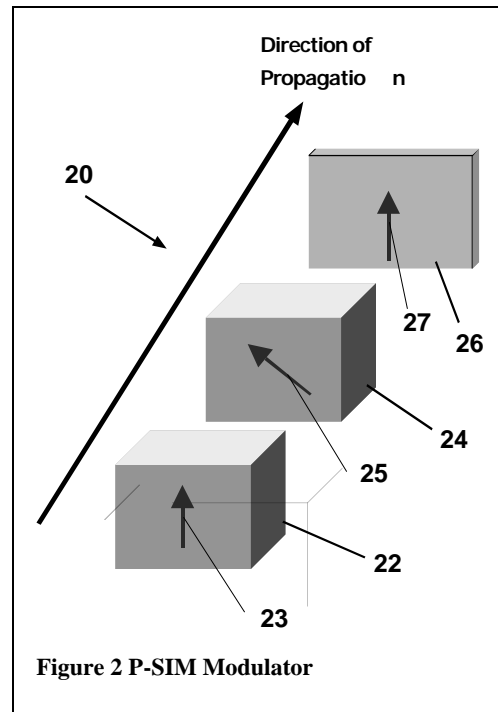
3.1. Overview of IRPHSI Project Underway

The InfraRed Polarimetric HyperSpectral Imager (IRPHSI) concept was originated by Aerodyne Research as part of a Phase 1 SBIR effort under the sponsorship of the U.S. Air Force (AFRL/Hanscom). The sensor simultaneously acquires polarimetric and hyperspectral imagery on a single focal plane array (FPA). The effort to fully design and build IRPHSI, under Air Force Phase 2 SBIR funding, is presently underway. Aerodyne is presently submitting a patent application for the underlying “Polarimetric Spectral Intensity Modulation” (P-SIM) design concept employed within IRPHSI. Our polarimetric hyperspectral imaging technique is equally viable in the visible or near-infrared bands, but our prototype sensor currently under development concentrates on the “thermal” infrared wavelengths spanning 8 to 12 micrometers.

3.2. IRPHSI Design Configuration

The difficulties of introducing snapshot polarimetry into hyperspectral imaging are obviated by employing IRPHSI’s P-SIM design concept. IRPHSI parsimoniously achieves time-space simultaneity without sacrificing spatial resolution, and employs a compact inexpensive configuration. The notional optical layout is shown in Figure 2 and Figure 3. In Figure 2 the key polarimetric spectral intensity modulator assembly 20 is presented, showing its components and their arrangements. A birefringent primary medium 22, in this instance crystalline CdS (Cadmium Sulfide), is the first component to receive the propagating polychromatic waveflux. The waveflux subsequently propagates into a birefringent secondary medium 24. The birefringent primary medium by nature possesses orthogonally resolved fast and slow axes of propagation. The nature of propagation through the medium is that a respectively shorter or longer propagation time delay is imposed upon the

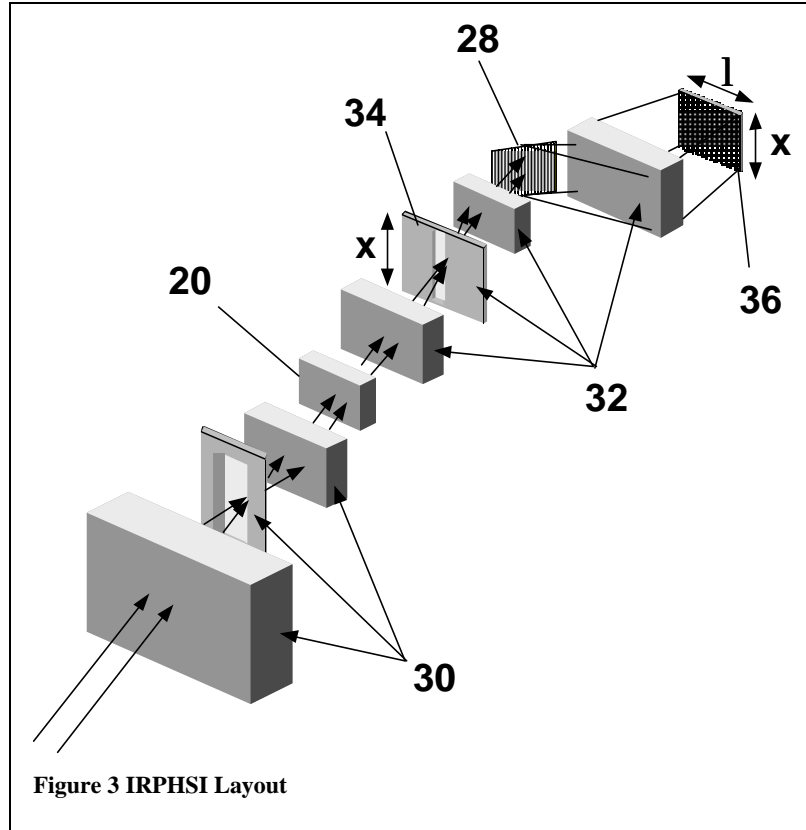
polarization component of the incident wave as resolved with respect to either axis. The primary medium is crafted such that its fast and slow axes lie within a plane substantially normal to the direction of propagation. The characteristics of the primary medium are chosen such that its propagation time delays are minimally dispersive or well tempered functions with respect to optical frequency within the desired spectral range.



The secondary medium is similar in aforementioned characteristics to the primary medium, albeit with two critical differences. First, the predetermined difference between its slow and fast propagation delays is normally to be discernibly different from that of the primary medium. This is typically arranged, if the two media are composed of the same material, by using differing lengths (as measured along the direction of propagation). Otherwise, a combination of length and material disparity is employed. The length of the secondary medium should often be considerably greater than that of the primary medium. Second, the secondary medium must be situated such that its own fast axis 25 is substantially displaced rotationally to form a finite acute (less than 90 degree) angle with respect to the fast axis 23 of the primary medium. Normally, this acute angle is set at 45 degrees.

Waveflux propagated through the secondary medium subsequently propagates through a linear polarizer 26. For optical infrared sensors, wire grid polarizers satisfy this and other predetermined typical criteria. They are fabricated by depositing metal within parallel grooves etched upon transparent planar substrates such as ZnSe. It is critical that the transmission axis 27 of the linear polarizer is oriented substantially rotated to form a finite acute (less than 90 degree) with respect to the fast axis 25 of the secondary medium. Normally, this acute angle is set at 45 degrees.

The IRPHSI sensor is composed by inserting the modulator assembly 20 described above into a slit-based line imaging spectrometer as shown in the layout of Figure 3. The inclusion of the modulator assembly enables this imaging spectrometer to simultaneously perform full elliptical Stokes polarimetry in addition to its normal function. This layout is comprised of input imaging optics 30, typically implemented with a light-collecting primary imaging optic, a field-stop, and collimating lens as illustrated. The imaged waveflux is relayed as a collimated beam through the modulator assembly 20, the output of which is subsequently analyzed by the arrangement of remaining conventional elements 32, 34, 28, and 36. The re-imaging optics 32 produce 2 principal focal planes. A field definition slit 34 is placed at the first (intermediate) focal plane. A dispersive element 28, in this instance a diffraction grating, is placed in the collimated beam region between the first and final focal planes of the re-imaging optics. The spectrally dispersed, 1-dimensionally spatially resolved line image is projected onto a photosensitive areal detector array 36.



With each readout of the photosensitive areal detector array 36, a spectrally and (1-D) spatially resolved image of the spectrometer slit 34 is acquired. To achieve spatial imaging in the direction perpendicular to the slit, sensor platform motion translates the slit across object space in “pushbroom” fashion. Alternately, for stationary platforms, a scanning mirror (not shown) may be engaged. Thus, assuming simultaneous synchronized sampling by all the elements in said photosensitive areal detector array 36 (assured by “snapshot” mode operation now commonly available), the complete sensor performs simultaneous spectropolarimetric 1-dimensional line imaging, while spatial imaging in the second (orthogonal) dimension is obtained in a time-sequential manner.

The effect of the modulator assembly 20 is to impress polarization-dependent modulation modes upon a spatial resolution element’s (hereinafter “pixel”) intensity spectrum. Appropriate signal processing of a pixel’s modulated intensity spectrum (as projected along a row of detector elements as shown, 36) allows recovery of its intensity spectrum and also deduces its full elliptical polarization state (Stokes vector).

3.3. Existing Experimental Confirmation of IRPHSI’s P-SIM Concept

The P-SIM concept has been experimentally confirmed at Aerodyne. The low cost of components for the visible spectrum was exploited by employing an inexpensive P-SIM modulator configuration (quartz waveplates and polaroid) within an off-the-shelf visible line imaging spectrometer. Figure 1 shows the measured areal focal plane response to a uniform line source for various imposed angles of linear polarization (impressed upon a light source via a polaroid film). Observe that the fringe pattern is more complex than a monotonic (though chirped) sinusoid that would otherwise result from a mere waveplate and polarizer typical of birefringent filter designs². Rather, as will be shown in what follows, P-SIM produces a multi-period fringe pattern enabling full elliptical polarimetry.

3.4. Mueller-Stokes Perspective of P-SIM Method

The following Mueller matrix-Stokes vector formulation of the sensor measurement equation provides one of several perspective for revealing the nature of operation. The derivation of the sensor measurement equation follows from expressing the *spectral* Mueller matrix of the modulator assembly and thereby relating its output Stokes vector to the incident Stokes vector.

The Stokes parameters are a method of characterizing the ergodic average (i.e., time average over many cycles of the principal frequency components of the wave) hence conventionally observable polarization state of partially polarized waveflux. The parameters are frequently assembled into a vector and are defined as follows:

$$\begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} \equiv \begin{bmatrix} a_1^2 + a_2^2 \\ a_1^2 - a_2^2 \\ 2a_1a_2 \cos \mathbf{d} \\ 2a_1a_2 \sin \mathbf{d} \end{bmatrix}$$

where a_1 and a_2 are the ergodic real-valued orthogonal field amplitudes and \mathbf{d} is the phase lag between instantaneous realizations. The first Stokes parameter is thus the conventional waveflux intensity.

Mueller matrices or scattering matrices, form a convenient method for describing the transformations on the Stokes vector as partially polarized waveflux propagates through a system. The resulting transformed Stokes vector is merely a matrix-vector product. Therefore, the modulated spectral intensity pattern is ascertained to be the first Stokes component (i.e. intensity) of the matrix-vector product formed by frequency-dependent Mueller matrix of the modulator assembly 20 and the incident waveflux Stokes vector. Attending to Figure 2, we can build up the expression for the resultant Mueller matrix of modulator assembly 20 by determining the matrix product of sequentially ordered standard Mueller matrix representations of the series of elements, namely the primary medium 22, secondary medium 24, and linear polarizer 26.

A retarder (delay element) behaves as a frequency-dependent phase-shifter, with denoted (frequency-dependent) phase shift angle \mathbf{f} . A uniaxial birefringent medium behaves as an orientationally selective retarder by virtue of its possessing orthogonally oriented fast and slow axes of propagation. The nature of propagation through the medium is that a respectively shorter or longer propagation time delay is imposed upon the polarization component of the incident wave as resolved with respect to either axis.

The *difference* between these propagation delays can be expressed as a frequency-dependent phase shift angle \mathbf{f} as follows:

$$\mathbf{f} = \frac{2\pi n \Delta n l}{c}$$

where \mathbf{n} is the optical frequency, Δn is the birefringence of the medium (i.e., the difference between the refractive indices of its slow and fast axes) and l is the physical length of the medium (as measured along the propagation direction), and c the speed of light in vacuum. The intent of the design is most directly realized when the birefringence of the medium Δn is nearly constant with \mathbf{n} within spectral range of interest (i.e., minimally dispersive), and this will be assumed in description of operation that follows. The Mueller matrix employed herein for a uniaxial birefringent medium corresponds to its fast axis oriented parallel with the system reference axis (i.e., the axis of reference for aforementioned Stokes vector definition field amplitude component a_1).

Consequently, we form the ordered products of the Mueller matrices corresponding to each element of the modulator assembly 20, accounting for rotation of axes with respect to the system reference axis which is taken to be the fast axis 23 of the primary medium. Denoting the rotation angle of the secondary medium as \mathbf{b} and that of the linear polarizer as \mathbf{q} , the Mueller matrix of 20 is:

$$M_{\text{modulator}} = M_{\text{polarizer}}(\mathbf{q}) \cdot M_{\text{rotation}}(-\mathbf{b}) \cdot M_{\text{uniaxial}}(\mathbf{f}_2) \cdot M_{\text{rotation}}(\mathbf{b}) \cdot M_{\text{uniaxial}}(\mathbf{f}_1)$$

A most useful design configuration for elliptical polarimetry is to configure element orientations to set \mathbf{b} to 45 degrees and \mathbf{q} to zero or 90 degrees. Assuming \mathbf{q} to be zero and solving for the output first Stokes (intensity), we obtain:

$$I_n(\mathbf{f}_1, \mathbf{f}_2) = \frac{1}{4} \begin{bmatrix} I & Q & U & V \end{bmatrix} \begin{bmatrix} 2 \\ 2 \cos \mathbf{f}_2 \\ \cos(\mathbf{f}_1 - \mathbf{f}_2) - \cos(\mathbf{f}_1 + \mathbf{f}_2) \\ \sin(\mathbf{f}_1 + \mathbf{f}_2) - \sin(\mathbf{f}_1 - \mathbf{f}_2) \end{bmatrix}$$

where, with l_1 and l_2 the lengths of the primary medium 22 and secondary medium 24:

$$\mathbf{f}_1 \equiv \frac{2pn\Delta n l_1}{c}$$

$$\mathbf{f}_2 \equiv \frac{2pn\Delta n l_2}{c}$$

Thus, the resultant intensity spectrum I_n is modulated by 4 sinusoidal modulation “fringe” modes: a DC (non- modulated) mode, a “carrier” mode corresponding to the \mathbf{f}_2 term, along with sum and difference frequency (upper and lower “sidebands”) modes corresponding to the $\mathbf{f}_2 - \mathbf{f}_1$ and $\mathbf{f}_2 + \mathbf{f}_1$ terms. A modulation mode’s periodicity is:

$$\Delta n = \frac{c}{\Delta n \cdot l}$$

with the periodicity units in bandwidth (Hz) per modulation cycle. Thus, the 4 modulation mode periodicities are:

DC

$$\Delta n_c = \frac{c}{\Delta n \cdot (l_2)}$$

$$\Delta n_{low} = \frac{c}{\Delta n \cdot (l_2 - l_1)}$$

$$\Delta n_{up} = \frac{c}{\Delta n \cdot (l_2 + l_1)}$$

For an otherwise white optical spectrum incident upon the modulator assembly 20, *sinusoidal* variations as a function of n will appear across the resultant intensity spectrum. For a non-white incident spectrum, the apparent variations will appear other than sinusoidal but nevertheless are determined by the measurement equation derived above. Note that for the given orientation configuration, the identified carrier mode frequency is set by the secondary medium.

3.5. SpectroPolarimetric Data Processing for P-SIM

The fundamental observation that modulator assembly 20 imposes 4 independent modulation modes (DC, carrier, and sidebands) upon resultant intensity spectrum evokes the following expectation. Inversion of the measurement equation given 4 or more n -varying spectral intensity samples should allow solving for the 4 unknown Stokes parameters of the sensor-incident waveflux. Given or providing the appropriate regularization conditions, this is indeed the case. There are a number of conceivable methods that will undoubtedly suggest themselves to those skilled in the art for retrieving the Stokes vector and also the conventional intensity spectrum, at varying spectral resolution, from the transduced modulated intensity spectrum pertaining to a given spatial image resolution element.

3.6. Polarimetric augmentation of CTIS design using P-SIM

Non-scanning computed tomography imaging spectrometry (CTIS) provides a means of 2D imaging snapshot spectrometry ^{ur3,4}. The essential notion is to use crossed transmissive phase gratings or a computer-generated holographic (CGH) disperser within the collimated image beam. Followed by re-imaging optics, this projects spatially/spectrally overlapped regions of various spectral orders. The effect mathematically is essentially to form various projections of the 3D space-spectral image cube onto various oblique 2D planes. These various “slices”, manifested by the various spectral orders projected on the focal plane detector, can be used to tomographically reconstruct the 3D spectral image cube voxels. However, the “missing cone” drawback of such imaging (due to a finite focal plane array, and other practical factors) limits the spectral and spatial resolution of the reconstruction. To add full-Stokes polarimetry to this design without disrupting its snapshot feature, we simply insert the P-SIM modulator into the collimated beam. The resulting spectrum will be modulated, and can be decoded for polarimetry, as evidenced in earlier discussion.

ACKNOWLEDGEMENTS

This work is sponsored by the U.S. Air Force Research Laboratory, Space Vehicles Directorate (AFRL/VSBM) at Hanscom AFB, Massachusetts, previously under SBIR contract F19628-97-C-0040 and presently under SBIR contract F19628-98-C-0024. Dr. Steve Lipson and Lt. Jason Mello are the USAF Technical Program Managers.

REFERENCES

-
1. L.B. Wolff and T.E. Boult, "Constraining object features using a polarization reflectance model", IEEE Trans. Pattern Analysis & Machine Intelligence 13:7 pp. 635-657 (1991).
 2. A.M. Title and W.J. Rosenberg, "Tunable birefringent filters", Opt.Engr., 20:815, 1981.
 3. M. Descour and E. Dereniak, "Computed-tomography imaging spectrometer: experimental calibration and reconstruction results", Applied Optics 34:22 (1995).
 4. M. Descour et al, "Demonstration of a computed tomography imaging spectrometer using a computer-generated hologram disperser", Applied Optics 36:16 (1997).

Reference Numerals in Drawings

20	optical embodiment of polarimetric spectral intensity modulator assembly
22	birefringent primary medium
23	fast axis of primary medium
24	birefringent secondary medium
25	fast axis of secondary medium
26	linear polarizer
27	transmission axis of linear polarizer
28	spectrum channelizer
30	input imaging optics
32	re-imaging optics
34	slit
36	photosensitive areal detector array
38	Descour disperser
40	re-imaging optics for tomographic area spectropolarimetric sensor
42	partitioning means
44	slow channel of primary delay unit
46	fast channel of primary delay unit
48	primary delay unit
50	interunit coupling means
52	secondary delay unit
54	mixing means
60	primary medium's fast pulse
62	primary medium's slow pulse
64	coupled media's slower pulse
66	coupled media's slow pulse
68	coupled media's fast pulse
70	coupled media's faster pulse
72	simulated spectral intensity input
74	simulated modulated output

Quantifying key trade-off between IR polarimetric discriminability versus pixel resolution against complex targets

Frank J. Iannarilli, Jr. and John A. Conant

Aerodyne Research, Inc., 45 Manning Rd., Billerica, MA USA 01821

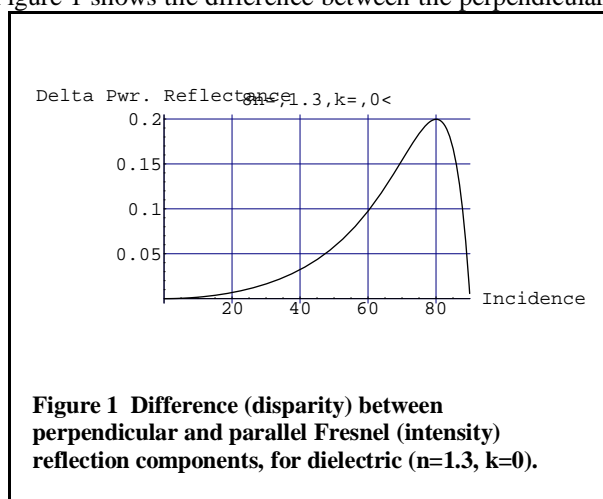
ABSTRACT

The augmentation of passive IR conventional and hyperspectral imaging sensors with polarimetric capability offers enhanced discrimination of man-made and geophysical targets, along with inference of surface shape and orientation. In our efforts to size the design of IR polarimetric hyperspectral imagers to various remote discrimination applications, we have ascertained critical relationships between polarimetric SNR and pixel sizing. This relationship pertains primarily to realms wherein the objects to be sensed will be marginally resolved spatially. The determination of such application-specific relationships is key to the design of effective polarimetric sensors. To quantify this key trade-off relationship, we have employed the latest developmental version of SPIRITS, a detailed physics-based signature code which accounts for the various geometric, environmental illumination, and propagation effects. For complex target shapes, detailed accounting for such effects is especially crucial to accurate prediction of polarimetric signatures, and thus precludes hand calculation for all but simple uniformly planar objects. Key to accurate polarimetric attribute prediction is our augmentation of the Sandford-Robertson BRDF model to a Mueller/Stokes formalism that encompasses representation of fully general elliptically polarized reflections and linearly polarized thermal emissions in strict compliance with Kirchhoff's Law. We discuss details of the polarimetric augmentation of the BRDF and present polarimetric discriminability-resolution trade-off results for various viewing aspects against a ground vehicle viewed from overhead.

Keywords: Optical infrared polarimetry discrimination polarization BRDF reflectance signature

1. OVERVIEW OF IR POLARIZATION PHENOMENA AND OBSERVABLE ATTRIBUTES

The fundamental principle behind polarimetric sensing is the fact that the parallel and perpendicular reflection coefficients are not, in general, equal. Figure 1 shows the difference between the perpendicular and parallel Fresnel (intensity) reflection components, for dielectric ($n=1.3$, $k=0$).



coefficients for a dielectric surface ($n = 1.3$, $k = 0$). The apparent signature of an elemental surface may exhibit a net polarization, the nature of which is determined by the surface optical properties, the surface temperature, the viewing angle of incidence and the effective temperatures and angular distribution of the incident radiation field. The impact of the optical properties can be characterized by the complex index of refraction, the surface roughness, and the "diffusivity". A sound representative model is one that partitions the exhibited behavior between "surface" (i.e., specular) and "volume" (i.e. diffuse) scatter components. The diffuse component (e.g., sub-surface volume multi-scatter) of reflected or emitted light is typically

unpolarized. The diffuse (i.e., rough) character of many natural background surfaces is the underlying reason why such backgrounds exhibit a decided lack of observable polarization. For the "surface" component, the (unscaled) reflectance coefficients for light polarized parallel and perpendicular to the plane of incidence are given by Fresnel's relations, dependent upon the index of refraction. Detailed discussion of the reflectance model is provided later below.

Reflected light will exhibit a net polarization vector (E-field) perpendicular to the plane of incidence, while conversely, emitted light will exhibit a net "parallel" polarization. Depending on the source intensities of surface reflected and emitted radiation, there may be a net partial polarization of observed radiation. The degree of such polarization with respect to the total exitant radiation intensity is herein termed the Degree of Polarization (DoP). For an isothermal surface and surrounding environment, the emitted and reflected polarized components cancel vectorially, yielding zero net DoP. However, with unequal reflected and thermally emitted intensities, which particularly occurs for at least some of an object's surface elements when located within angularly anisotropic radiative environments (e.g., warm earth/cold sky radiance fields), such cancellation is typically incomplete, yielding a finite net DoP. For such situations, the angle of the (E-field) polarization vector with respect to some reference direction is termed the Angle of Polarization (AoP), and depends upon the relative surface orientation. In certain (relatively unusual) configurations, a third polarimetric attribute may be observable, that of the ellipticity of polarization (EoP). Discussion of the various types of polarization descriptors and means of measurement can be found in Collett.¹ In what follows, we will assume negligible EoP, and thus that DoP and AoP correspond to linear polarization.

2. SENSOR DESIGN MOTIVATIONS FOR POLARIMETRIC SIGNATURE PREDICTION

As indicated above, the IR polarimetric observables are strongly dependent on configuration variables. In prior years there had been a vigorous if somewhat restricted debate about whether IR polarimetric observables were so fleeting and evanescent that it was not worth building sensors for their exploitation. Although computational modeling evidence indicated sufficient persistence over much of the typical range of terrestrial environmental conditions, ever-accumulating field measurements provided the decisive corroboration. At this point in time, there is general acceptance of the utility of polarimetric sensing. Nevertheless, the relevance of signature prediction to polarimetric sensing remains strong, particularly for sensor design.

Implied in the earlier discussion regarding the vectorial addition of polarimetric observables (e.g., Stokes vectors) is a fact of fundamental significance for polarimetric sensor (non-imaging) field-of-view (FOV) or (imaging) pixel instantaneous FOV (IFOV) sizing. The polarimetric observables from two differently oriented surfaces may be of significant intensity when measured separately in turn. Nevertheless, when these same surfaces are measured together in combination, the net polarimetric observables may yield negligible intensity, precisely due to vectorial combination. This explains why a polarimetric imaging sensor that measures a sphere that completely falls within a pixel's subtense will register no polarization signal, while a similar sensor with higher resolution pixels that spatially resolve the sphere will register otherwise. Thus, it is evidently of critical importance to match sensor pixel sizing to polarimetric feature spatial scale.

An object such as an aircraft or ground vehicle can be modeled as a collection of elemental surfaces or facets. The computation (prediction) of the net polarization attributes (e.g., polarized radiance, DoP, AoP) of an object, as observed within any given (partially or totally subtending) angular view subtense, is essentially an exercise in the detailed accounting per the phenomenology described above, with the vectorial components due to each surface element within the subtense properly accumulated to yield the net polarimetric signature. However, only a computer model can adequately account for the complex interplay of parameters required to compute such quantities for a complex shape such as a vehicle, especially if these quantities are to be spatially resolved (imaged).

3. IR POLARIZATION SIGNATURE PREDICTION CODE CAPABILITIES

3.1. IR polarization signature prediction using SPIRITS-AC3

SPIRITS (which stands for "Spectral and Inband Radiometric Imaging of Targets and Scenes") is a well-established and validated optical signature prediction code owned by the U.S. Government and principally developed at Aerodyne Research by one of us (J. Conant). Over the past 15 years, it has undergone numerous enhancements for various application versions. The most recent versions are denoted SPIRITS-AC, with "AC" indicating features and modules specifically useful in modeling aircraft signatures. Nevertheless, the code is sufficiently flexible to allow importing of surface temperatures computed by specialized heat transfer codes, which facilitates applicability to various other types of vehicles and objects.

A “polarized” version of SPIRITS, called “POLAR”, was first generated by Aerodyne in 1986. The POLAR code included two separate incompatible BRDF (bi-directional reflectance distribution function) models, for “polar” and “non-polar” use. The polar BRDF routines were based upon the Maxwell-Beard model^{2,3}, while non-polar computations utilized the Sandford-Robertson (“S-R”) model as with other SPIRITS versions. The capabilities of this POLAR model evolved in tandem to those of the SPIRITS model itself, all the while remaining a separate code due to the non-unified BRDF modeling. Although valid for specular, non-metallic coatings, the POLAR code implementation of the Maxwell-Beard BRDF model lacked an explicit coupling between the specular and diffuse components.

To include proper treatment of coatings possessing a diffuse reflectance component as well as metallic (e.g., complex refractive index) coatings, the most current developmental version of SPIRITS has thus been extended to compute both polarized and non-polarized target body radiances, and is denoted as SPIRITS-AC3dr1.⁴ It improves on and supplants previous versions of POLAR (polarized SPIRITS) in two ways: it is based on the latest developmental non-polarized version (AC2Dr2), and it has an improved polarized reflectance-emittance model. This model combines the Sandford-Robertson approach (used previously in SPIRITS) with the Fresnel/Mueller equations, resulting in a model which includes the most important polarization properties, is easy to use, and gives total radiances identical to those of version AC2Dr2. The new version has been given to U.S. Government (JANNAF) control, and is expected to comprise the latest configuration of mainstream SPIRITS.

3.2. Characteristics and advantages of the polarized Sandford-Robertson BRDF treatment

The Sandford-Robertson (“S-R”) model as included within SPIRITS includes emission, transmission, and reflection. Reflectance is divided into quasi-diffuse and quasi-specular portions with separate sub-models. Total reflectance and emittance are analytically coupled to conserve energy, with their relative balance an adjustable function of incidence angle. We model the specular reflections as polarized, while the diffuse reflections are depolarizing. Future upgrades could add in polarization to the diffuse contribution. At present, since we have focussed concern on opaque surfaces, transmissions are taken as neutral; they neither polarize nor depolarize the transmitted light, but merely attenuate all Stokes components by the same factor. A more detailed discussion of how we polarized the S-R BRDF treatment is contained in [4]. An overview is presented below.

In polarizing the S-R BRDF treatment, there were two main constraints we have adhered to. First, we constrain the Stokes 1 element (total radiance) to strictly match that yielded by the (non-augmented) S-R model. In this way we ensure that all previous (non-polarized) SPIRITS tests and data validations remain relevant and repeatable. Moreover, the new SPIRITS-AC3 may be run in either polarized or non-polarized mode, and will give identical results in the total radiance.

The second applied constraint is to meet the requirements of thermal equilibrium (discussed below), keeping a proper balance between polarized emissions and reflections. Our previous polarization studies have shown that in most cases the total polarization state of a 3D object results from a subtle balance of vectorially competing polarized emissions and reflections, modulated by varying surface element orientations. Specular reflectance and emission from a surface are cross-polarized (for unpolarized incident light), and they will cancel to the degree that the surface’s Planck emission matches the incident radiance field. The remaining polarization will be a “small difference of two large numbers”, and thus sensitive to modeling infidelities.

In contemplating the compliance with Kirchoff’s Law, it is important to conceptualize the situation of thermal equilibrium, which dictates that incoherent unpolarized radiation is incident upon the surface. Polarized incident light is by definition non-equilibrium. Since unpolarized incident radiation will reflect as linearly polarized radiation, the emittance must be linearly polarized in such a way as to cancel the reflected polarization and yield a net unpolarized exitance under equilibrium conditions. This also implies that elliptically polarized emission is forbidden. [This is true even though at the microscopic level when a primary reflection followed by a secondary exitant reflection yields elliptically polarized light. In this case, the primary surface will also emit linearly polarized light that exactly cancels the linear polarization of the primary reflection. Thus, the total light exitant from the primary surface onto the secondary surface is unpolarized.] This relationship between directional (polarized) emittance and scattering coefficients is similarly established by Peake.⁵ A more generalized relationship for translucent anisotropic media in non-equilibrium is derived using Landau/Lifshitz fluctuation-dissipation theory in Yueh and Kwok.⁶ A treatment in [7] establishes these fundamental properties of the BRDF from field (coherence) theory.

Unlike other signature models, the polarization predictions of SPIRITS properly include the effects of finite-width lobe specular reflectance, non-Lambertian diffuse reflectance, and polarized directional emittance derived in strict compliance with energy conservation and Kirchoff's Law. Less capable approaches typically attempt to represent surface reflectance merely with a simple Mueller matrix derived directly from the complex refractive index. Such approaches are ill-equipped to account for the effects of angularly variable diffuse reflectance, which complicates the maintenance of Kirchoff's Law compliance for the balance between reflected and emitted polarized light.

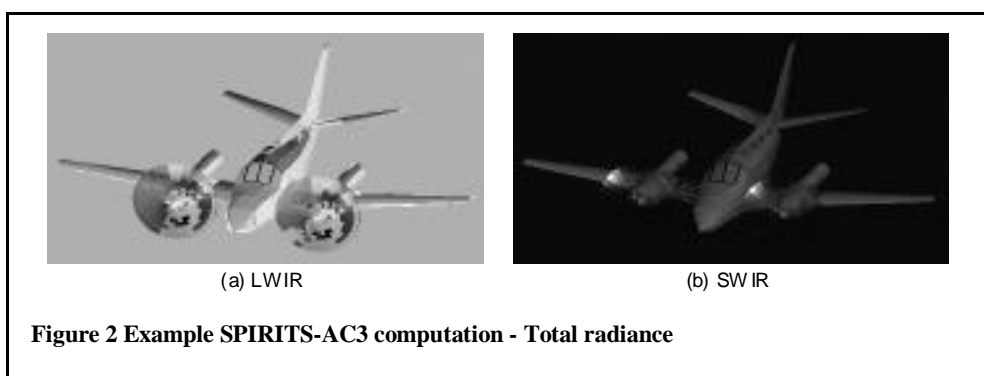
The accuracy of predicted polarized radiance due to specular reflections of extended sources is fundamentally dependent on the accuracy of the source radiance assumed. SPIRITS versions AC2 and above employ a panoramic environmental (earth/cloud/sky) representation with inwelling spectral radiances properly varying in both zenith and azimuth. This level of fidelity (as compared say to simplifications that assume hemispherically isotropic upwelling earth and downwelling sky radiance) is oftentimes critical for polarization signature prediction.

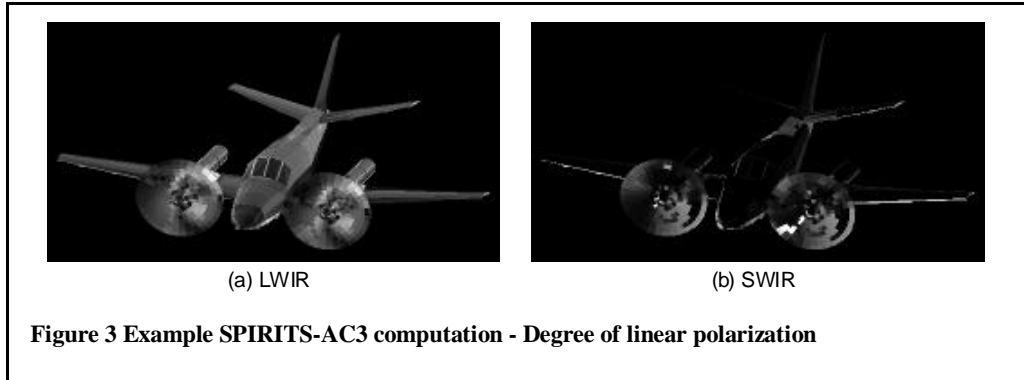
3.3. Examples of SPIRITS-AC3 polarized signature computations

An example computation using SPIRITS-AC3 is shown in Figure 2 and Figure 3. This employed a representation of a Beech Duke commercial aircraft. In this representation the propellers are modeled as a translucent disk with radially varying normal vector, capturing the effect of their time-averaged spin on a staring sensor and also the effect of propeller variable pitch. A set of realistic paints was applied, and a full radiating environment was used. Computations were made in both the LWIR (1000 – 1500 cm^{-1}) and SWIR (4000 – 5000 cm^{-1}) regions to show both thermal-dominated and solar-dominated images.

Figure 2 shows the total (Stokes 1) radiances in both bands. The LWIR image is dominated by thermal emission, which is fairly uniform across the body. Most variation across the image is due to varying amounts of reflected warm earth and cold sky, with reflected earth causing the lighter regions on the sides. The translucent time-averaged propellers reflect either earth or sky depending on their local orientation, which varies radially (propeller pitch) and azimuthally (propeller spin). The SWIR image is dominated by reflected sunlight, with glints from the propellers.

Figure 3 displays the degree of linear polarization ("DoLP"), color-coded from 0.0 (black) to 28% (white) for both bands. The DoLP is largest for facets seen near-grazing angles. Since the primary aircraft paint was given significant extinction (imaginary index) there is no Brewster Angle where a facet (in the SWIR) is 100% polarized; the angle of maximum linear polarization occurs around 80° off-normal, or 10° from grazing. Furthermore, the net DoLP will tend toward a maximum when there is minimal cancellation of the competing emission and reflection. Since the employed set of Sandford-Robertson BRDF model parameters has emittance decrease toward grazing while specular reflection increases, a DoP peak at grazing is to be expected.





4. MULTISCALE POLARIMETRIC SIGNATURE COMPUTATIONS

While SPIRITS-AC3 is an essential tool for computing the apparent Stokes imagery of complex targets, it is inefficient to iterate its computations merely in order to span across various pixelizations (spatial resolutions). Instead, our approach for gauging the impact of spatial resolution is to perform a multi-resolution decomposition of the high-resolution Stokes imagery output from SPIRITS-AC3. At each spatial resolution level, the corresponding Stokes image planes are used to compute images of various polarimetric measures such as DoP. This of course is the physically correct manner of reckoning; merely performing a multi-resolution decomposition on, say, the hi-resolution DoP image itself is incorrect.

The multi-resolution decomposition is performed using an augmentation of the classical (oversampled) Gaussian pyramid.⁸ The “pyramid” terminology is appropriate since each resolution level’s image is decimated by factor 2 (in each dimension) relative to the preceding level. Conceptually “stacking” the various resolution level imagery along a vertical axis yields a pyramid, with the finest resolution level corresponding to the original image (level “0”) at the bottom, proceeding to ever coarser resolutions toward the top. An image at the “i-th” resolution level is formed by lowpass filtering the “(i-1)-th” level image then subsequently decimating the filtered result by factor 2. The low-pass filter is typically a 2D Gaussian kernel – we employed an even-number of taps (6x6).

It can be shown that a pixel at the “i-th” resolution level corresponds to the output of an ever more finely interpolated “effective” filter kernel spanning $2^{i-1} \cdot \text{kernelWidth}$ pixels along a given dimension of the original (level 0) image. Furthermore, the “weight” of the effective interpolated kernel is preserved – we normalized our 6x6 kernel taps to sum to 1.0. The composite “phase” of this effective kernel “laydown” within the original level 0 image can be controlled by selecting the “odd” or “even” samples (for each dimension) at each decimation step. This successive downsampling-with-averaging property of the Gaussian pyramid exactly mirrors the pixel response to a scene for a physical imaging sensor as its IFOV resolution is coarsened.

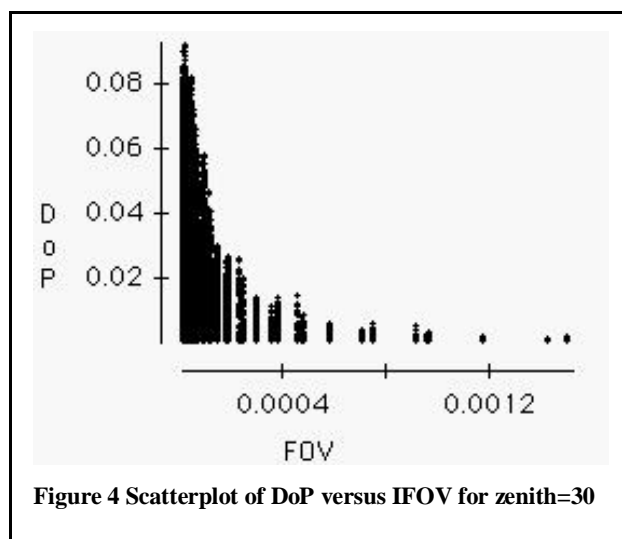
To statistically assess the impact of varying IFOV size on polarimetric discriminability, we treated each of the non-zero valued pixels (i.e., target & boundary samples) within a given multiresolution level DoP image as dwell samples pertaining to the corresponding IFOV size. At each level, we include in the sample ensemble all 4 sets of sampling “phase”, i.e., the “odd” and “even” samples for each dimension at each decimation step. Although strictly speaking, it is the polarized radiance and not the DoP that determines polarimetric SNR, for our particular example given its fairly high target effective emissivities and warm environmental reflections, SNR is sufficiently proportional to DoP across all total radiance instantiations. When reporting effective IFOV values, we quote the IFOV as the full-width-half-maximum (FWHM) of our effective (Gaussian) kernel, which we take to be half its total width. Its total width spans ± 2.5 standard deviations (s.d.), so the FWHM spans 2.5 s.d. (± 1.25 s.d.).

We took care in our implementation of such computations to preclude careless missteps that might tarnish the results. The lowpass filtering was performed using FFTs, which implies performing a circular convolution. To preclude the associated wraparound and edge effects, we padded the Stokes imagery, placing the target in the center of the pad. Pad values set to the background radiance were employed for s1 imagery (total radiance), while zero-valued pads were employed for s2-s4 (this assumes unpolarized background). Since the successive decimation performed by the Gaussian pyramid operations would eventually cause enough pad decimation to threaten wraparound effects, we adaptively re-padded during the “ascent” of the pyramid. Padding was maintained so that the distance between the target boundaries and the image borders was always

greater than the kernel width. One might dub this augmentation of the pyramid computations as the “interrupted Gaussian pyramid”. For efficiency of coding, all implementation was performed in MATLAB.

5. POLARIMETRIC OBSERVABILITY VERSUS IFOV SIZE – EXAMPLES

We model a military ground vehicle, which for expedient modeling purposes we assume is uniformly coated with a typically high emissivity specular paint. The inquiry regards observers above the ground at the following viewing geometries. Observers are swept in azimuth about the vehicle in steps of 40 degrees starting from 0 degrees (head-on), for the following discrete zenith angles: 30, 45, 60, 75, and 85. Zenith angle is measured between the local vertical (at the vehicle)



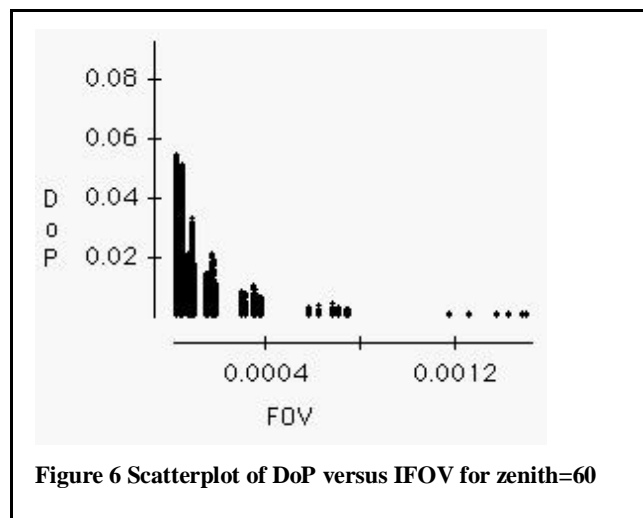
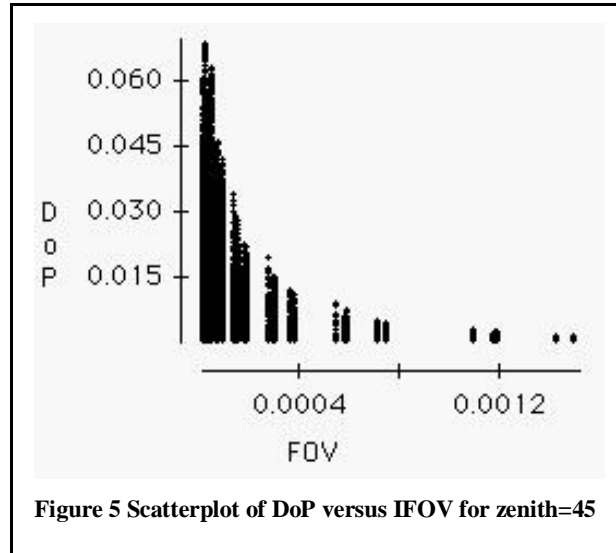
and the vector from vehicle to observer; thus 0 degrees corresponds to a directly overhead observer, while 90 degrees is a horizontally viewing observer. Signatures were computed for the 8-12 micrometer band from a range of 10km through a clear sky U.S. Standard Atmosphere, rural haze, with multiple scattering as computed by MODTRAN. The vehicle was quiescent with temperatures computed for a sunny diurnal cycle point of 1500 local time.

A number of competing factors confound simplistic expectations regarding the dependence of the distribution of pixel DoP on zenith, azimuth, and varying IFOV size. Although incidence angle, thus Fresnel disparity hence elemental DoP, increases with zenith for horizontally oriented surfaces, it decreases for those vertically oriented. Furthermore, as the Fresnel disparity of horizontally oriented surfaces increases with zenith, their projected area decreases as the reciprocal of cosine(zenith). Thus, a statistical analysis of the behavior of DoP with the factors zenith, azimuth, and IFOV size is warranted and presented below.

2D scatterplots of DoP versus IFOV lineal subtense (1 plot for each given zenith) are shown in Figure 4 through Figure 8. Regression analysis revealed no clear trend with respect to azimuth angle (even when considering trimmed quantiles), so it is a hidden factor in what we present here (although azimuth induces a slight systematic variation on realized IFOV values due to the image formation operations within SPIRITS). Each point plotted represents a pixel from within each of the 4 grid sets of pixels that fall across the target for a given IFOV size. A grid set corresponds to 1 of 4 possible dyadic sampling phases (e.g., northeast, northwest, southeast, or southwest shift by half an IFOV width). All grid set pixels are plotted, even those with meager target fill factors (e.g., those that barely include the target edge). Thus, for nearly all IFOV bins the DoP sample values will span a range extending down to near zero. Detection will be dependent largely on high fill factor pixels with correspondingly higher DoPs (i.e., not diluted by unpolarized background). Therefore, the most relevant “response surface” is what one may visualize as “draping” across the tops of the “spikes” of samples within the scatterplots.

One prominent robust trend is (visually) evident from inspection of the scatterplots. For a fixed zenith angle, the observable DoP values decrease at least superlinearly and nearly quadratically with lineal IFOV subtense. This makes sense in that the projected area increases quadratically. For this vehicle and scenario, the highest DoPs occur for zeniths of 75 and 85 and, relative to these, are somewhat lower for zeniths of 30, 45, and 60. Interestingly, the maximum DoPs achieved appear to be lowest for 60 degree zenith, with an increase in DoP as the zenith value approaches the extremes of zenith. This may be related to the fact that the neighborhoods near the extremes of zenith (0 and 90) represent the best opportunity for near-

grazing incidence viewing of horizontal or vertical surfaces (hence maximal Fresnel disparity), although as earlier remarked this competes with other effects.



An IFOV (“FOV” in the plot figures) of 400 microradians (0.0004) corresponds to a projected length of 4 meters (lineal) at a 10km range. The size of the vehicle is nominally about 4m width by 7m length by 3m height. An IFOV projected footprint of between 0.5m-1.0m (lineal) (50-100 microradians) appears to be about the coarsest sizing that maintains DoP at or greater than 5% (corresponding roughly to 10:1 polarimetricSNR), before the DoP drops precipitously to negligible noise-equivalent (e.g., 0.5%) levels. (Note: the axes of the plots are displaced outward from the data origin for clarity).

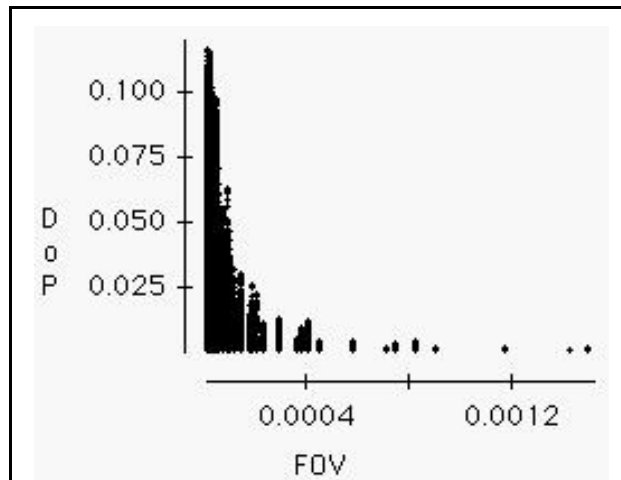


Figure 7 Scatterplot of DoP versus IFOV for zenith=75

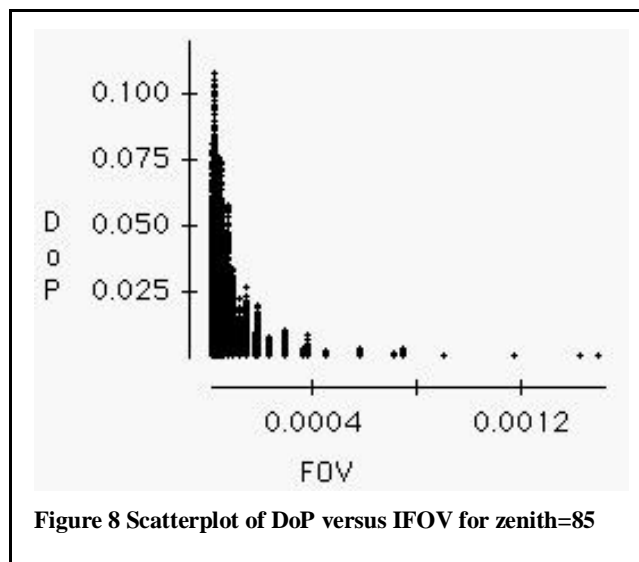


Figure 8 Scatterplot of DoP versus IFOV for zenith=85

6. CONCLUSION

These results, though rendered for a singular scenario and certainly not exhaustive, should infuse a cautionary approach into polarimetric sensor design for remote sensing and surveillance. Specifically, we demonstrate that the spatial subtense of the detector(s) is a critical design parameter – if their sizing is chosen carelessly, a negligible polarimetric signal may result when deployed against the applications of interest. These IFOV size sensitivity results need to be validated by comparison with actual measurement data. Nevertheless, a well-validated polarimetric signature prediction model appears slated to play an essential role in careful polarimeter designs of the future.

REFERENCES

1. E. Collett, *Polarized Light*, Marcel Dekker, Inc., New York (1993).
2. J.R. Maxwell and S.F. Weiner, "Polarized emittance Vol 1: Polarized bidirectional reflectance with Lambertian or non-Lambertian diffuse components", US Army Ballistic Research Lab Report BRL-CR-154 (1974).

-
3. K.K. Ellis, "Polarimetric Bidirectional Reflectance Distribution Function of Glossy Coatings", J.Opt.Soc.Am.A, 13:8 (1996).
 4. J.A. Conant and F.J. Iannarilli, Jr., "The augmentation of SPIRITS-AC2r2 for elliptical polarization", presented at JANNAF SPIRITS Users Meeting (1998).
 5. W.H. Peake, "Interaction of electromagnetic waves with some natural surfaces", IRE Trans. Antennas and Propagation (Special Supplement), December 1959.
 6. S.H. Yueh and R. Kwok, "Electromagnetic fluctuations for anisotropic media and the generalized Kirchoff's law", RadioSci., 28:4, July-August 1993.
 7. J. Greffet and M. Nieto-Vesperinas, "Field Theory for Generalized Bidirectional Reflectivity: Derivation of Helmholtz's Reciprocity Principle and Kirchoff's Law", J.Opt.Soc.Am. A 15:10 (1998).
 8. P.J. Burt and E.H. Adelson, "The Laplacian pyramid as a compact image code", IEEE Trans. Communications, 9:4:532-540 (1983).

In this work we consider the solution of multiple linear systems of equations of the form

$$Ax^{(j)} = b^{(j)}, \quad j = 1, \dots, J,$$

where $A \in \mathbb{C}^{n \times n}$ is large, sparse, and non-Hermitian and the $b^{(j)}$'s are all available simultaneously. Such problems arise, for instance, in the numerical solution of frequency-domain electromagnetic wave scattering; in this case, the right hand sides might correspond to different incident fields on the scatterer induced either by plane waves incident at various angles or by excitation sources at different locations. We propose a new technique, based on the quasi-minimal residual (QMR) algorithm, to solve these types of problems. Our approach is efficient in both computation time and storage.

The naive approach of solving each of the J linear systems independently using a Krylov subspace method does not take advantage of the fact that the $b^{(j)}$'s, and hence the $x^{(j)}$'s, may be closely related due to the underlying physical nature of the problem. By closely related, we mean that the solution to the j th system has large components in the directions of the Krylov subspace generated from one of the other systems. Projection-type techniques for both the Hermitian and non-Hermitian cases that specifically exploit such shared information have been proposed (e.g., see [?, ?] and the references therein).

Another alternative is to use a block algorithm, such as block QMR [?, ?], to solve the systems simultaneously. However, this approach can be more expensive in terms of storage than projection techniques because the length of the recurrences to update the iterates depends on the number of right hand sides, or, in the case of deflation [?], the number right hand sides corresponding to the deflated Krylov subspaces. Further, the residuals themselves are not available at every iteration. Also, if a deflation technique is used, a deflation tolerance must be specified in advance, and we have found in experiments that the performance and convergence of the systems depends in an unpredictable way on this value. Therefore, we pursue the idea of a projection-type technique and then present a block variant of our algorithm that exploits good properties of the block QMR algorithm while preserving the basic properties of our projection-type technique.

The idea of a projection-type technique is to choose one of the J systems to be the “seed” system and apply some type of Krylov subspace method to it. For the non-Hermitian case, choices include conjugate gradient for least squares (CGLS), biconjugate gradient (BiCG), conjugate gradient squared (CGS), and generalized minimal residual (GMRES). As the relevant subspaces are generated, the approximations to the other systems are simultaneously updated by projecting the residual onto a particular subspace and by either enforcing a Galerkin-type condition [?, ?] or by minimizing the projected residual [?]. Our algorithm is similar to the project-minimize approach but has certain advantages over the algorithm in [?]; namely, we do not need to store the basis vectors, we do not need to predetermine a subspace dimension, and the approximate solutions and residuals are cheaply computed and available at every stage of the algorithm because they are updated with short-term recurrences.

Our algorithm proceeds as follows. First, we select one system, say system

j , to serve as “seed” and apply QMR (without lookahead) to the seed system. In the following, we use $r_0^{(j)}$ to denote the initial residual to system j given the starting guess $x_0^{(j)}$ and $r_k^{(j)}$ to denote the residual of system j after k iterations. As QMR progresses to the $k+1$ st iterate, it generates bi-orthogonal bases for two k -dimensional Krylov subspaces, $K_{k+1}(A, r_0^{(j)})$ and $K_{k+1}(A^T, r_0^{(j)})$. We denote the respective bases by the columns of V_{k+1} and W_{k+1} . By bi-orthogonality, $W_{k+1}^T V_{k+1} = D_{k+1} = \text{diag}(\delta_1, \dots, \delta_{k+1})$. For the seed system, the k th iterate is obtained by minimizing the norm of a quasi-residual, which amounts to solving a least squares problem involving a $(k+1) \times k$ tridiagonal matrix. That quasi-residual can also be thought of as a projected residual; that is, if $r_k^{(j)}$ denotes the k th residual of the seed, the k th iterate is taken to be $x_0^{(j)} + V_k z_k^{(j)}$, where $z_k^{(j)}$ denotes the vector in \mathcal{C}^k which minimizes the norm of $D_{k+1}^{-1} W_{k+1}^T r_k^{(j)} = D_{k+1}^{-1} W_{k+1}^T (r_0^{(j)} - A V_k z_k^{(j)})$.

We extend this idea to solve the remaining systems. Thus, if $r_k^{(i)}$ denotes the k th residual of the i th non-seed system, the corresponding k th iterate is determined to be $x_0^{(i)} + V_k z_k^{(i)}$ where $z_k^{(i)}$ minimizes the norm of $D_{k+1}^{-1} W_{k+1}^T r_k^{(i)}$. Alternately, the iterates to the non-seed systems can be thought of as minimizing a quasi-minimal residual. After the seed system converges, the current iterates are used as starting guesses and the process is repeated by selecting a new seed from among those that have not converged. The whole process repeats until all the systems have converged. Although it is not obvious from this discussion, we note that no extra Lanczos vectors need to be stored to form the k th iterates of the non-seed systems. We expect our algorithm to have better convergence behavior than BiCG-projection approaches just as QMR behaves more stably than BiCG on a single problem.

The dominant cost in our algorithm is the total number of matrix-vector products. In the worst case, if the solutions have very little shared information, we expect the cost to be comparable to the method of solving each system separately. However, if the Krylov subspace generated by the seed and the Krylov subspace generated by a non-seed are close, we expect that only a few restarts, relative to the number of right hand sides, will be necessary, and thus significantly reduce the number of matrix-vector products computed overall. Along the lines of the analysis in [?], we show that under certain conditions, when the process is restarted with a new seed, the convergence rate of the residual of that seed system behaves as if the extreme ends of the of the spectrum of A are cut off. We are currently working on bounds for the remaining residuals after each restart. Indeed, our numerical results indicate that the computational savings of our algorithm over the approach of solving each system separately can be significant.

As mentioned, the performance of the single seed projection-type algorithms relies on presence of common information among the right hand sides. If little such information is found, one way found to generalize our algorithm to address this difficulty is to take m ($1 < m < J$) systems as seed, say those with indices in I_s , and use block QMR to solve the seed systems. Then, as with the single

seed algorithm, we project the residuals of the non-seed systems and solve the corresponding minimization problems. Thus, the Krylov subspace in which approximate solutions are sought is the km dimensional subspace

$$K_{km}(A, B_s) = \text{span}\{B_s, AB_s, A^2B_s, \dots, A^{k-1}B_s\}, \quad B_s = [b^{(i_1)}, \dots, b^{(i_m)}], \quad i_m \in I_s.$$

We use the implementation in [?], which incorporates deflation to drop converged seed systems and to delete (nearly) linearly independent vectors in $K_{km}(A, B_s)$ and $K_{km}(A^T, B_s)$. Methods for selecting optimal choices for m and I_s are topics for future research.

Therefore, if the right hand sides of the seed systems are not very closely related, they must each be contributing unique information to the Krylov subspace, and we expect convergence of the seed systems to occur for small k . Further, we expect that initial guesses of non-seed systems obtained by projection onto this subspace will be better. As long as the number of seed systems is not too large, there are still short term recurrences which generate the solution estimates at each iteration. We are currently exploring the theoretical convergence properties of our multiple seed approach.

ABSTRACT FOR the SIAM ANNUAL MEETING in TORONTO 1998

Ill-conditioned matrices with Toeplitz-type structure arise from the discretization of certain ill-posed problems. We use a preconditioned Krylov subspace method to compute a regularized solution to this linear system given noisy data. Our preconditioner is a rank- m approximation to the Cauchy-like matrix obtained by applying a fast, symmetry-preserving orthogonal transform to the original system. We show that the preconditioner has desirable properties, give a method for applying it quickly, and illustrate its effectiveness on examples.

ABSTRACT FOR SIAM ANNUAL MEETING IN Atlanta, 1999

We consider the solution of large, sparse, non-Hermitian, linear systems with multiple right hand sides. We use QMR to solve the seed system and generate bi-orthogonal Krylov subspaces; approximate solutions to non-seed systems are generated by minimizing their appropriately projected residuals. The process is repeated by selecting a new seed from among the non-converged systems and using previously generated approximate solutions as initial guesses. Convergence theory is given and algorithmic efficiency is illustrated on examples.

ABSTRACT FOR PLEMMONS BIRTHDAY CONFERENCE, 1999

Discrete ill-posed problems in the form of large linear systems or least squares problems occur in a variety of applications. Due to the presence of noise, regularization methods must be used to determine a solution that approximates the noise-free solution. A wise choice of regularization parameter, which controls conditioning, is crucial to obtaining a good approximate solution. If the dimension of the discrete problem is large, however, selecting a parameter and solving the regularized problem can be costly. For this reason, we compute an approximate solution by projection onto a smaller dimensional subspace via Krylov subspace-based iterative methods. We show how techniques designed for choosing the regularization parameter for the original problem can also be used effectively on a regularized form of the projected problem. We prove some results on the approximate equivalence of this approach of regularizing after projection to other forms of regularization, which operate on the larger problem directly. The success of these techniques and the computational savings is illustrated in examples.

FOR TALK AT UMASS DARTMOUTH, 1999

Many problems in image processing require the solution of a linear system or least squares problem resulting from the discretization of a first kind integral equation. The exact solution to such problems is often hopelessly contaminated by noise, since the discretized problem is very ill conditioned and noise is usually present in the data; hence, regularization must be employed to approximate a solution. In this talk, we will consider a wavelet-based regularization scheme for use in linear image restoration problems. We introduce the L-hypersurface method for determining the multiple regularization parameters associated with our regularization scheme. Since the computational cost of selecting parameters by means of the L-hypersurface can be prohibitive, we also explore cost efficient variants. The effectiveness of our regularization approach and parameter selection techniques will be illustrated on several examples.

Hyperspectral Imaging Using AOTF and NIR Sensing of Buried Objects and Landmines

J. Mobley¹, G. H. Miller¹, P. M. Kasili^{1,2}, C. DiMarzio³, and T. Vo-Dinh^{1,2(*)}

¹Oak Ridge National Laboratory, Oak Ridge, TN 37831-6101

²University of Tennessee, Knoxville, TN 37916

³Northeastern University, Boston, MA 02115

(*) Author for Correspondence: tvo@ornl.gov

Abstract

The detection of landmines and buried objects requires methods that can cover large areas rapidly while providing the required sensitivity to detect the optical and spectroscopic contrasts in soil properties that can reveal their presence. These conditions on contrast and coverage can be met by capturing images of the soil at wavelengths which are sensitive to the properties modified by the presence of buried objects. In this work we investigate both imaging and scanning methods which may have some utility for the detection problem. In the imaging approach, we capture hyperspectral reflection images using an acousto-optic tunable filter (AOTF) and fluorescence images using a long-pass filter. For the scanning method, we acquire data point-by-point over a two-dimensional grid with a single emitter/detector pair. The results illustrate the potential of these two approaches for detection of landmines and buried objects.

I. Introduction

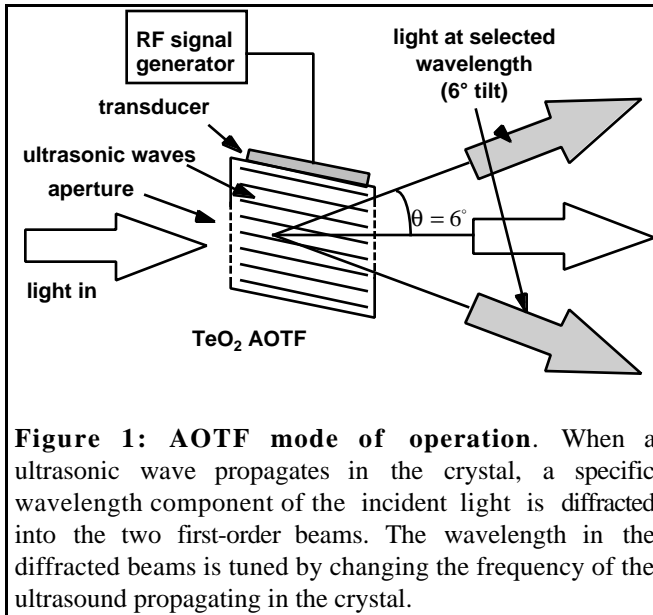
Indirect methods of detecting buried objects involve looking for differences between undisturbed soil and soil directly above a buried object. The detection of landmines and buried objects requires methods that can cover large areas rapidly while providing the sensitivity to the contrasts in soil properties that can reveal the presence of foreign objects. The conditions on coverage can be met by acquiring images of large areas or rapidly scanning over the ground. The contrast is provided by acquiring the data at specific wavelengths or over continuous spectral bands which are particularly sensitive to the properties affected by the presence of buried objects. In this work we look at both imaging and point-by-point (scanning) methods which can be adapted for use in the detection problem. The images were acquired by a charge-coupled device (CCD). Two types of images were captured: (1) narrowband reflection images using an acousto-optic tunable filter (AOTF) and (2) fluorescence images using a long-pass filter. The scanning method employed a near-infrared (NIR) emitter/detector pair where data were captured point-by-point over a two-dimensional grid.

II. Imaging-Based Methods for Object Detection

IIa. Reflection Imaging with an Acousto-Optic Tunable Filter

By taking two-dimensional images of a target zone, a large area can be spectrally mapped quickly given that the contrast between the background soil and burial sites are sufficiently strong to allow rapid detection. One method of increasing contrast is to image at/over wavelengths where the differences in reflection or fluorescence properties are maximized. These types of images can be captured rapidly by using an acousto-optic tunable filter (AOTF) for wavelength discrimination. The AOTF is

a narrowband optical filter with a electronically-selectable passband and a two-dimensional aperture. It can quickly jump between remote wavelengths or be rapidly scanned across some continuous spectral range. Band-limited (i.e., a spectral range wider than the resolution width of the filter) imaging can be performed by scanning the AOTF across a specific range of

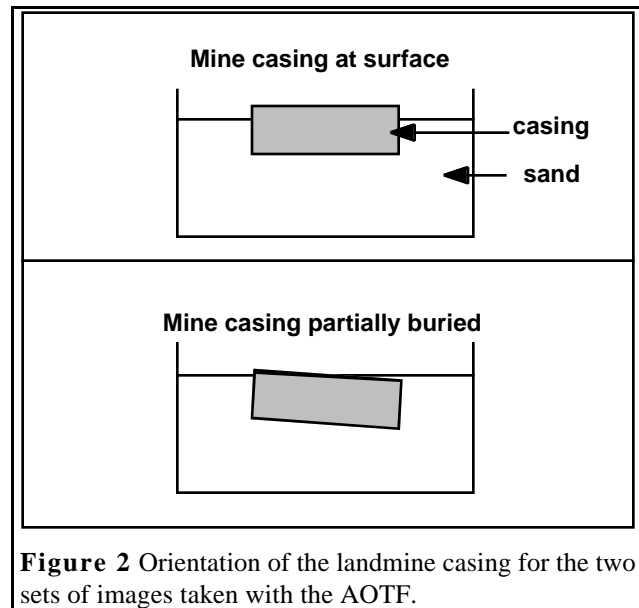


light is discriminated by the TeO₂ AOTF used here.

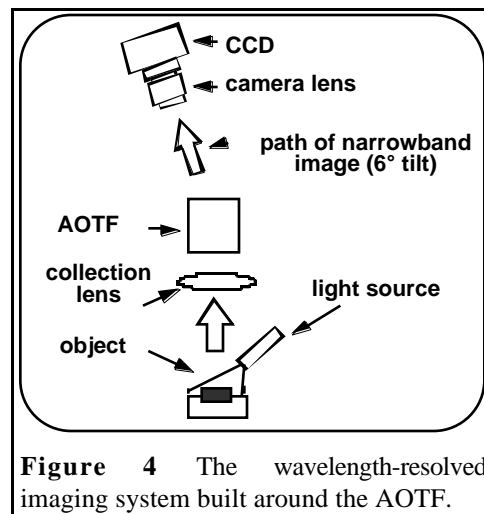
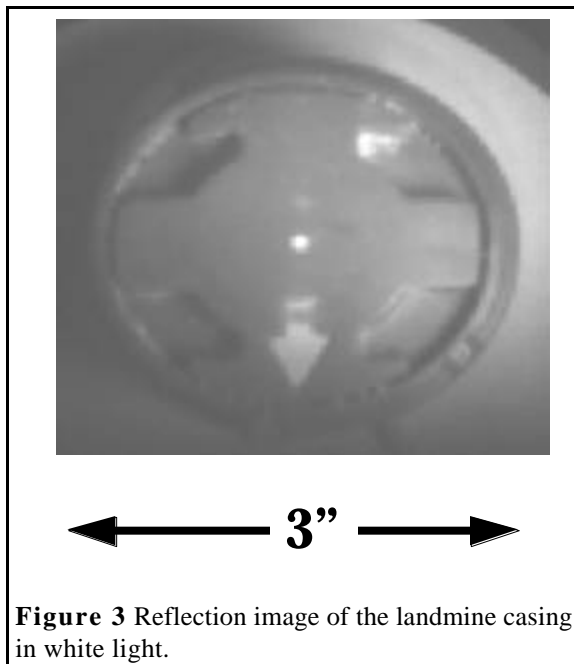
In this work our imaging system was limited to wavelengths shorter than 700 nm. Thus to demonstrate the capabilities of an AOTF-based system to make wavelength-resolved images, we have examined a landmine casing both above and partially buried in sand (see Figure 2) illuminated by an incandescent lamp and imaged at wavelengths from 525 to 700 nm. A white-light image of the casing is shown in Figure 3. The imaging system is diagrammed in Figure 4 and employs a TeO₂ AOTF (Brimrose, Inc. tunable from 450 to 700 nm), a thermoelectrically cooled CCD (Santa Barbara Instruments, Model ST-6) and a camera lens with adjustable aperture and focus. At each wavelength, in addition to the landmine casing, an image of a white card was also recorded along with the landmine casing for use in data normalization. The normalization process is diagrammed in Figure 5.

wavelengths while exposing the image-capturing elements (e.g., film or charge-coupled device array (CCD)). Our laboratory has previously reported on the use of AOTF/CCD based systems for hyperspectral imaging[1-3].

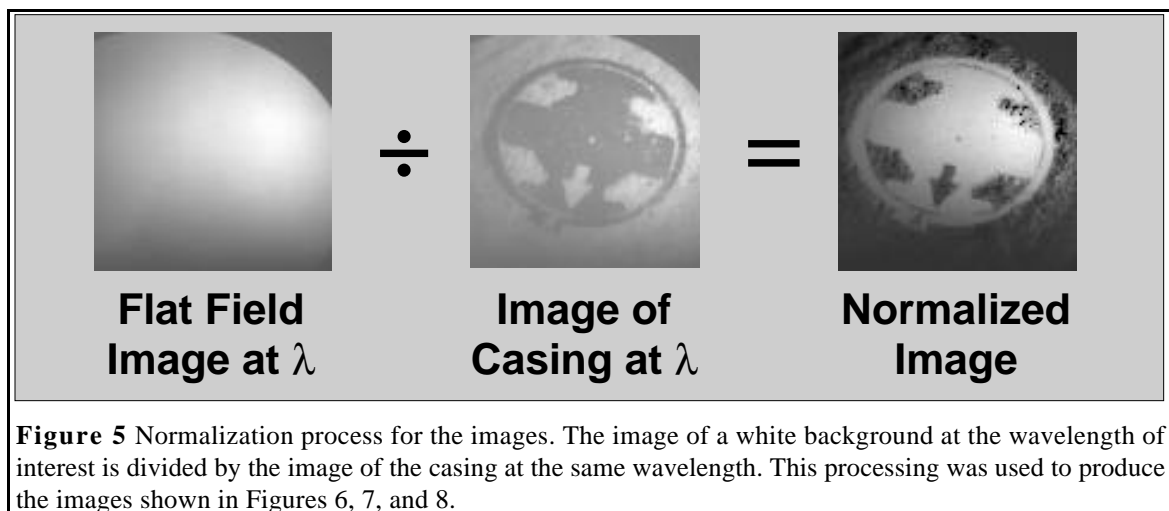
The AOTF operates by propagating ultrasonic waves through an anisotropic crystal. The strains induced by the ultrasound act to spatially modulate the indices of refraction of the crystal. The device is essentially a three-dimensional grating whose spacing can be tuned by changing the wavelength of the ultrasound in the crystal. Since the AOTF has no moving parts and can be rapidly moved to the desired wavelength, it is ideal for use in portable instrumentation where spectral selectivity and timely feedback are required. Figure 1 shows the how the



The normalized images for the exposed casing are displayed in Figure 6. The casing has five indentations on its surface (four around the rim and one arrow-shaped) which were filled with sand in the wavelength-resolved images. The best images in



terms of clarity and contrast between the sand and casing are the three images taken at 600 nm, 625 nm, and 650 nm respectively. The images of the partially buried casing, as displayed in Figure 8 (Figure 7 gives the orientation of the casing as seen in Figure 8), showed similar results. For these, the 625-nm and 650-nm images display the best contrast. The resulting contrasts in these reflected-light pictures are of course related to the respective colors of the casing and sand. The casing is mostly green and as one moves to longer wavelengths (starting from 525 nm) the casing reflects less light while the background sand reflects more evenly across the spectrum. Thus the contrast between the two increases as one moves from green wavelengths to the red end of the spectrum. Future work will involve investigating local reflection properties of the soil at NIR wavelengths. By performing narrowband or wavelength-resolved imaging, the contrast in soil properties can be maximized.



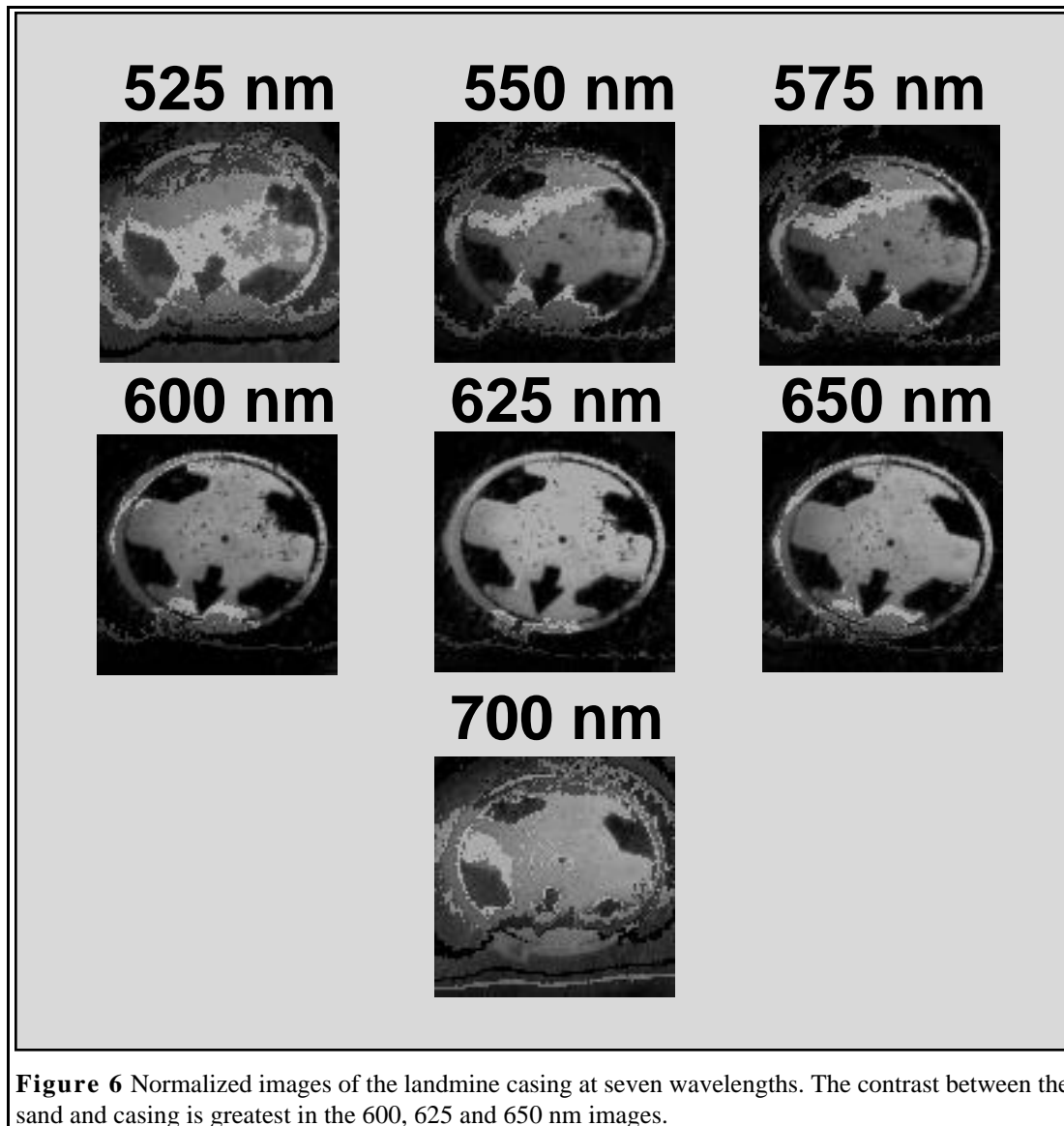


Figure 6 Normalized images of the landmine casing at seven wavelengths. The contrast between the sand and casing is greatest in the 600, 625 and 650 nm images.

Iib. Fluorescence Imaging

In addition to reflected light, we also captured images of the fluorescence emission from the partially-buried object. In this mode, the AOTF was replaced by a 525-nm long pass filter, the focusing lens/camera were axially aligned with the collection lens and the object was illuminated by the 488-nm line of an Argon laser. The captured image is shown in Figure 9. The high contrast between the casing and sand is readily apparent. If a condition of the soil above a buried mine can be detected via fluorescence, this method could be useful in low-light situations where the fluorescence can be measurably detected over ambient reflected light.

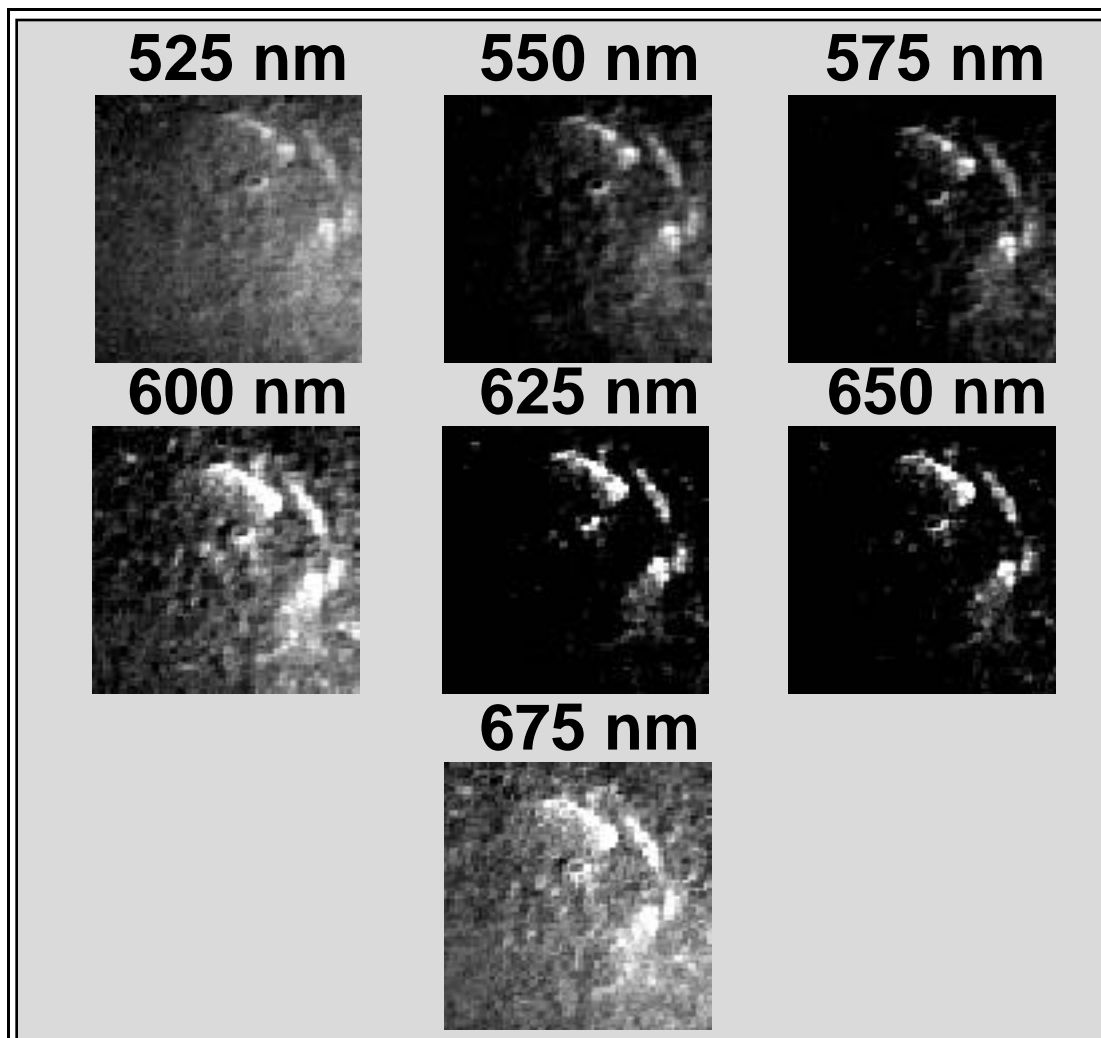
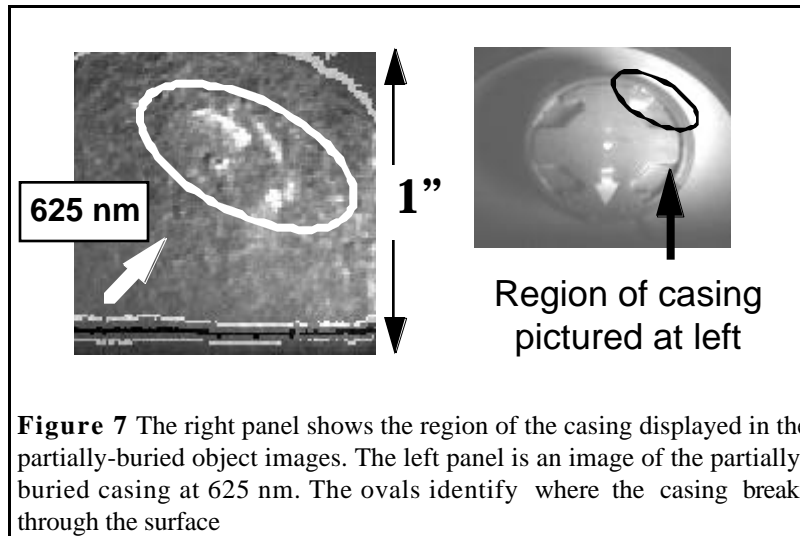
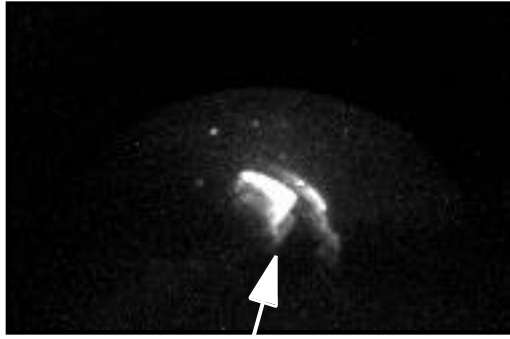


Figure 8 Normalized reflection images of the partially-buried landmine casing at seven wavelengths. The degree of contrast between the sand and casing is greatest in the 625 and 650 nm images.



Exposed portion of casing

Figure 9 Fluorescence image of the partially-buried landmine casing. The sample area was illuminated with the 488 nm line of an Argon laser. A 520 nm long pass filter was used to remove the background.

III. Point-by-Point Scanning in the NIR

In contrast to the direct imaging methods, we also performed a point-by-point reflection-mode scan in a 2-D grid pattern at NIR wavelengths over the surface of sand with wet and dry areas. The idea behind this experiment is that the soil above the buried object may contain less moisture - the flow of moisture being drawn up to the surface (due to concentration gradients stemming from surface evaporation) may be affected by the presence of the object. A casing for a landmine was buried under 5 mm of material in a tray that measured 30x16x8 mm (LxWxD). In order to simulate the dry and wet areas found in the field, a piece of plastic wrap was used to separate the dry material from the wet. Using an infra-red emitter/detector matched pair operating at 915 nm, a light-emitting diode (LED) and a photodetector were mounted side-by-side on perforated board and soldered in place. The orientation was chosen to maximize the detection of reflected IR from the infrared LED. The photodetector has a sensitive area of about 1 mm² while the LED has a collimating lens mounted in place and a radiant power of 0.5 mW. The detector and experiment are depicted in Figure 10.

The photodetector was used in the passive mode and voltages were measured with a digital multi-meter. The device was positioned using an optical rail system marked in mm. The LED and photodetector were held about 30 mm above the surface of the soil and a reading was taken, in a darkened

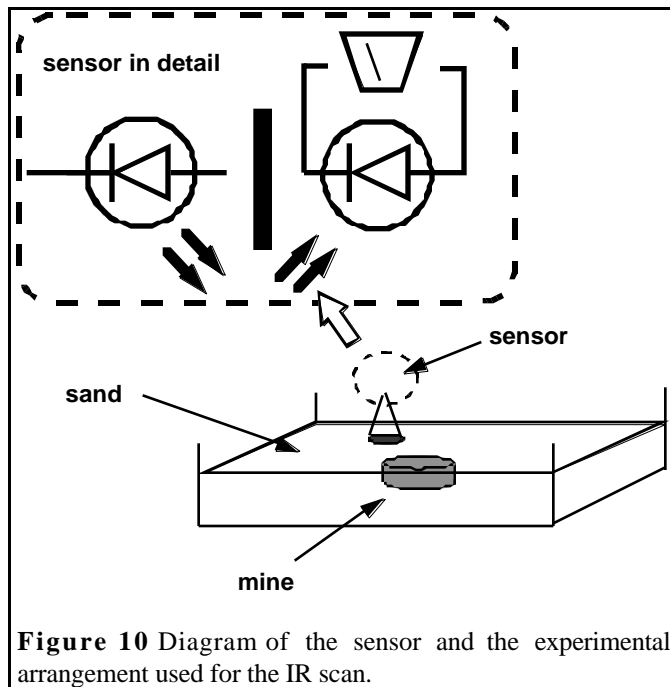


Figure 10 Diagram of the sensor and the experimental arrangement used for the IR scan.

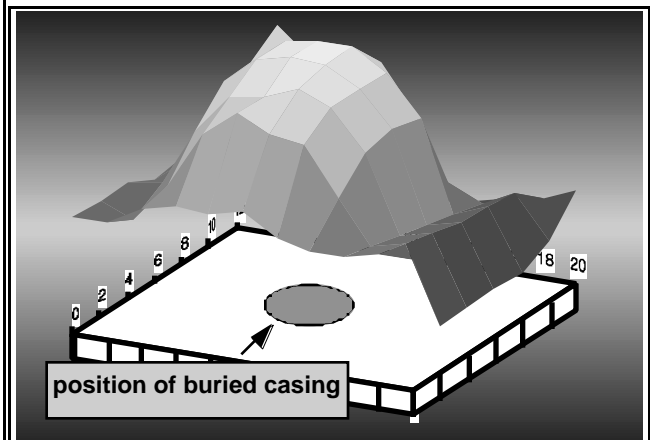


Figure 11 Result for the IR scan of the buried object. The x and y axes are position and the z-axis is the photodiode voltage in linear units.

room, every 20 mm in the x and y direction. The surface scanned in this procedure covers a 12x20 cm area. The resulting image is illustrated in Figure 11.

The experiment provided an image of medium resolution where the wet sand reflected enough radiant energy to produce a signal on the order of 27 - 40 mV and the dry sand that of 50 - 60 mV. The variation of the signal at any given point was about one millivolt. The device was able to differentiate between wet and dry material and produce an image that roughly corresponds with that of the buried landmine. Further work will focus on improving the sensor and more accurately simulating field conditions in the laboratory.

Conclusion

In this paper, we have reported on some preliminary work evaluating methods that could be used in the detection of buried objects. We examined imaging techniques using both reflected and fluorescent light as well as point-by-point scanning with a single infrared emitter-detector pair. Each of the methods shows some promise and will be evaluated under more stringent conditions in the future.

Acknowledgments

This research is sponsored by the US Army MURI Project (US ARMY ARO/DAAG-55-97-1-0013) and by the Office of Biological and Environmental Research, U.S. Department of Energy under contract DE-AC05-96OR22464 with Lockheed Martin Energy Research Corporation.

References

1. F. Moreau, D. M. Hueber, and T. Vo-Dinh, Remote Spectral Imaging System (RSIS) Based on an Acousto-Optic Tunable Filter (AOTF), *Instr. Sci Tech.*, 24 (3), pp. 179-193 (1996).
2. F. Moreau, S. M. Moreau, D. M. Hueber, and T. Vo-Dinh, Fiber-Optic Remote Multisensor System Based on an Acousto-Optic Tunable Filter (AOTF), *Appl. Spectrosc.*, 50 (10), pp. 1295-1300 (1996).
3. A. D. Campiglia, D. M. Hueber, F. Moreau, and T. Vo-Dinh, Phosphorescence Imaging System Using an Acousto-Optic Tunable Filter and a Charge-Coupled Device, *Analytica Chimica Acta*, 246, pp. 361-372 (1997).

Utilizing NIR Emitter/Detector Pair for Detecting Buried Objects

G. H. Miller¹, S. C. Culbertson^{1,2}, J. Mobley¹, C. DiMarzio³, and T. Vo-Dinh^{1,2(*)}

¹Oak Ridge National Laboratory, Oak Ridge, TN 37831-6101

²University of Tennessee, Knoxville, TN 37916

³Northeastern University, Boston, MA 02115

(*)Author for Correspondence: tvo@ornl.gov

Abstract

The sand above a shallow buried object can be differentiated from the surrounding sand by detecting the difference in relative water content. Moist sand absorbs more near-infrared (NIR) incident light than dry sand. A light emitting diode operating at 900 nanometers and a photodiode sensitive to NIR radiation formed the emitter/detector pair used in this study. The pair were mounted side-by-side and scanned over a surface in a two-dimensional grid as readings were collected point-by-point. The results indicated that this simple NIR emitter/detector pair discriminated between sands of varying water content and possessed an imaging resolution of 4 millimeters.

I. Introduction

Remote detection of a shallow-buried-object requires identifying a property to distinguish the undisturbed soil and the soil directly above the buried object. One characteristic that may be exploited for such differentiation is the water content of the surface soil. The water content of the soil directly above a buried object can differ from that of undisturbed soil and the difference can be detected using near-infrared reflectance. The detection of landmines and buried objects requires methods that can cover large areas rapidly while providing required sensitivity to detect the optical and spectroscopic contrast in the soil properties that reveal their presence. These conditions on contrast and coverage can be met by capturing images of the soil at wavelengths which are sensitive to the properties modified by the presence of buried objects. Eventually these techniques can be extended to rapidly scan large areas by using NIR photography. Previous work has shown how powerful an imaging system is in detecting buried objects [1]. The work in this paper illustrates the usefulness of the NIR optical region for detecting buried objects and leads directly to NIR imaging techniques.

II. Experimental

In this study, a probe was constructed to examine the ability of NIR reflectance to detect differences in the moisture content of sand. The emitter and detector were mounted side-by-side on perforated circuit board with a current limiting 1K resistor and a 9V battery for the source. The detector was used in the passive mode and had a 12 millimeter long opaque tube affixed to its casing. This tube protected the photodiode from side scatter of the LED and enhanced the image resolution. Light reflected off of the sand surface was captured by the photodiode, which generated a voltage level that was read by a digital multimeter.

Sand (Quikrete Play Sand) was placed into a plastic tub that measured 13X7X3.5 inches. Plastic petri dishes measuring 10 millimeters high and 35 millimeters in diameter were placed in the sand so that their tops came flush with the surface. On the average, one of these petri dishes holds 13.3 grams of sand. Three of these dishes were placed in the tub and various amounts of

water were added to each dish. The emitter/detector pair was mounted onto the end of an optical rod and this mounted onto an optical rail scored in millimeters. The configuration positioned the pair to be looking down onto the surface. This experimental configuration was used to optimize the distance between the surface and the emitter/detector pair to determine the response of the device to various amounts of water in sand and the resolution of the optimized system.

Resolution was determined by finding a bright reflective surface (white paper) and a good absorber (black paper) and mounting them to produce a sharp boundary perpendicular to the axis of travel of the emitter/detector pair. A step size of 2 millimeters provided sufficient accuracy to determine the imaging resolution of the system. The resolution of the detector was adjusted by placing a black plastic tube over the active element to create a limiting aperture. By increasing (decreasing) the length of the tube, the resolution is increased (decreased) as well. A longer tube length, however, is also accompanied by a decrease in signal. We determined the resolution of a 12 millimeter tube to be on the order of 4 millimeters (Figure 1). We found 4 millimeter resolution gave us a reasonable balance of spatial sensitivity and signal-to-noise. The final experimental configuration is illustrated in Figure 2.

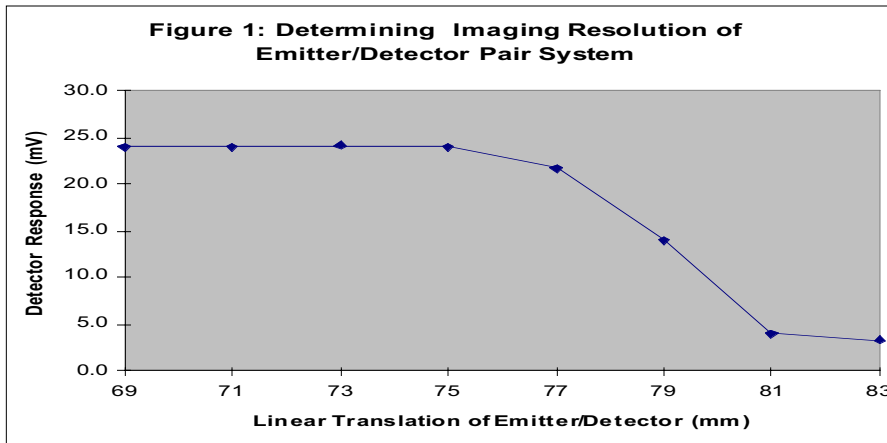
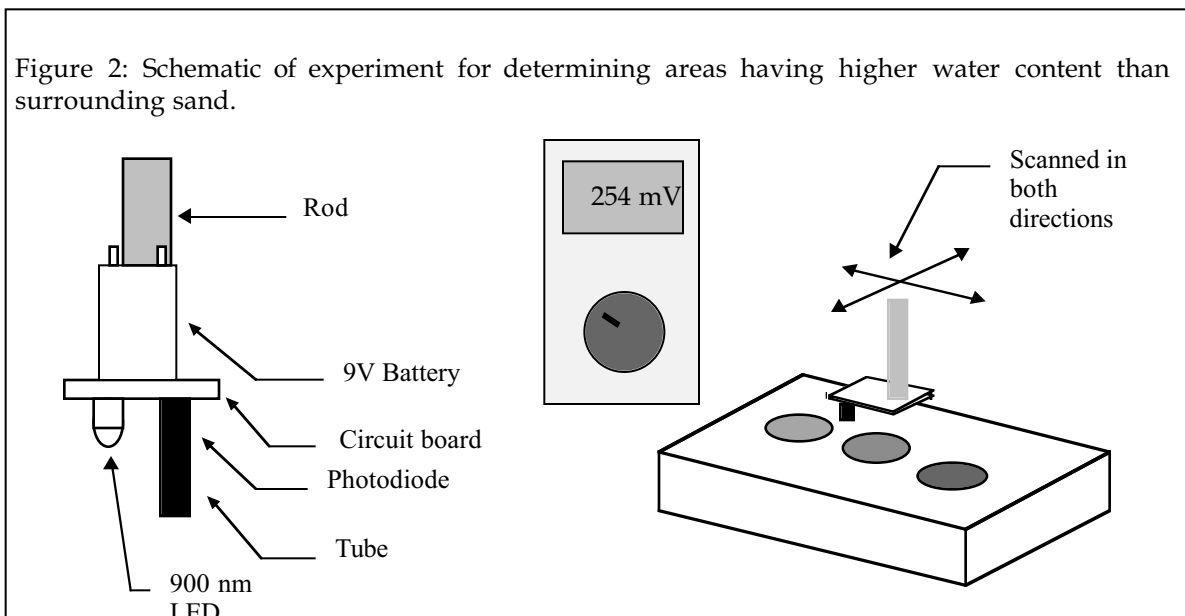


Figure 2: Schematic of experiment for determining areas having higher water content than surrounding sand.



The scans were made in a darkened room and the device was moved in 5 millimeter increments in both directions. The sand was used as is; there were no special preparations made such as baking or washing. To determine the amount of water per gram of sand three petri dishes were chosen at random, tared, filled with sand (overfill and scrape level method), and weighed when filled. Each was weighed filled three times generating nine data points. The weights were consistent as seen in table 1.

Table 1

	Petri Dish 1	Petri Dish 2	Petri Dish 3
Net Weight 1	13.26 g	13.66 g	12.96 g
Net Weight 2	13.15	13.63	13.02
Net Weight 3	13.02	13.54	13.39
Mean \pm St. Deviation	13.14 \pm 0.12 g	13.61 \pm 0.06 g	13.12 \pm 0.23 g

All nine results give a mean weight of 13.29 ± 0.27 gram. In preparing this proof-of-principle experiment, it was desirable to maintain a constant distance between the detector and the surface of the sand. In accomplishing this, material was moved into and out of the petri dishes several times as the surface and the dishes were scraped over by a steel ruler. The leveling procedure makes it impossible to weigh a known amount of material into a petri dish. However, the standard deviation of the weights and the consistency between dishes shows that relatively little error was introduced by not weighing the sand in each dish. Figure 3 gives a graph of the response of the emitter/detector pair when viewing sands containing various amounts of water. Dry refers to untreated sand. The unit (milliliters/gram) given in the graph refers to the amount of water added to untreated sand.

Figure 3: Response of System to Soils Containing Various Amounts of Water

mL / gr	Mean	St. Dev.	Data 1	Data 2	Data 3	Data 4	Data 5
(Dry) 0	15.5	0.122474	15.6	15.5	15.3	15.5	15.6
0.037	13.26	0.320936	13.3	13.4	13.5	12.7	13.4
0.075	11.8	0.291548	11.4	11.9	11.8	12.2	11.7
0.113	11.12	0.130384	11.2	11.1	11.2	11.2	10.9
0.15	10.6	0.2	10.7	10.4	10.5	10.5	10.9
0.188	10.58	0.130384	10.5	10.6	10.7	10.7	10.4
0.226	10.22	0.083666	10.2	10.3	10.3	10.1	10.2

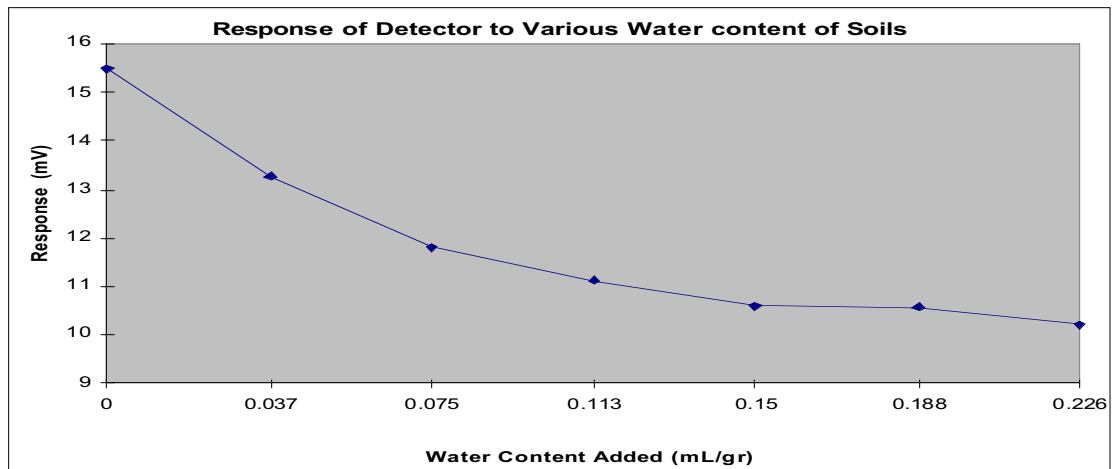


Figure 4 illustrates the image resulting from a point-by-point scan of the surface. The petri dishes contained sands of higher water content than the surrounding sand. This image was produced by moving the detector in 5 millimeter increments. A grayscale was assigned to the numeric values so that the dry sand produced a light gray pixel and the wettest sand produced a black pixel.

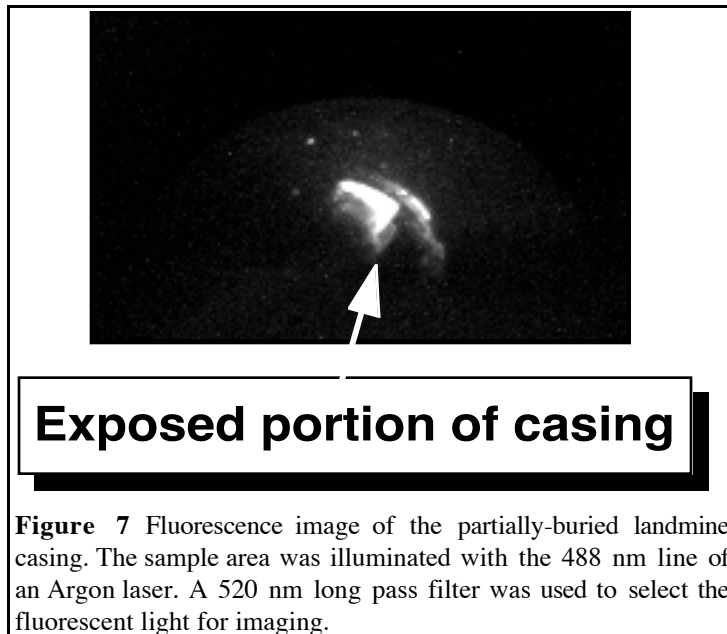
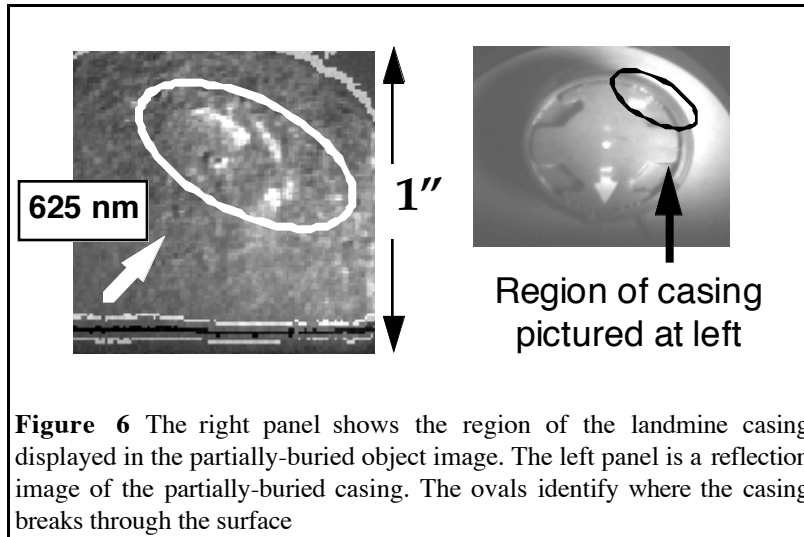
The data may be represented in a three-dimensional form (Figure 5). The regions of higher water content are illustrated as having greater height and the variation in the dry sand is illustrated in the baseline. The values of the z-axis are in descending order to produce peaks where there is greater water content. The value of the signal is smaller in these regions.

Figure 4: Two-Dimensional Representation: Response of the NIR Detector to Sand With Various Water Content. The x and y axes represent the distance in millimeters and the gray scale is the voltage in millivolts.

Figure 5: Three-Dimensional Representation of Wet/Dry Sand. The x and y axes represent the distance in millimeters and the z-axis is the voltage in millivolts.

III. Imaging Approach

By taking two-dimensional images of a target zone, a large area can be spectrally mapped quickly given that the contrast between the background soil and burial sites are sufficiently strong to allow rapid detection. The idea is to increase contrast by collecting images at/over wavelengths where the differences in reflection or fluorescence properties are sufficiently large. To illustrate the application of imaging methods to the landmine detection problem, images of the fluorescence emission from a landmine casing partially buried in sand were captured. The 488-nm line from an Argon laser illuminated the object and a CCD camera captured the image through a 525-nm long pass filter. The orientation of the casing is shown in Figure 6. One of the captured images is shown in Figure 7. The high contrast between the casing and sand is readily apparent.



IV. Conclusion

In this paper, we have reported some preliminary work evaluating methods that could be used in detection of buried objects. We showed the signal of the NIR detector to be sensitive to water content of sand. We illustrated that such a device can produce images of regions having different moisture. This study illustrates the potential of an NIR imaging system to detect buried objects. Such a system could be composed of a charge-coupled device (CCD) in combination with a device to select a specific wavelength of light (e.g. an acoustic optical tuning filter).

V. Acknowledgements

This research is sponsored by the US Army MURI Project (US ARMY ARO/DAAG-55-97-1-0013) and by the Office of Biological and Environmental Research, U.S. Department of Energy under contract DE-AC05-96OR22464 with Lockheed Martin Energy Research Corporation.

VI. References

1. J. Mobley, G. H. Miller, P. M. Kasili, C. DiMarzio, and T. Vo-Dinh, Hyperspectral Imaging Using AOTF and NIR Sensing of Buried Objects and Landmines, *Env. Monit. and Remed. Technol.* 1999, Vol. 3534.

FDTD Wave Propagation in Dispersive Soil Using a Single Pole Conductivity Model

Carey M. Rappaport, Shuang Wu and Scott C. Winton
Center for Electromagnetics Research, Northeastern University, Boston, MA

Abstract—In FDTD modeling of lossy, dispersive soil for subsurface imaging and detection applications, the electric flux and the current are convolutions of $E(t)$ with $\epsilon(t)$ and $\sigma(t)$ respectively. To avoid these memory-intensive computations, the convolutions can often be accurately and simply modelled as second order difference equations. In particular, by matching the corresponding Z -transform of the E-field/current relation to frequency-dependent conductivity results in a ratio of polynomials in Z^{-1} (where $Z = e^{j\omega\Delta t}$). A good fit to measured soil data over two decades in frequency is possible using only a single pole, two zero conductivity model. Compared to a similarly accurate three-term Debye model, this one-pole model requires one-third the storage of previously computed field values.

Index terms—FDTD, Soil Modeling, Dispersion, Mine Detection

I. BACKGROUND

There is growing interest in simulating wave propagation for underground microwave and RF sensing and imaging applications. Soil is a good candidate for the finite difference methods since it is usually inhomogeneous, lossy, and has an irregular surface boundary. It has been challenging to accurately compute wideband wave behavior in realistically modelled soil because of its dispersive nature, requiring either many individual frequency domain calculations or a robust deconvolution of $E(t)$ from $D(t)$ in the time domain. A simple dispersive variant of the FDTD algorithm which includes the effects of frequency-dependent conductivity and dielectric constant enables the use of this prevalent and efficient modeling method.

The conventional approach to modeling dispersion in soil approximates the frequency domain dispersive complex dielectric constant with rational functions (Debye or Lorentz models) of $j\omega$ [1,2], multiplies the constitutive relation by the denominator and inverse Fourier transforms the result. We improve on this method by modeling solely the conductivity as a simple rational function of the Z -transform [3,4], based on the observation that the frequency variation of the real dielectric constant does not significantly affect either the real propagation constant β or the decay rate α . By modeling σ in terms of powers of the Z -transform variable Z^{-1} (which readily transforms to time delays), the conversion of the generalized dispersive Ohm's Law $\mathbf{J}(Z) = \sigma(Z)\mathbf{E}(Z)$ to the time domain is particularly straightforward.

Manuscript received June 1, 1998.

This work has been supported by The Army Research Office, Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013

II. MODEL FORMULATION

To preserve the conceptual elegance and simplicity of the FDTD method, it is important to keep the media model to at worst second-order. In previous reports, a quite satisfactory two-pole, two-zero rational function conductivity model was presented. This current work shows that a model with a single pole can often be just as good. The right hand side of Ampere's Law $\nabla \times \mathbf{H} = \partial \mathbf{D} / \partial t + \mathbf{J}$ Z -transforms to: $[(1 - Z^{-1}) / \Delta t] \epsilon_0 \epsilon_{Av} \mathbf{E} + \sigma(Z) \mathbf{E}$ (where it is assumed that ϵ_{Av} is frequency independent). The time-dependent conductivity may thus be modelled as

$$\sigma(Z) = \frac{J(Z)}{E(Z)} = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2}}{1 + a_1 Z^{-1}} \quad (1)$$

Both the real and the imaginary components of $\sigma(Z)$ depend on the frequency sampling interval Δt and the coefficients of the rational function (a_1 , b_0 , b_1 , and b_2). The imaginary component corresponds to an effective permittivity. This model is implemented by fitting $\text{Re}\{\sigma\}$ to measured conductivity and $\text{Im}\{\sigma\} / \omega \epsilon_0$ to measured real dielectric constant less an average value ϵ_{Av} . To simultaneously solve for b_0 , b_1 , and b_2 , an initial guess is made for a_1 . The conductivity and dielectric constant at three representative frequencies for the measured data and the model are equated. Further simple optimization is performed by trial and error, varying either a_1 or one of the three frequencies. Complex wave number $k(\omega)$ is then calculated from both the model and the measured data, and the real propagation constant β and the decay rate α are derived from $k = \omega / c \sqrt{\epsilon_{Av} - j\sigma / \omega \epsilon_0}$.

Once the a and b parameters of σ have been determined, integration into the FDTD algorithm is straightforward. If the temporal average of the current density \mathbf{J} is used, Ampere's law in the Z domain becomes:

$$\nabla \times \mathbf{H}(Z) = \epsilon_{Av} \frac{1 - Z^{-1}}{\Delta t} \mathbf{E}(Z) + \frac{1 + Z^{-1}}{2} \sigma(Z) \mathbf{E}(Z) \quad (2)$$

Using (1) in (2) and noting that Z^{-1} corresponds to the time index shift $n \rightarrow n - 1$, the E-field update equation for a 2-D Transverse Magnetic ($\mathbf{E} = \hat{z} E_z$) simulation is:

$$E_z^{n+1} = \frac{1}{\epsilon_0} (-e_1 E_z^n - e_2 E_z^{n-1} - e_3 E_z^{n-2} + \Delta H) \\ \Delta H = \frac{\Delta t}{\epsilon_{Av}} \left(-I_x^{n+\frac{1}{2}} - a_1 I_x^{n-\frac{1}{2}} + I_y^{n+\frac{1}{2}} + a_1 I_y^{n-\frac{1}{2}} \right) \quad (3)$$

where $e_0 = 1 + b_0 \Delta t / 2 \epsilon_{Av}$, $e_1 = a_1 - 1 + (b_0 + b_1) \Delta t / 2 \epsilon_{Av}$, $e_2 = -a_1 + (b_1 + b_2) \Delta t / 2 \epsilon_{Av}$, $e_3 = b_2 \Delta t / 2 \epsilon_{Av}$ and I_x and I_y are the spatial first differences of the H_x and H_y fields with respect to y and x (spatial indices have been suppressed for clarity).

The one-pole model of $\sigma(Z)$ requires storing I_x and I_y (and I_z in a three-dimensional simulation) a single additional time compared with non-dispersive FDTD. For each additional conductivity pole, this same amount of storage space must be additionally allocated. Fig. 1 shows the fit of the model (at $\Delta t = 20$ ps) to the conductivity and dielectric constant data measured from Puerto Rican clay loam (PRCL) with 1.4 g/cm^3 density and 10% moisture [5]. The difference between the imaginary conductivity of the model and the measured ϵ (lower panel) is used to determine the constant real soil permittivity ϵ_{Av} . Fig. 2 shows the fit of the full model to β and α . The maximum amount of error is less than 20%, which is close to the measurement error of the soil sample.

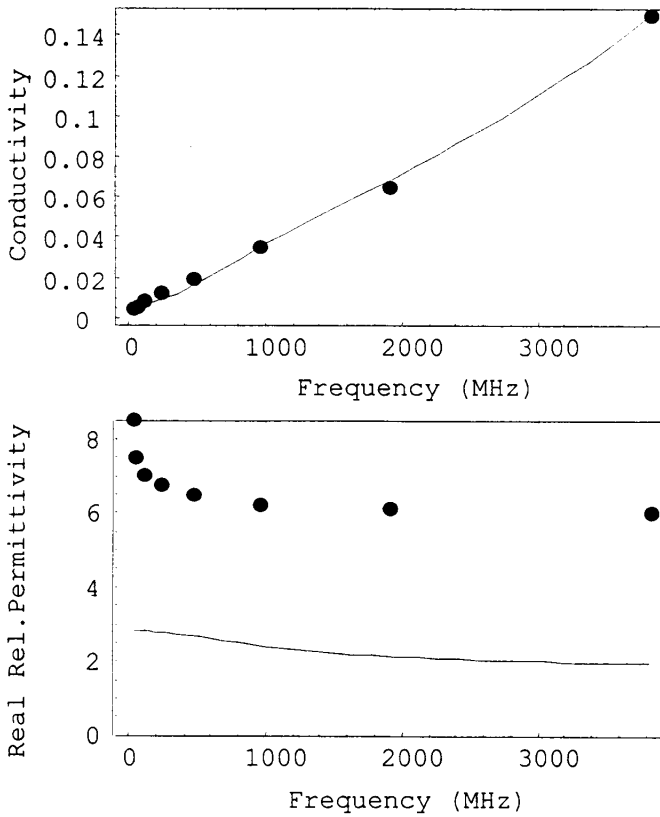


Fig. 1. (a) Measured (circles) versus single pole model (continuous curves) of $\sigma(Z)$: (a) real conductivity and (b) imaginary conductivity (real relative dielectric constant) for Puerto Rican clay loam, 10% moisture, 1.4 g/cm^3 density [5]. The constant real dielectric constant ϵ_{Av} is the average difference between the curve and the measured circles of (b).

The use of the average value of the current density is well established and produces accurate results. However, using only the present value of the current density in the update equation results in the savings of another storage location. To facilitate this savings without sacrificing accuracy, $\sigma(Z)$ is multiplied by $(1 + Z^{-1})/2$ and a new set of model parameters is determined to fit this

product over the frequencies of interest. In this case, (3) is additionally simplified with $e_0 = 1 + b_0 \Delta t / \epsilon_{Av}$, $e_1 = a_1 - 1 + b_1 \Delta t / \epsilon_{Av}$, $e_2 = -a_1 + b_1 \Delta t / \epsilon_{Av}$, and $e_3 = 0$, and this dispersive model requires only one additional storing of each field value compared to non-dispersive FDTD. This compares with two field storings for the (2,2) model and three field storings for the comparably accurate three-term Debye model.

Similar modeling for several different types of soil with widely varying density and moisture [5,6] at $\Delta t = 20$ ps is summarized in Table 1. All models have wave numbers in agreement with measured values within 20%.

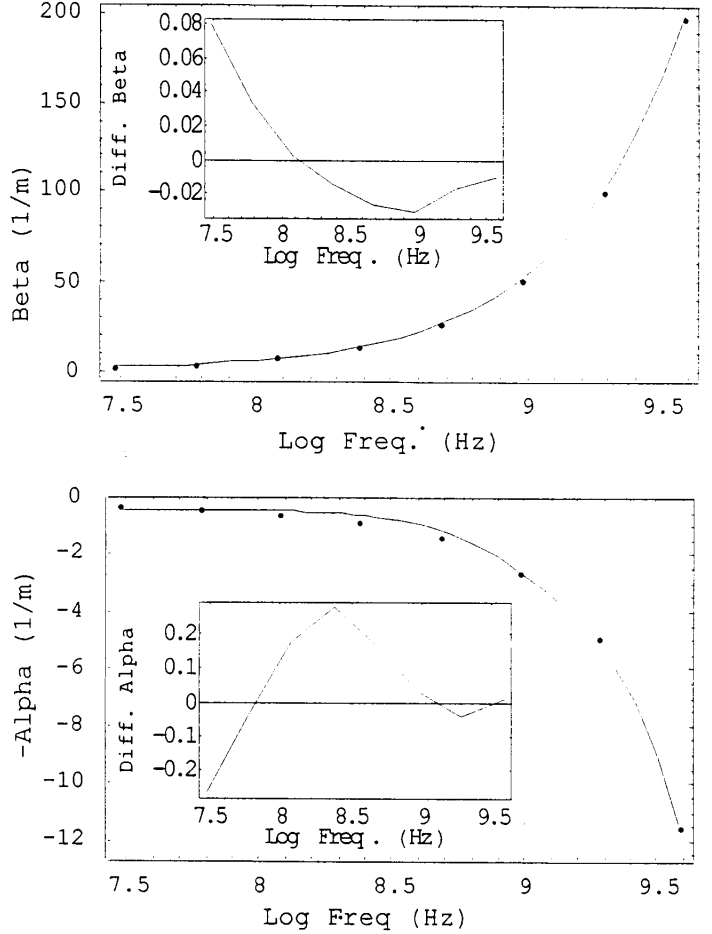


Fig. 2. Measured (circles) versus model for single pole model of $\sigma(Z)$ with constant ϵ_{Av} (continuous curves): (a) Real and (b) imaginary parts of wave number for Puerto Rican clay loam, 10% moisture, 1.4 g/cm^3 density [5] as a function of frequency. Insets show normalized error for this model.

III. STABILITY CONSIDERATIONS

The Courant stability analysis in [3] shows that in a source-free homogeneous medium, the time-space transform of the wave equation for a plane wave ($\Delta x = \Delta y = \Delta z = \Delta$) is:

$$(Z - 1)F(Z) = -4r^2 s^2 \quad (4)$$

where from (2), $F(Z) = \epsilon_{Av}(1 - Z^{-1}) + (\Delta t/2\epsilon_0)(1 + Z^{-1})\sigma(Z)$, and $r = c\Delta t/\sqrt{\epsilon_{Av}}\Delta$, and $s^2 = \sin^2 k_x\Delta + \sin^2 k_y\Delta + \sin^2 k_z\Delta$ with maximum value of s equal to the spatial dimension of the FDTD simulation. Expressing $F(Z)$ in rational form:

$$F(Z) = \epsilon_{Av} \frac{(e_0 + e_1 Z^{-1} + e_2 Z^{-2})}{1 + a_1 Z^{-1}} \quad (5)$$

Using (4), (3) becomes a third order polynomial in Z^{-1} . The solution to this equation yields one real root and two complex roots that are a function of Δ . Stability requires $|Z| < 1$. For Δ within the range of 4.6mm to 120mm, all three roots are within the unit circle.

NUMERICAL TEST CASES

Several numerical experiments were performed to validate this method. The simulations were run in one dimension for simplicity. The choice of Δt is critical since it defines the model, determines the range of usable frequencies and determines the limits on Δ through the Courant condition. For these experiments, the spatial increment was 2.5 mm. The incident field is a modulated gaussian pulse, used to excite many frequencies simultaneously. The modulation frequency is 3 GHz for the Alicia and A.P. Hill soils and 1.5 GHz for the PRCL soil sample.

In order to compare the results of the FDTD simulations with the measured data, the frequency-dependent wave number k must be extracted. Fast Fourier Transforming the time domain fields gives

$$k_{\text{FDTD}} = -\frac{1}{j\ell} \ln \frac{E_z(\omega, \ell)}{E_z(\omega, 0)} \quad (6)$$

Fig. 3 shows the real and imaginary parts of the wave number versus frequency for measured values and either the "Time-Averaged" or "Model-Averaged" k_{FDTD} . These latter "Model-Averaged" simulations use only the current value of the conduction current J^n . The parameters in Fig. 3 correspond to the three separate sites of Table 1. Both the real and imaginary wave numbers agree well for each soil, for both types of models across the entire two decades of bandwidth from 45 MHz to almost 4 GHz. It should be noted that the "Time-Averaged" FDTD computation is more robust, with less sensitivity to excitation function and Courant number.

CONCLUSIONS

A minimal-storage time-domain model for frequency-dispersive soil based on a constant real dielectric constant and a (2,1) rational function of Z^{-1} conductivity function has been developed and tested. By pre-multiplying the measured conductivity data by the Z-transform of the time average function, the resulting model avoids requiring conduction current time averaging in Ampere's law. Without this time averaging,

TABLE I
ONE-POLE CONDUCTIVITY PARAMETERS

Puerto Rican Clay Loam: $m = 10$, $d = 1.4^\dagger$					
av?	a_1	b_0	b_1	b_2	ϵ_{Av}
no	-0.88	0.91625	-1.67662	0.761072	4.18775
yes	-0.9	0.7983	-1.4695	0.67176	4.282
A.P. Hill, Firing Point 22: $m = 19.3$, $d = 1.527$					
no	-0.975	1.51947	-2.97284	1.45362	5.25834
yes	-0.97868	1.5473	-3.036	1.4888	4.7731
Bosnia, Test site Alicia: $m = 25.3$, $d = 1.263$					
no	-0.925	1.76106	-3.32102	1.56193	5.03815
yes	-0.93	1.6325	-3.0827	1.4521	4.9831

$^\dagger m$ is percent moisture content of the soil, d is soil density (g/cm³).

a one-pole conductivity model requires just one additional store of electric and magnetic field compared to the non-dispersive FDTD algorithm. This is half the required additional storage of the previously reported (2,2) model, and even less than standard Debye and Lorentz models. In addition, fitting parameters to a single, real conductivity function — rather than to both parts of a complex permittivity function — is quite advantageous.

Numerical simulations on measured data show that this simple model is efficient and accurate across a wide frequency band for both real and imaginary parts of wave number, giving good predictions of velocity and decay rate. A stability analysis shows that the model is stable for one, two, and three dimensions, for a wide range of grid spacings.

REFERENCES

- [1] T. Kashiwa and I. Fukai, "A Treatment by the FD-TD Method for the Dispersive Characteristics Associated with Electronic Polarization", *Microwave and Guided Wave Letters*, vol. 16, pp. 203-205, June 1990.
- [2] O. Gandhi, "A Frequency-Dependent Finite Difference Time Domain Formulation for General Dispersive Media", *IEEE Transactions on Microwave Theory and Techniques*, vol. 41, pp. 658-665, April 1993.
- [3] W. Weedon, and C. Rappaport, "A General Method for FDTD Modeling of Wave Propagation in Arbitrary Frequency-Dispersive Media", *IEEE Transactions on Antennas and Propagation*, pp. 401-410, March 1997.
- [4] C. Rappaport and S. Winton, "Modeling Dispersive Soil for FDTD Computation By Fitting Conductivity Parameters", *12th Annual Review of Progress in Applied Computational Electromagnetics Symposium Digest*, pp. 112-118, March 1997.
- [5] J. Hipp, "Soil Electromagnetic Parameters as Functions of Frequency, Soil Density, and Soil Moisture", *Proceedings of the IEEE*, vol. 62, pp. 98-103, January 1974.
- [6] J. Curtis, "Dielectric Properties of Soils: Various Sites in Bosnia", *U.S. Army Corps of Engineers, Waterways Experimental Station Data Report*, 20 August 1996.

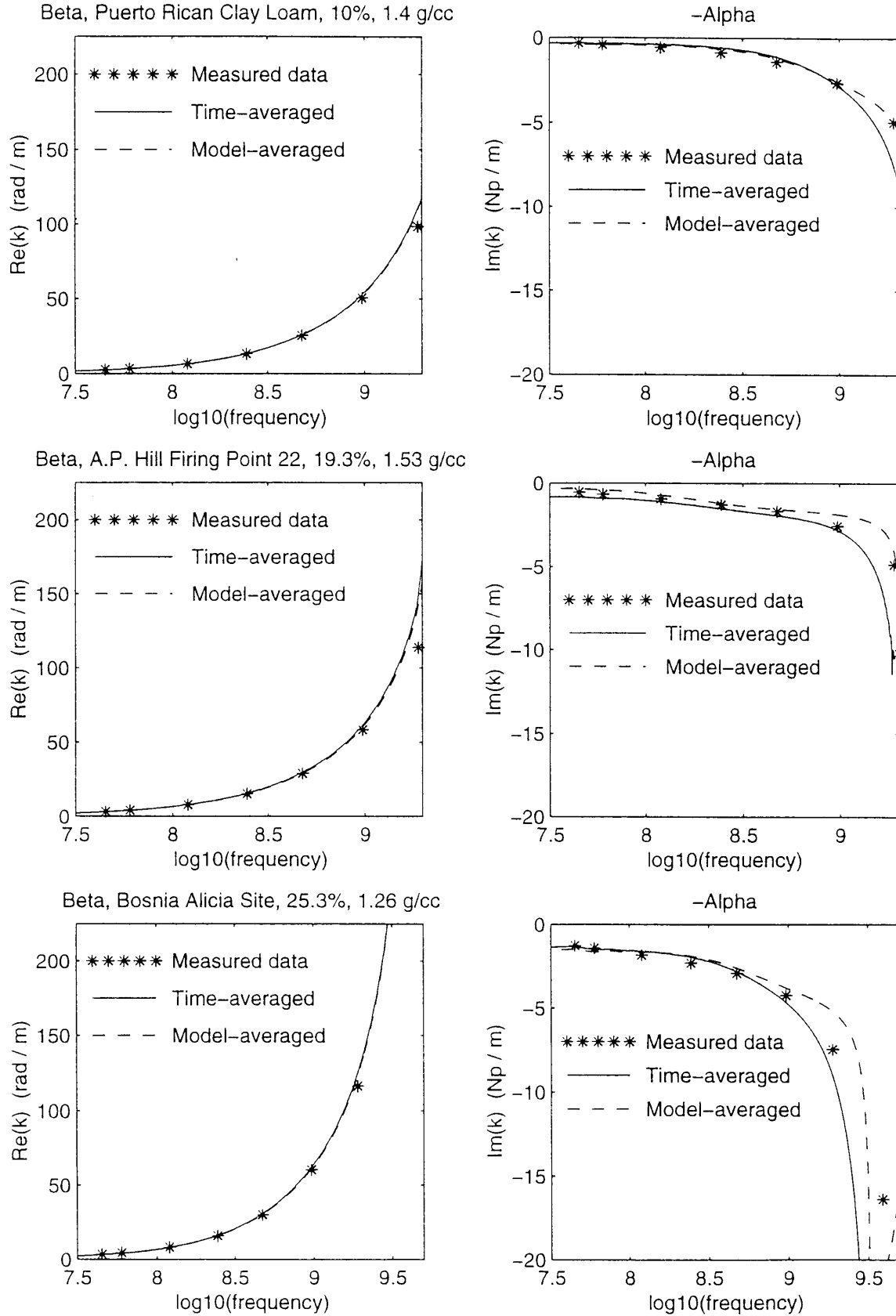


Fig. 3. Measured (*) versus FDTD computations of real and imaginary wave number using "Time-averaged" (solid) and "Model-Averaged" (dashed) models for single pole model of $\sigma(Z)$ with constant ϵ_{Av} for three soils with varying moisture and density.

Modeling GPR Signal Degradation from Random Rough Ground Surface



Carey Rappaport, Magda El-Shenawee, and He Zhan
Center for Subsurface Sensing and Imaging Systems
Northeastern University
Boston, MA 02115

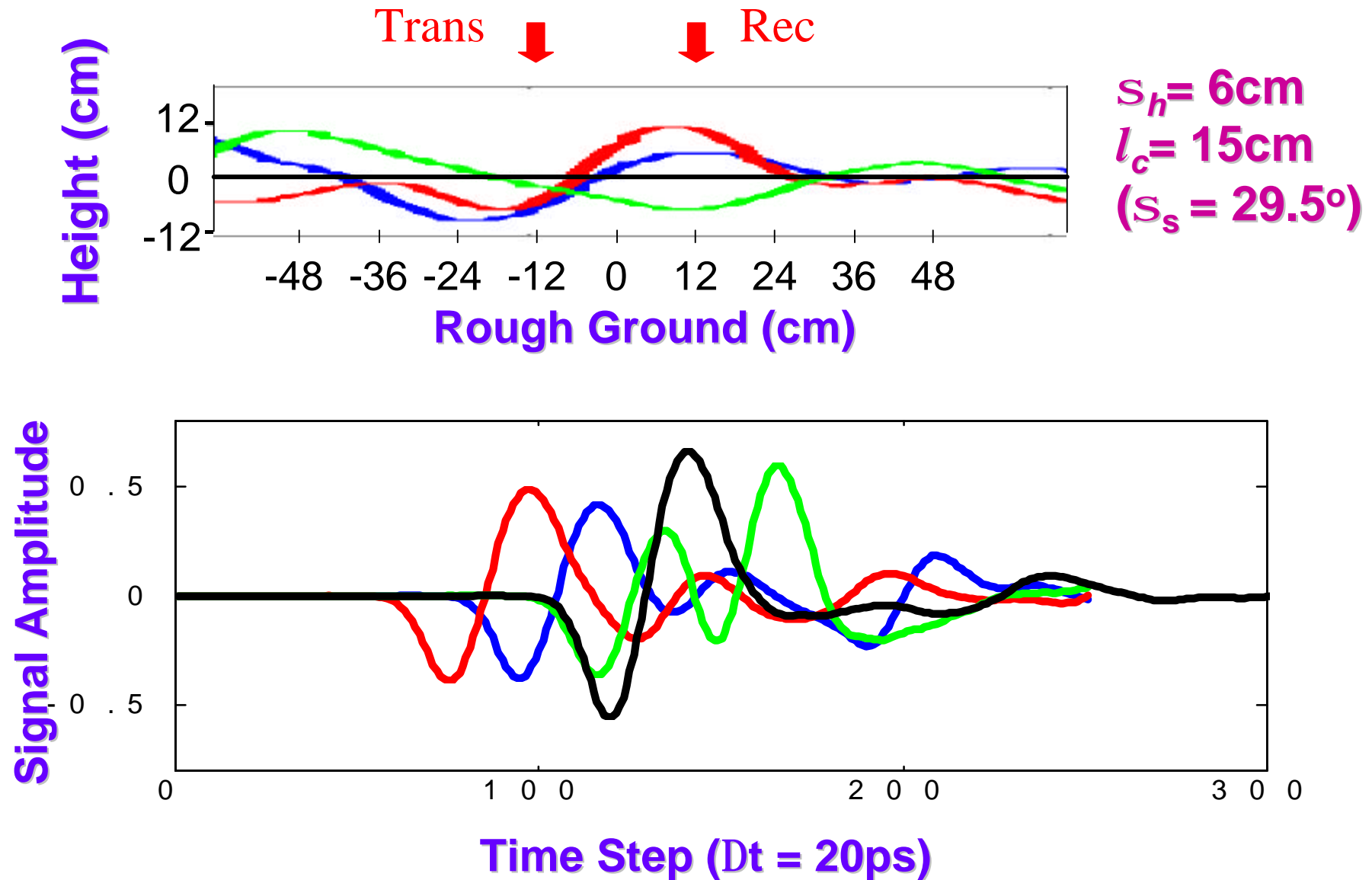
IGARSS, Honolulu Hi, July 2000

Outline

- Problem: Why is it important to characterize the nearfield rough surface effects on impulse GPR signals?
- Modeling methodologies
- Features of ground surface clutter signal
- Separating surface clutter from target signal.

* Demining MURI Supported by ARO

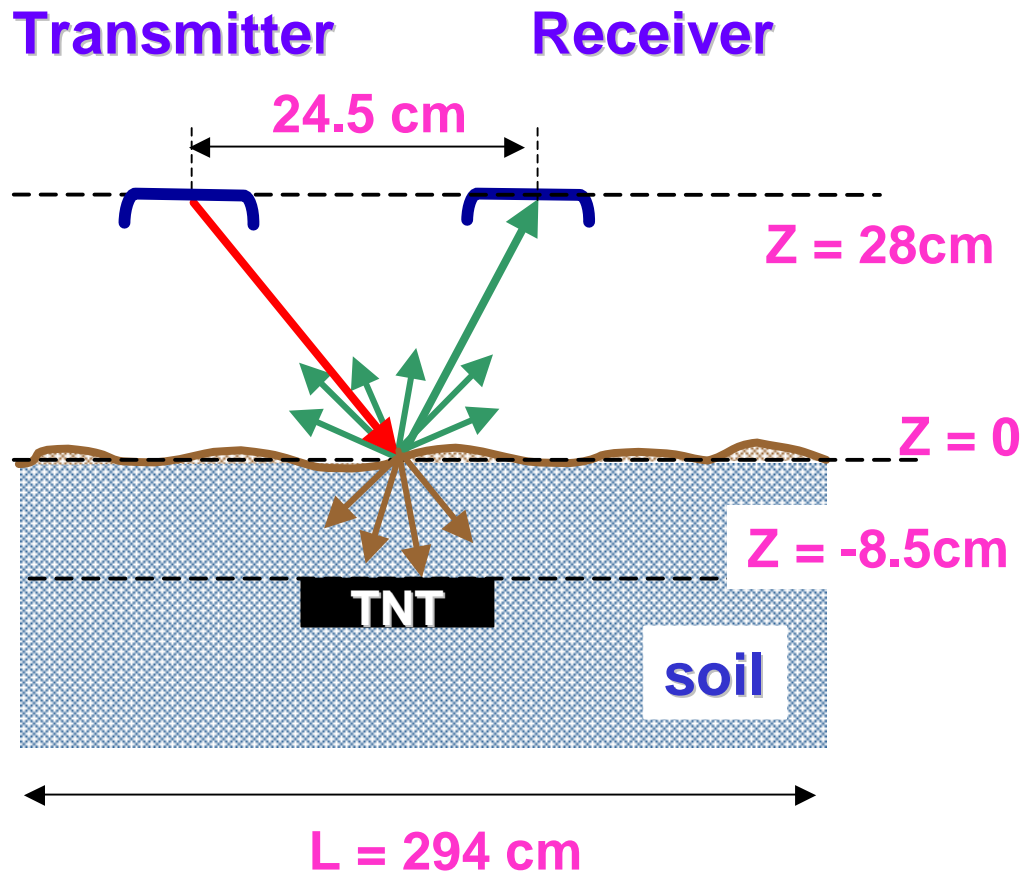
Effect of Ground Roughness on Ground Scattered Pulse



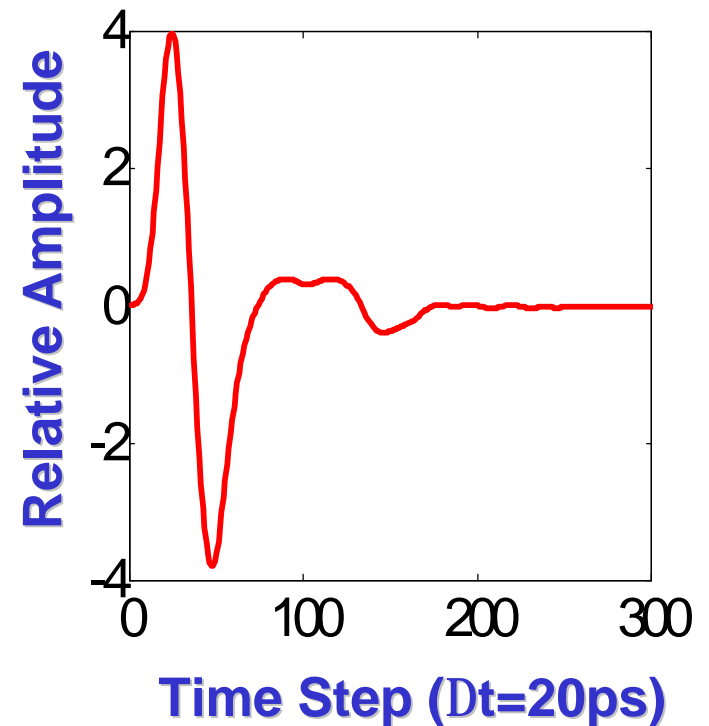
Rough Ground Clutter Signal Characterization

- Signals from rough ground vary considerably
 - Pulse shape depends on roughness and T/R position [DISTORTION]
 - Peak depends on particular T/R position [SHIFT]
 - Overall amplitude varies [SCALING]
- Monte Carlo simulation can model these relevant features
 - 2D FDTD model
 - Real measured impulse GPR excitation and dispersive soil
 - 500 different rough surface realizations

Impulse Ground Penetrating Radar Specifications



Simulation Space



Measured Incident Pulse

Rough Surface Statistical Specification

Gaussian PDF for surface height z with zero mean and rms height = s_h :

$$p(z) = \frac{1}{s_h \sqrt{2\pi}} \exp\left(-z^2 / 2s_h^2\right)$$

Surface height auto-correlation function for surface points x_1, x_2 :

$$R(x_d) = \frac{\langle f(x_1)f(x_1 + x_d) \rangle}{s_h^2}, \quad x_d = x_2 - x_1$$

$$R(x_d) = \frac{1}{s_h^2} \int_{-\infty}^{\infty} W(K_x) \exp(iK_x x_d) dK_x$$

W is the surface profile spectral density function

Rough Surface Statistical Specification (con'd)

For Gaussian spectrum W:

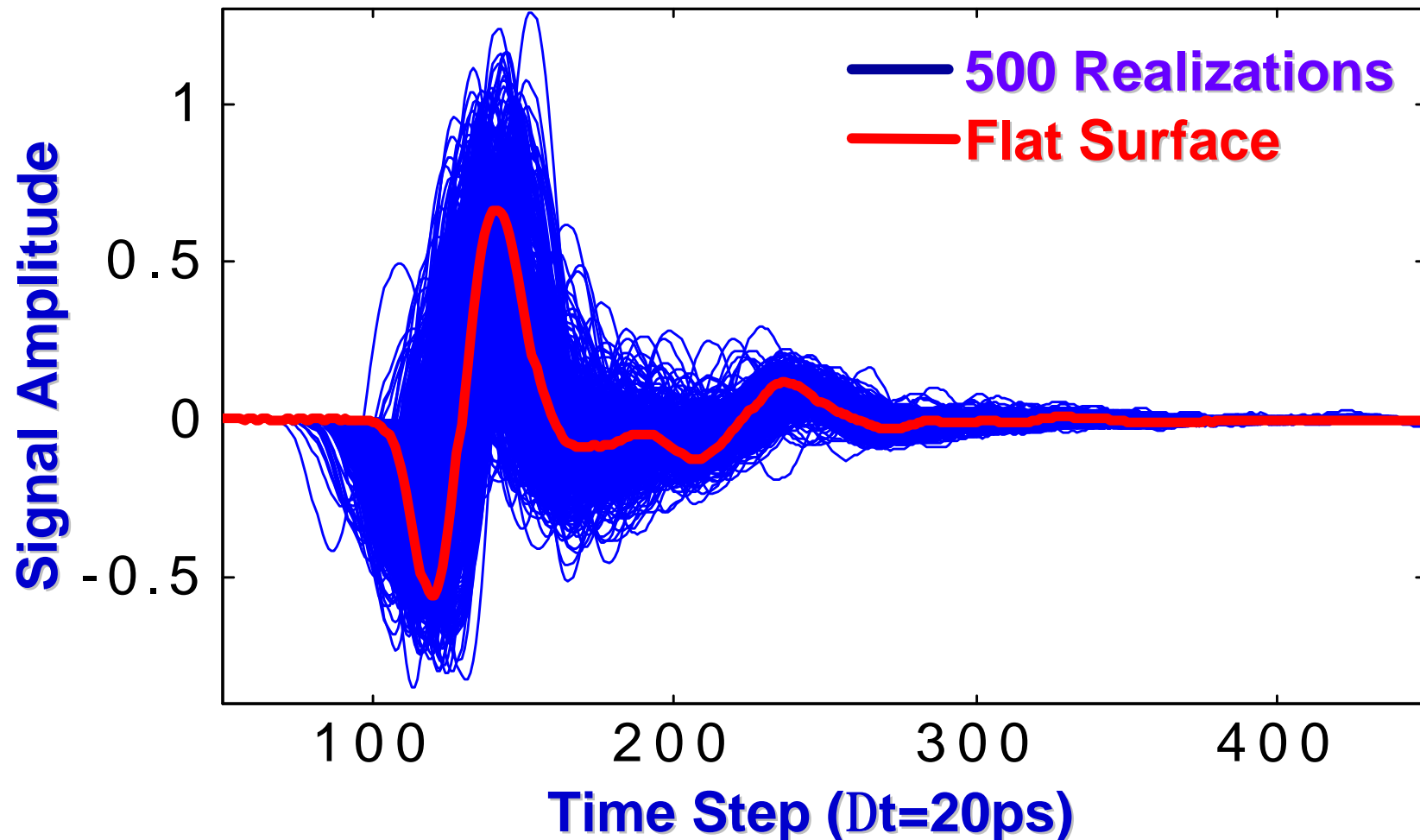
$$R(x_d) = \exp\left(-\frac{x_d^2}{l_c^2}\right)$$

Gaussian joint PDF for surface heights (z_1, z_2) with zero mean and rms height = s_h :

$$p(z_1, z_2) = \frac{1}{2ps_h^2 \sqrt{1 - R^2(x_d)}} \exp\left(-\frac{z_1^2 - 2R(x_d)z_1z_2 + z_2^2}{2s_h^2(1 - R^2(x_d))}\right)$$

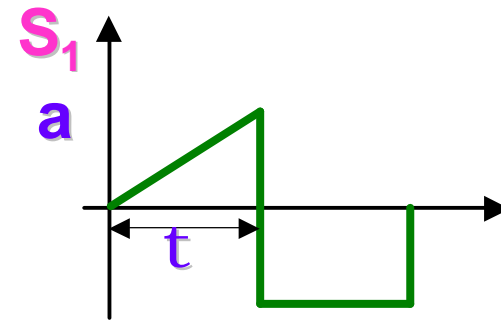
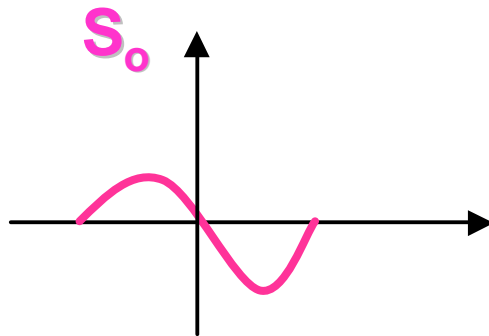
Raw Computed Signals

Rough Surface Parameters: $s_h = 3\text{cm}$, $l_c = 10\text{cm}$, ($s_s = 23^\circ$)

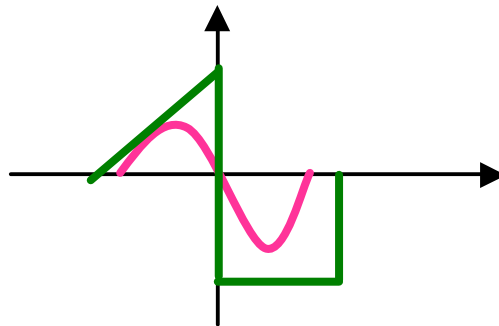


Alignment of Randomly Delayed Signals

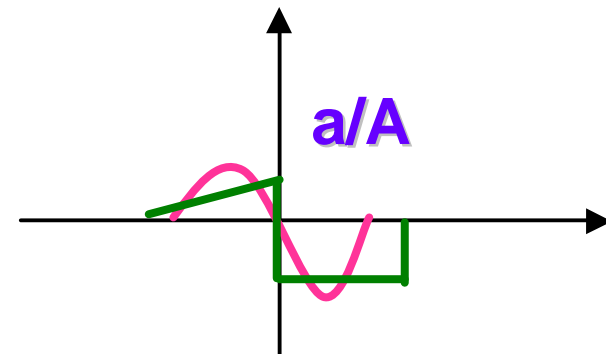
Given two signals S_0 and S_1



1) Shift by t :



2) Scale by A :



Cross-Correlation Signal Alignment

Normalized cross-correlation functions between flat 'f' and rough 'i' surface:

$$C_{fi}(m) = \sum_{n=1}^{N-|m|} S_f(n) S_i(n+m) / \sqrt{\sum_{n=1}^N |S_f(n)|^2} \sqrt{\sum_{n=1}^N |S_i(n)|^2} \quad \text{For } m > 0$$

$$C_{fi}(m) = C_{if}(-m) \quad \text{For } m < 0$$

Relative Scaling:

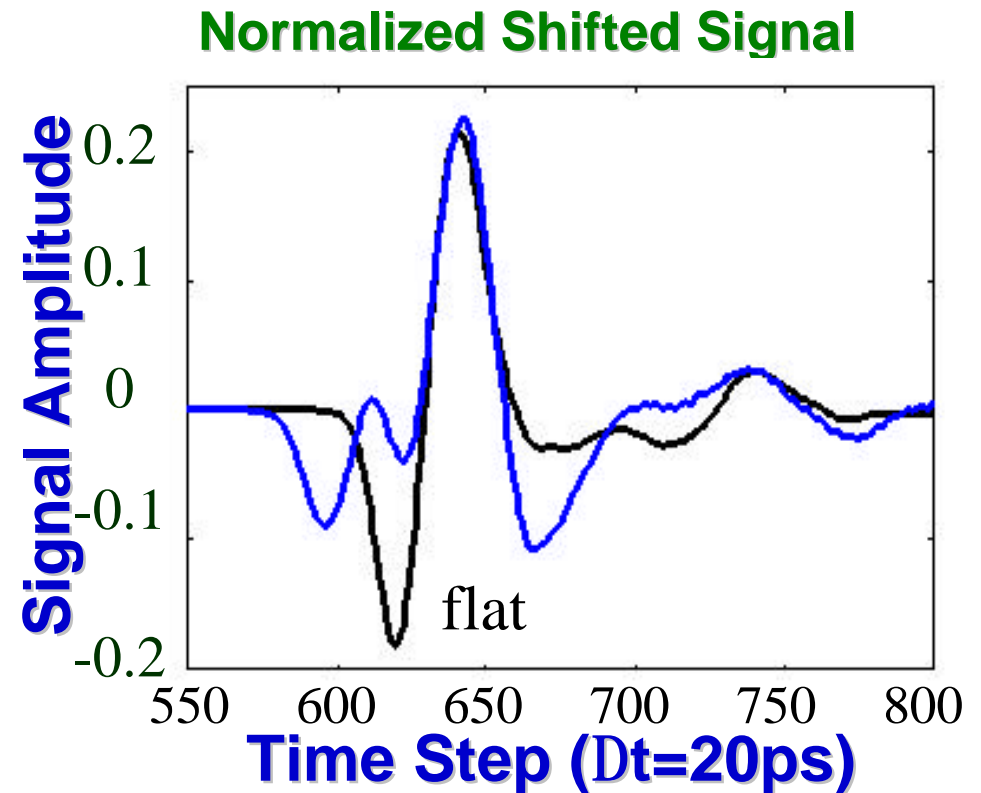
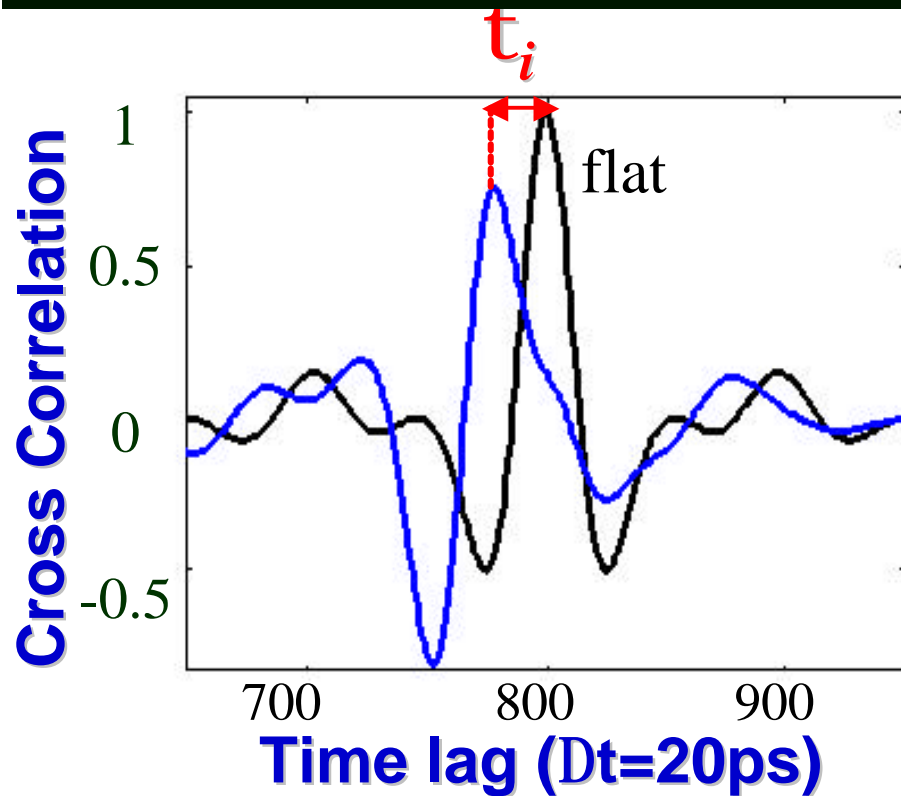
$$A_i = \sqrt{\sum_{n=1}^N |S_i(n)|^2} / \sqrt{\sum_{n=1}^N |S_f(n)|^2}$$

where $i = 1, 2 \dots M$.

M = number of surface realizations.

N = number of time steps in each signal.

Alignment Example



For realization # i (e.g.# 378):

Shift= $t_i = 22$ Dt, Scaling factor= $A_i = 0.88$

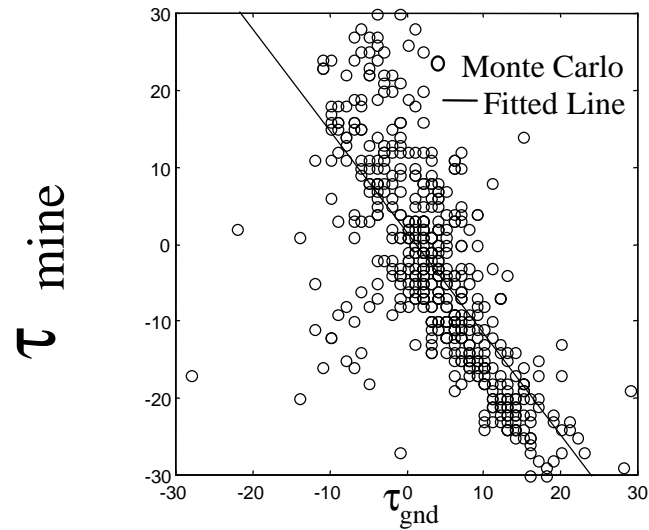
For $i=1,2,\dots,500$: Signal Statistics: 1- Average Shift= \bar{t}

2- Average Scaling= A

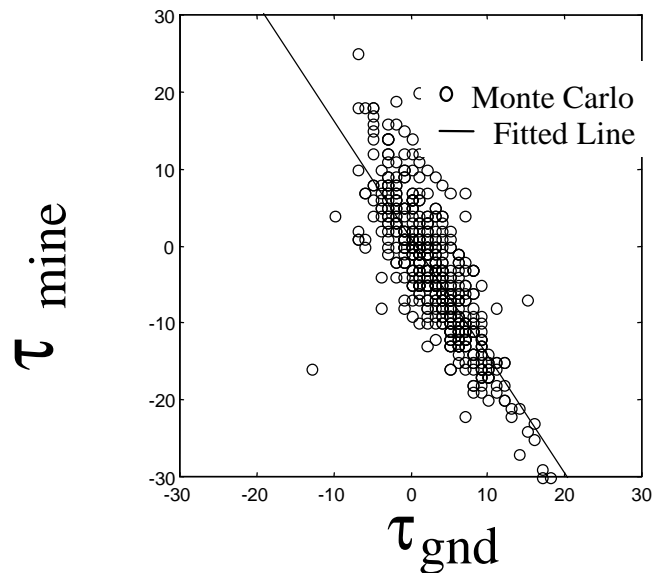
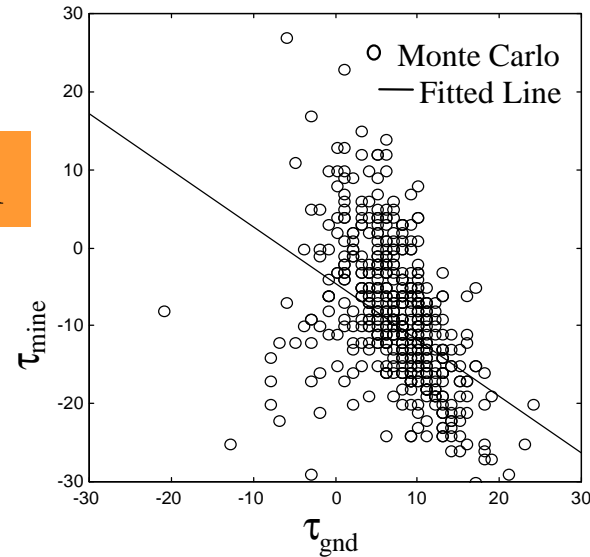
Determining Differential Target Delay from Detected Ground Clutter Delay

- If ground surface height varies, amount of soil between air and nominal target depth varies
- For a given target depth, longer path in air implies shorter path in soil; earlier target signal arrival.
 - Earlier arrival proportional to differential soil propagation velocity
- Verify this effect by comparing ground-only signals with ground-and-target signals:
 - Align ground clutter signals to find τ_{gnd}
 - Align ground-and-target signals each with ground-only signal subtracted w.r.t. target peak to find τ_{mine}

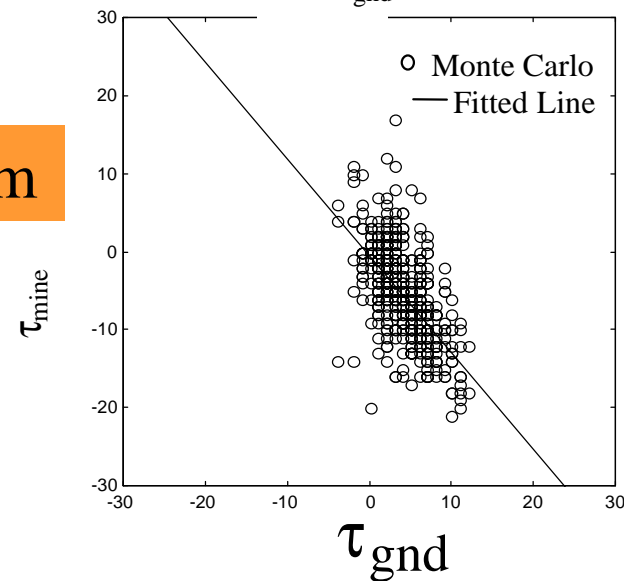
Correlation of Ground Surface Delay and Target Delay



$\sigma_h = 3 \text{ cm}$



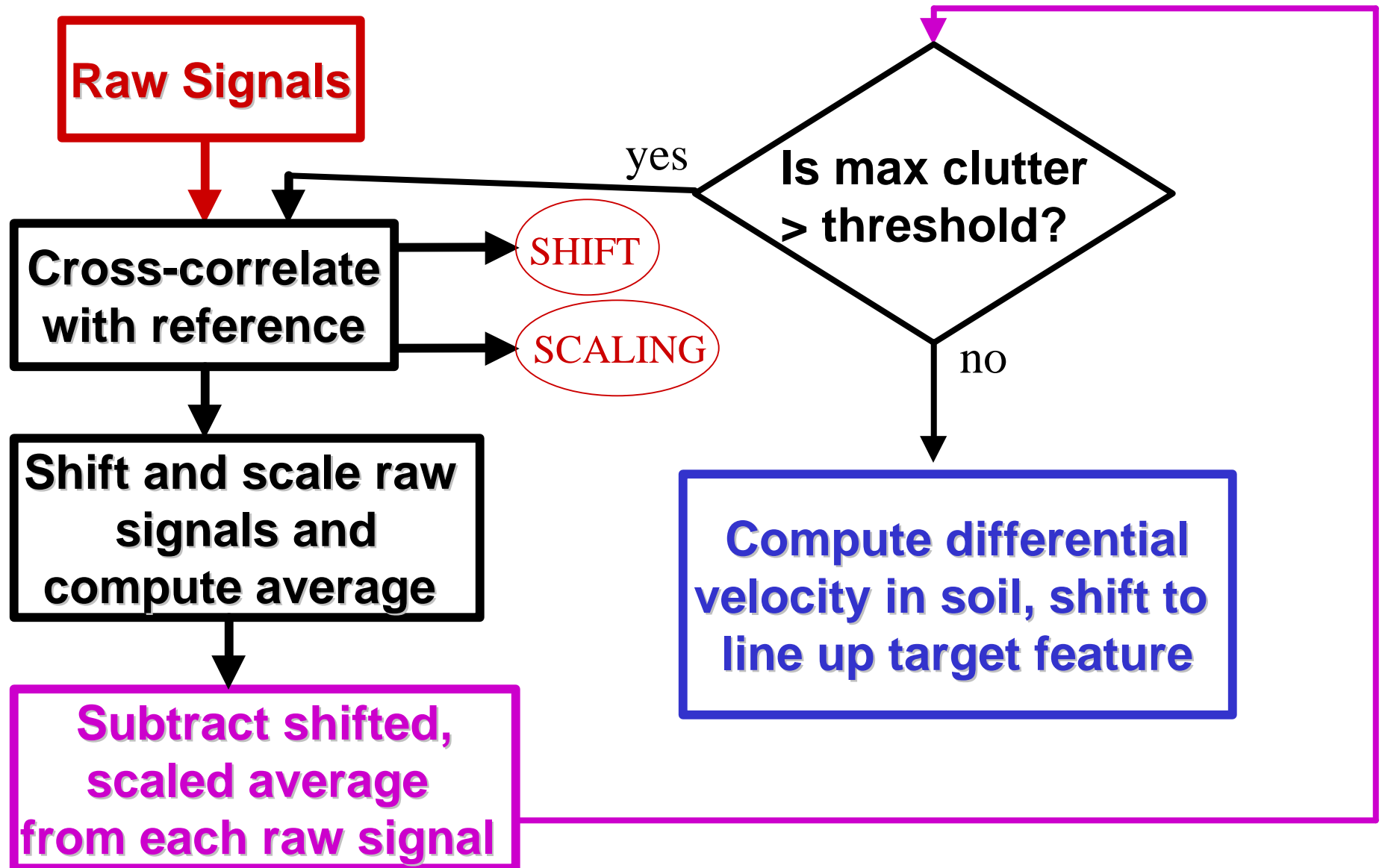
$\sigma_h = 2 \text{ cm}$



$L_c = 10 \text{ cm}$

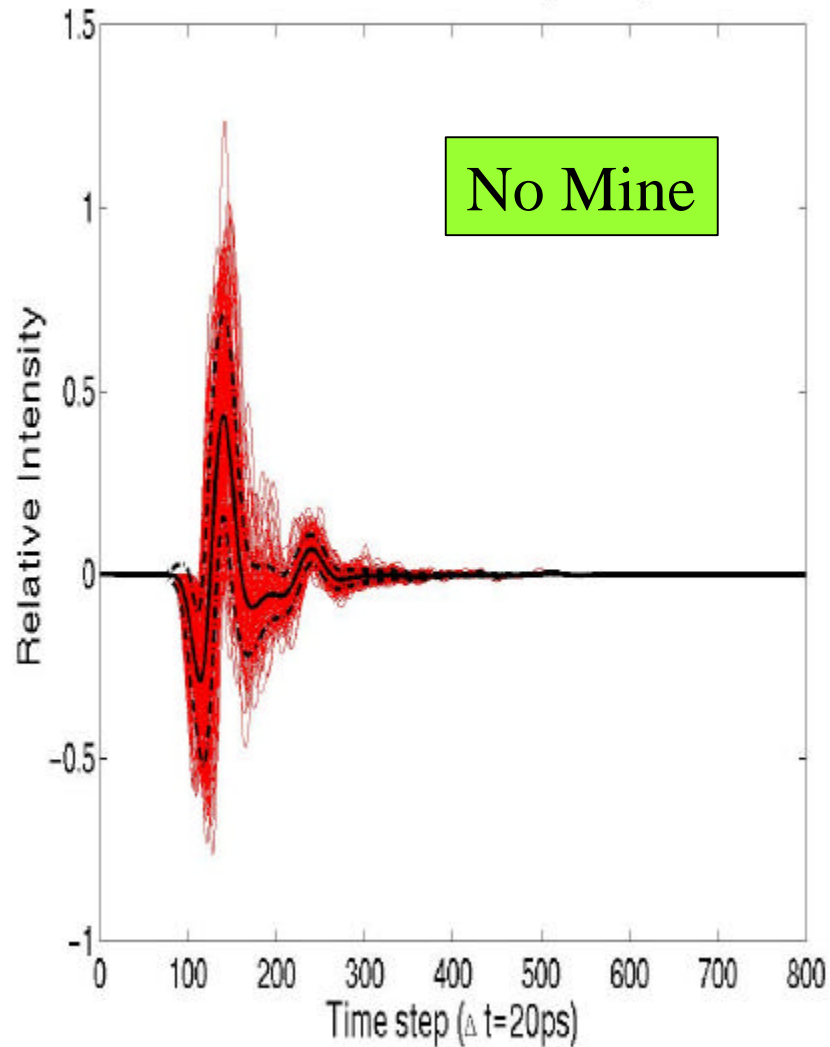
$L_c = 3 \text{ cm}$

Separating Mine Response from Ground Clutter

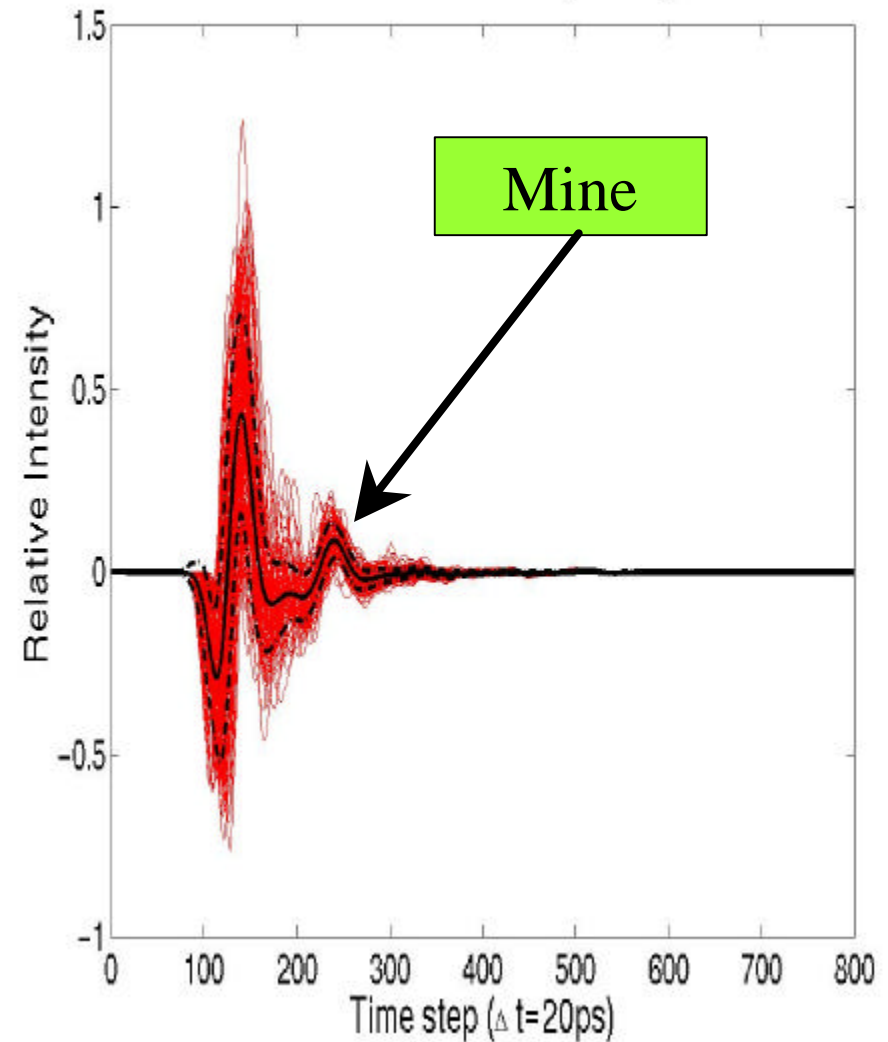


Original Signal Averages Obscure Mine Signal

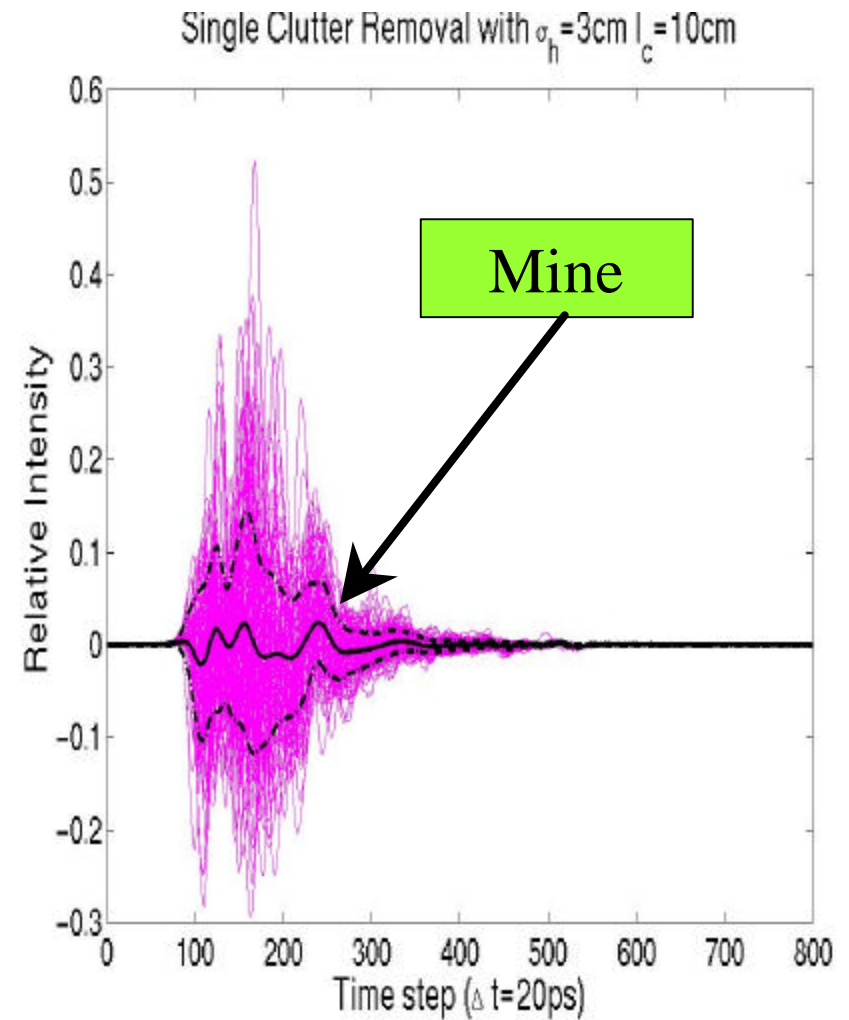
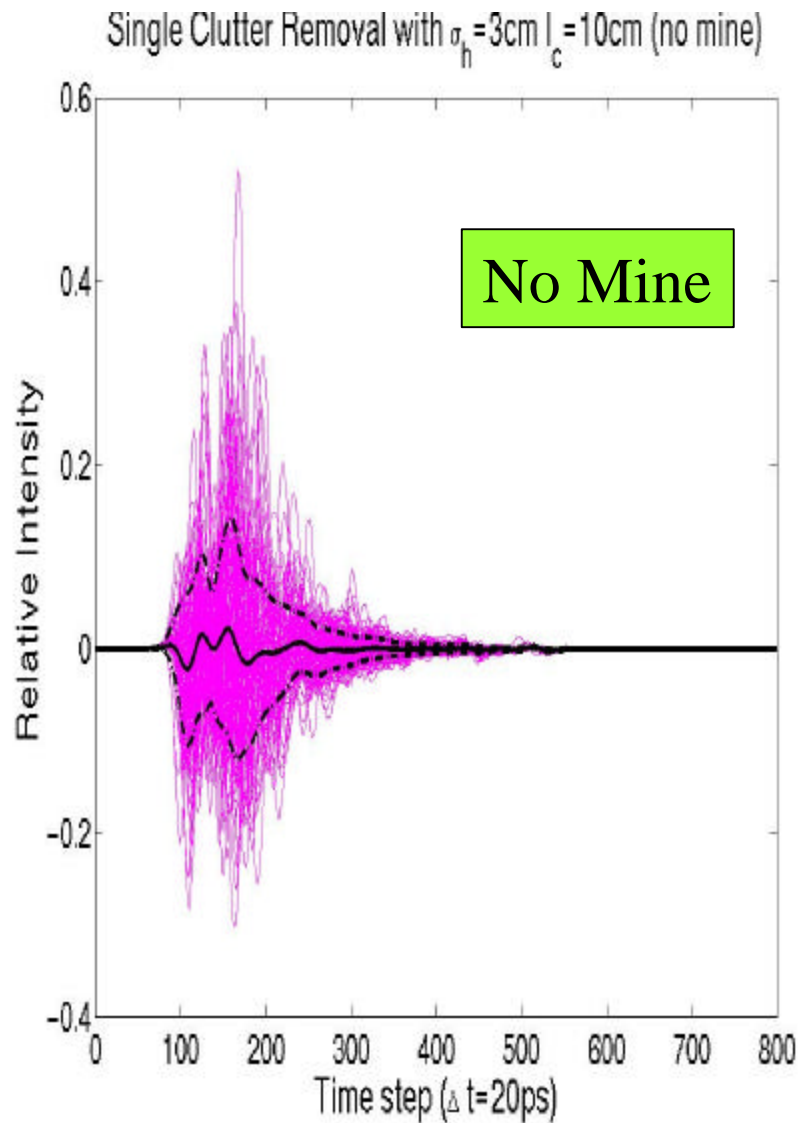
Original Nomine Signals with $\sigma_h = 3\text{cm}$ $l_c = 10\text{cm}$



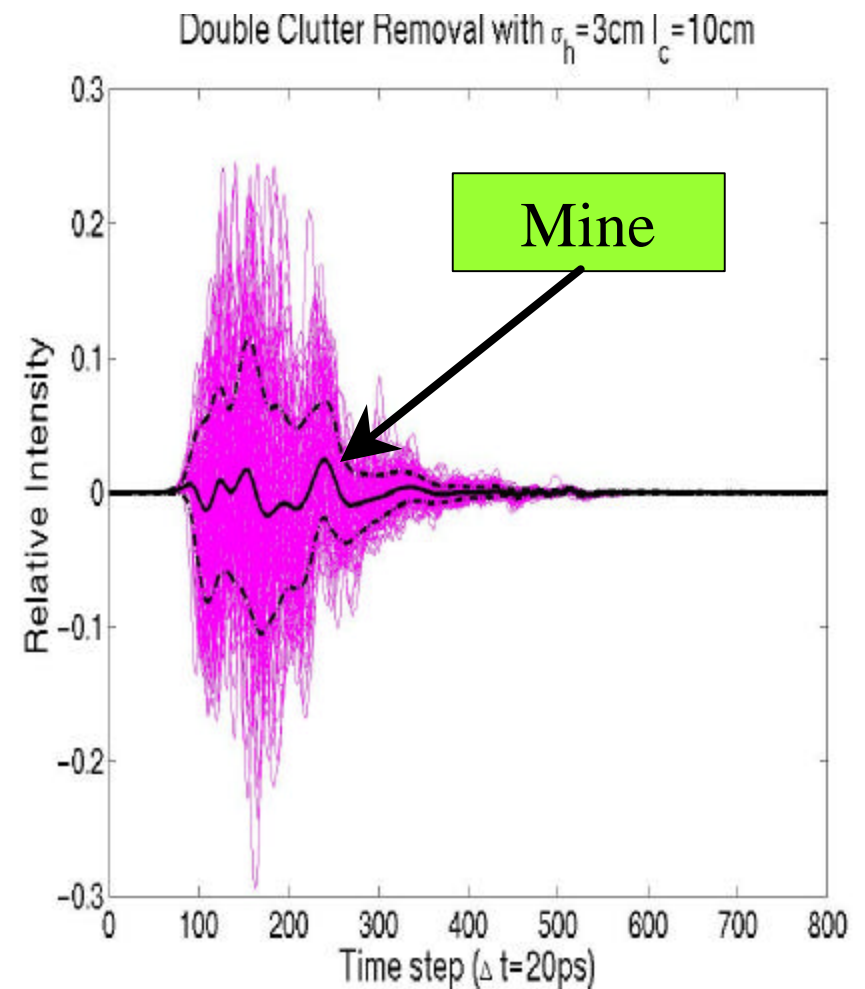
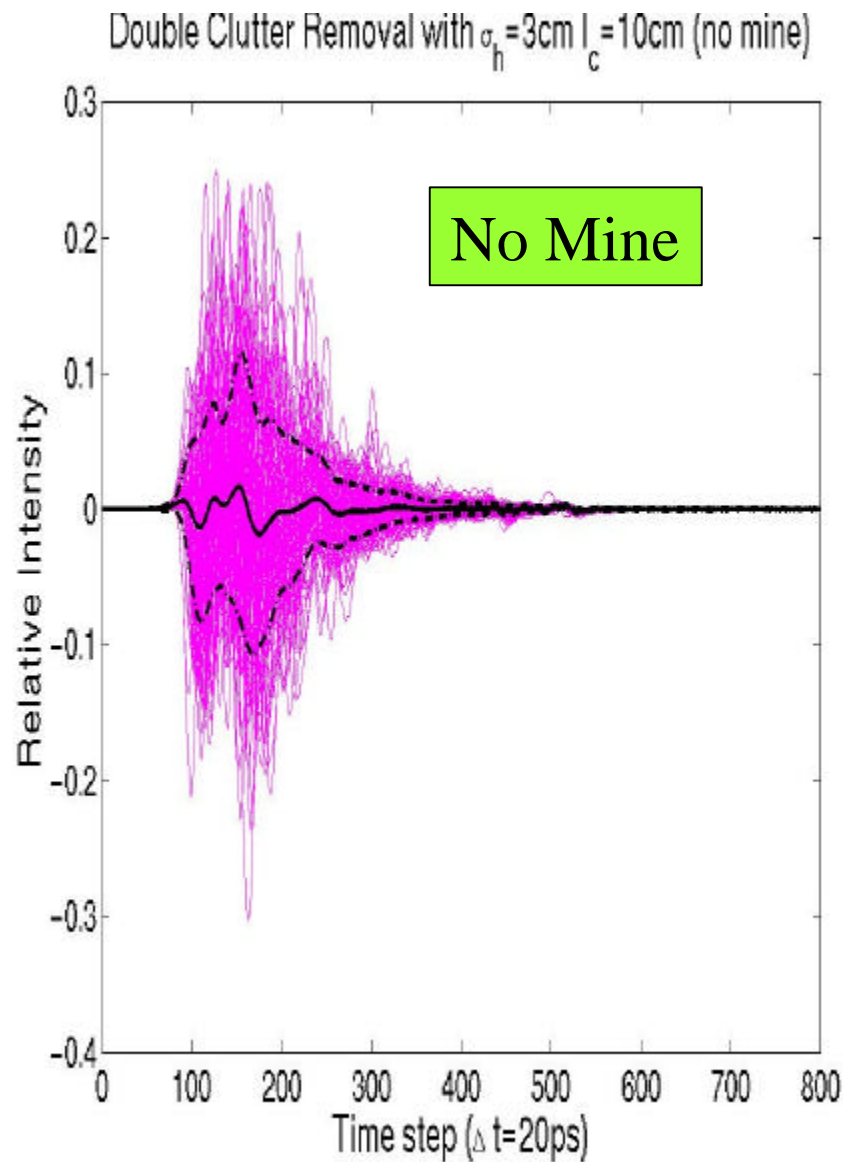
Original Mine Signals with $\sigma_h = 3\text{cm}$ $l_c = 10\text{cm}$



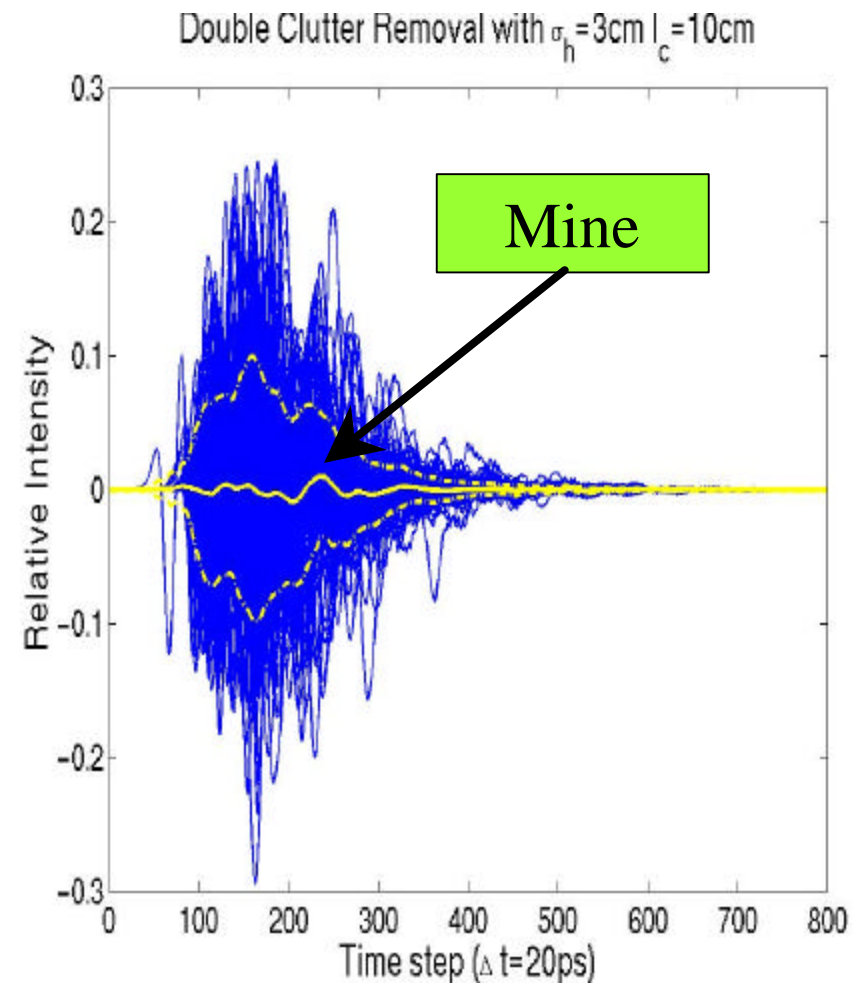
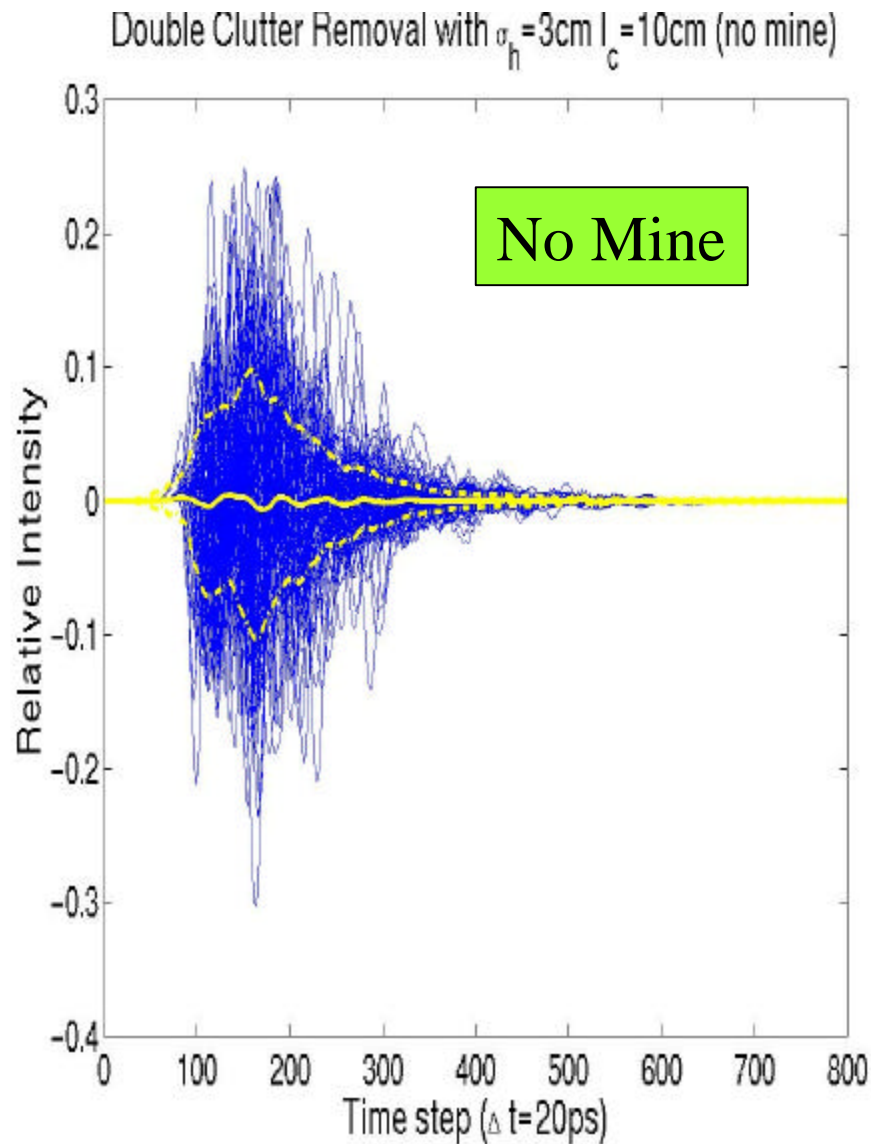
Single Ground Clutter Signal Removal



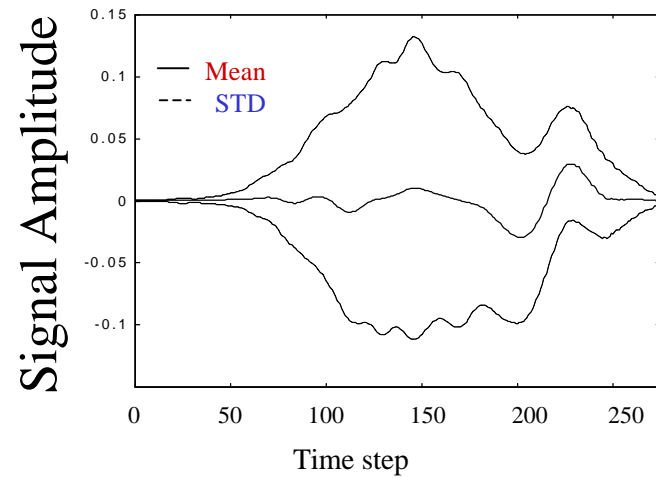
Double Ground Clutter Signal Removal



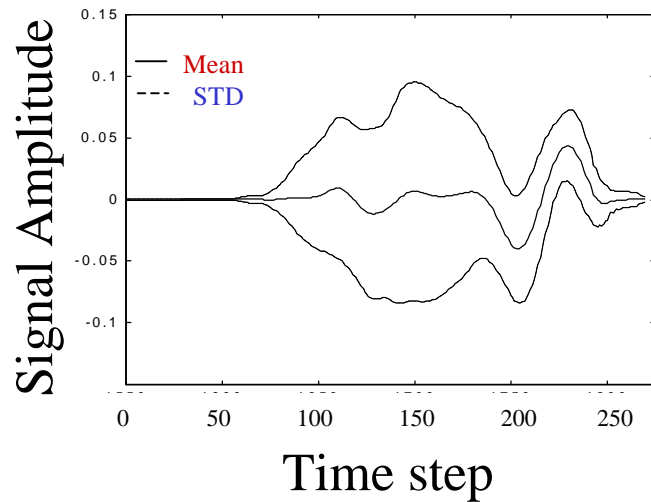
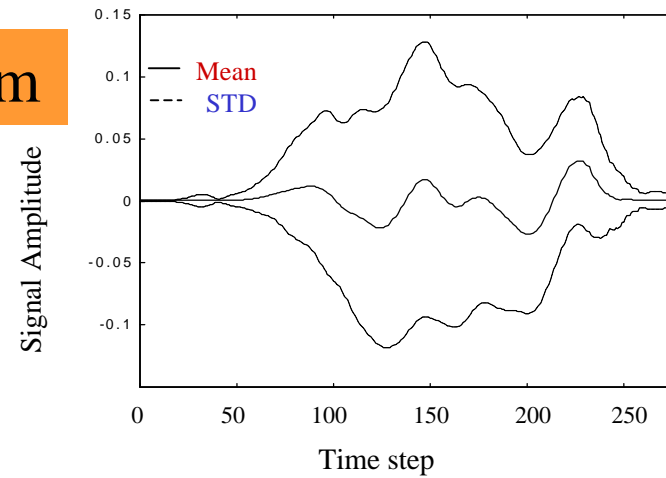
Realigning Signals to Presumed Mine Position



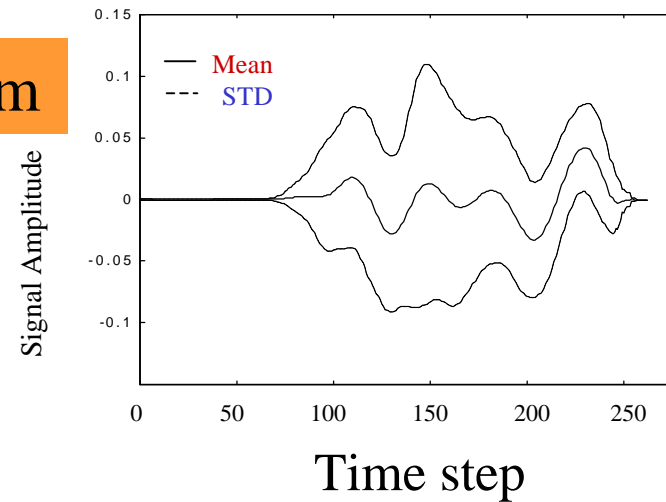
Average Mine Scattered Signals



$\sigma_h = 3$ cm



$\sigma_h = 2$ cm



$L_c = 10$ cm

$L_c = 3$ cm

Conclusions

- Rough ground clutter can be effectively analyzed in terms of:
 - Signal time delay
 - Signal amplitude increase/reduction
 - Pulse distortion
- Signal distortion and time delay depend directly on both surface roughness and roughness scale.
- Target signals are difficult to separate from ground clutter signals because they are small and occur near clutter peaks.
- Identifying and removing the ground signal -- and realigning with the propagation time of a given soil -- enhances the target signal.

Dual Frequency Microwave-Enhanced Infrared Thermography

By

Tianchen Shi, Gerhard O. Sauermann, Carey M. Rappaport, Charles A. DiMarzio

CENSSIS, the Center for Subsurface Sensing and Imaging Systems
235 Forsyth, Northeastern University, Boston, MA 02115

Voice #: (617) 373-8570

Abstract

Humanitarian landmine detection and clearance is one of the most challenging, difficult and time-consuming tasks to be completed with existing technologies. Infrared (IR) Imagery images differences in heat transfer on the surface of the soil due to a buried object. Based on sunlight heating, it is only useful at certain times. Microwave heating has been proposed to enhance the thermal signature, but it is limited by surface roughness.

We have proposed a method called *Dual Frequency Microwave Enhanced Infrared Thermography*. Heating with microwaves instead of natural sunlight leads to a number of advantages, such as more efficient heating, more feature parameters like frequency, modulation and incident angle. Moreover, two different frequencies are used consecutively, and the heating image subtracted to minimize background noise introduced by the rough, irregular surface of the ground itself, and vegetation covering the ground. The dependence of scattered fields on frequency makes this possible.

A 2-D computational model of this method has been developed to simulate real-world landmine detection. The model includes two main parts: a microwave model to get an image of absorbed power density and a 2-D thermal diffusion model to get temperature distribution on the surface of the ground. Finally, to evaluate the performance of a system using this method, ROC (Receiver Operational Characteristic) curves are used.

Microwave-Enhanced Infrared Thermography

by

Charles A. DiMarzio

Carey M. Rappaport

Li Wen

Gerhard O. Sauermann

Center for Electromagnetics Research

Northeastern University

Boston, Massachusetts 02115

and

Herman E. Scott

Aerodyne Research, Inc.

45 Manning Road

Billerica, MA 01821-3976

ABSTRACT

Microwave heating of the ground can enhance infrared signatures of buried objects. The extent of the enhancement depends upon many parameters including wavelength, polarization, angle of incidence, and properties of the soil and the object. We show that angle of incidence and dielectric properties of the object are important, with some analytical and experimental results.

Key Words: Microwave heating, infrared imaging, landmine detection.

1. INTRODUCTION

We have previously reported analytical and experimental results showing that microwave heating of the ground can produce enhancements of the infrared signature of buried objects [DiMarzio, *et. al.*, Sept. 1998]. While diurnal variations in natural heating provide some useful signatures, the microwave heating offers the opportunity to enhance the signals by heating the soil below the surface and the object itself. Two mechanisms contribute to the unique signatures of buried objects. The first, evident at short times, is a modification of the electromagnetic field of the microwave source by the presence of an object with dielectric properties which contrast with those of the ground. This results in changes in the heating pattern of the soil including that near the surface, and appears almost immediately. At later times, an image appears which is quite useful for identifying

the shape of the buried object. This signature depends on the difference in absorption of microwaves between the object and the background, and on the contrasting thermal properties. This signature remains for many minutes after heating is ceased. The combination of these two signatures offers a powerful new tool for the detection and characterization of buried mines and mine-like objects.

2. VARIATION WITH ANGLE AND POLARIZATION

One of the most obvious discriminants is the combination of angle of incidence and polarization. The two-dimensional model was run with a dielectric mine and different angles of incidence for TE and TM polarizations, all at 2.54 GHz. Figure 1 shows some of the results for different angles with TM polarization. The color scale shows the amount of absorbed power. The axes in each plot are spatial, with the vertical representing depth and the horizontal representing transverse position. The microwave energy is incident as a plane wave from the top left at the indicated angle with respect to the vertical. The major result is that the absorbed power decreases with increasing angle as a result of the changing Fresnel reflection at the surface. However, a smaller effect is also noted, in that the region of maximal absorption moves away from the beam with increasing angle. The absorption is almost entirely in the soil rather than in the mine, which has a lower absorption coefficient. The soil heating pattern at the surface is rather complex, resulting from an interference pattern between the incident beam and the scattered one. In a cluttered environment, this heating pattern may make it difficult to detect buried objects directly using the microwave-enhanced infrared signatures. On the other hand, the pattern provides information about the field distribution, which may prove useful in conjunction with imaging using ground-penetrating radar.

A similar result was obtained with the model using TE polarization. Differences do exist between the two, but they are in all cases less than 10 percent of the total, and are much less than that at the surface. Thus these differences would make unreliable indicators of buried objects. A slight advantage exists for TM polarization as a result of the decreased Fresnel reflection at the surface at angles near the pseudo-Brewster angle. However, this effect is not very strong in materials such as soil with high absorption. While the reflectivity for TM polarization does reach a minimum at the pseudo-Brewster angle, the minimum is not very deep, and the ratio of Fresnel reflection between TM and TE polarization is not far from unity. Because the dielectric constant is high, the wave bends strongly toward the normal, and the interaction of the wave with the target is similar for the two polarizations. This result would change for targets without circular symmetry in a three-dimensional model, but it is unlikely that the difference could be exploited effectively to distinguish between mines and other objects under the surface.

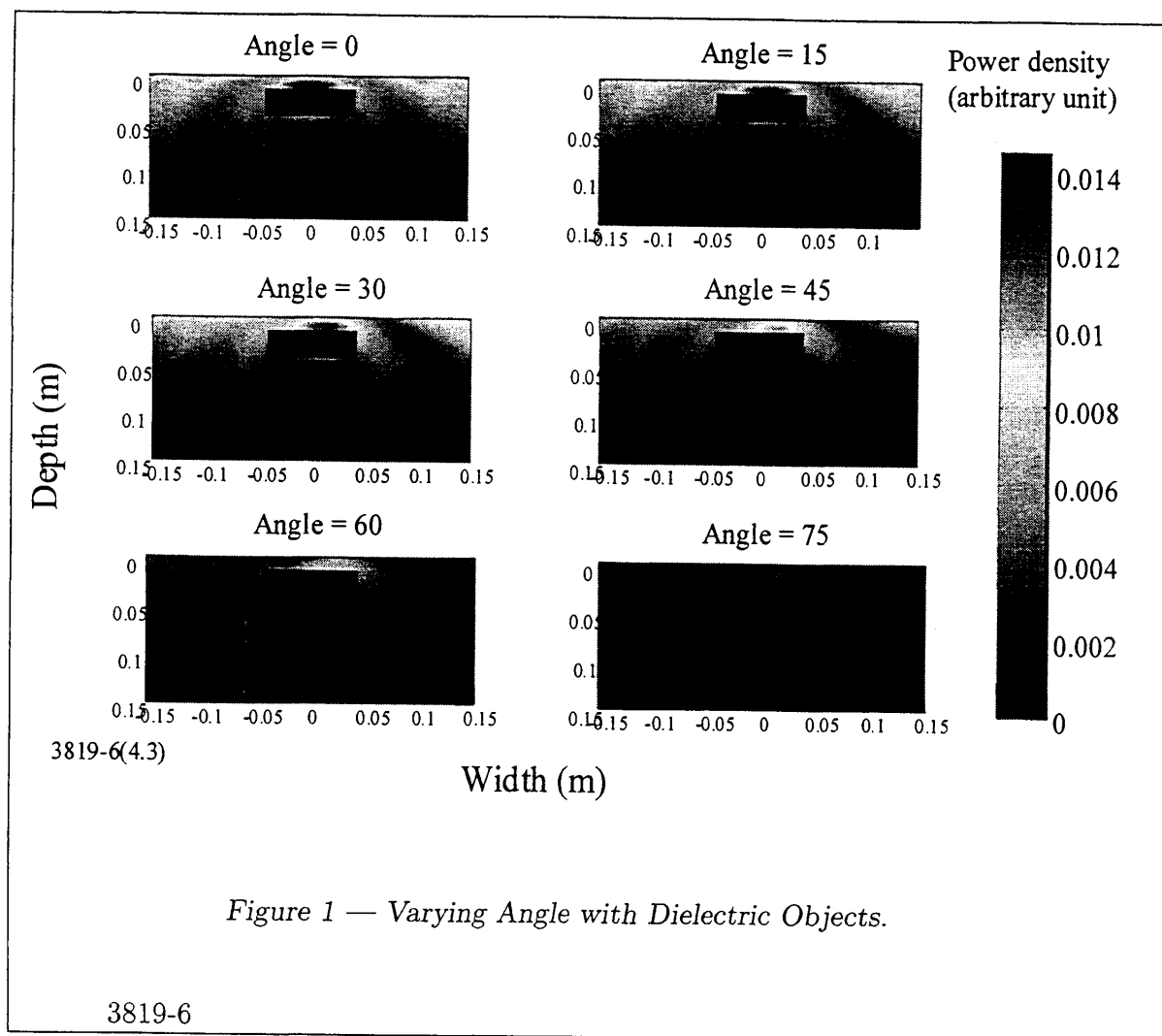
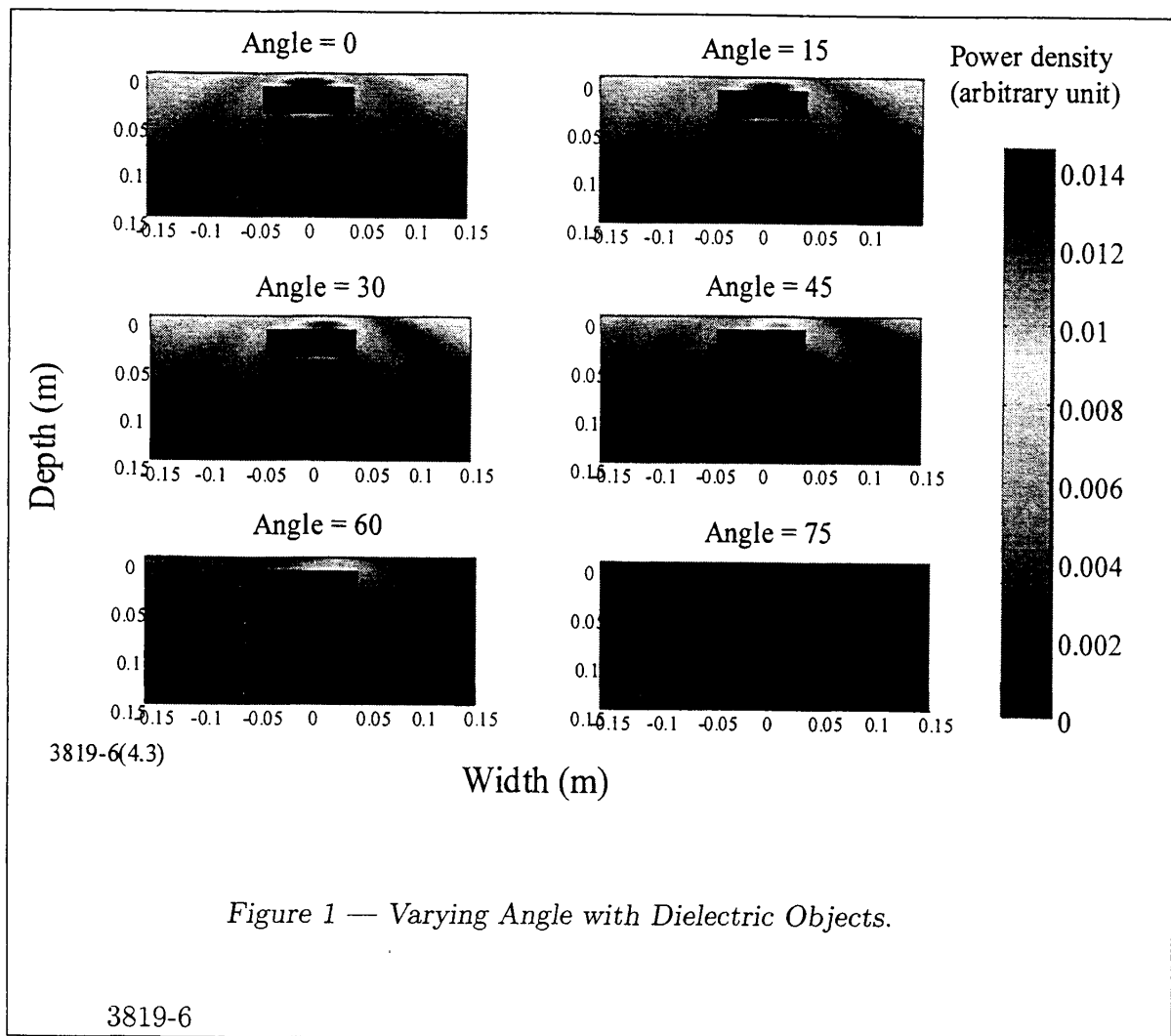


Figure 1 — Varying Angle with Dielectric Objects.

3. VARIATION WITH MICROWAVE ABSORPTION

Results similar to those of Figure 1 were obtained with absorbing targets, but the energy was absorbed primarily in the target rather than the background. This results in a considerably different time history than is observed for the dielectric mines. Figure 2 shows three different objects, having low, medium, and high absorption. The top images are in the same configuration as those in Figure 1, and show the distribution of absorbed power. The bottom images show how surface temperature varies with time. The horizontal axis is the same as in the top figures, and the vertical axis represents time. The color shows the temperature at the surface.

For all objects, heating begins immediately with the application of power at the top of the time axis, and the interference pattern is readily evident. In a three-dimensional



3. VARIATION WITH MICROWAVE ABSORPTION

Results similar to those of Figure 1 were obtained with absorbing targets, but the energy was absorbed primarily in the target rather than the background. This results in a considerably different time history than is observed for the dielectric mines. Figure 2 shows three different objects, having low, medium, and high absorption. The top images are in the same configuration as those in Figure 1, and show the distribution of absorbed power. The bottom images show how surface temperature varies with time. The horizontal axis is the same as in the top figures, and the vertical axis represents time. The color shows the temperature at the surface.

For all objects, heating begins immediately with the application of power at the top of the time axis, and the interference pattern is readily evident. In a three-dimensional

example, this pattern would result in rings around the object. When the power is turned off at time zero, the ground containing the low-absorption object begins to cool almost immediately.

For the more absorbing objects, the surface temperature continues to rise after the power is turned off, as heat from the object is conducted toward the surface. Figure 3 shows a typical temperature time history from the model for the most absorbing object shown in Figure 1, and below that, experimental results for a hockey puck, 76 mm in diameter and 25 mm thick buried in dry sand. The temperature of the surface continues to rise for 10 minutes after the power is turned off. Each graph on the right shows the temperature as a function of time for a column near the center of the corresponding color plot.

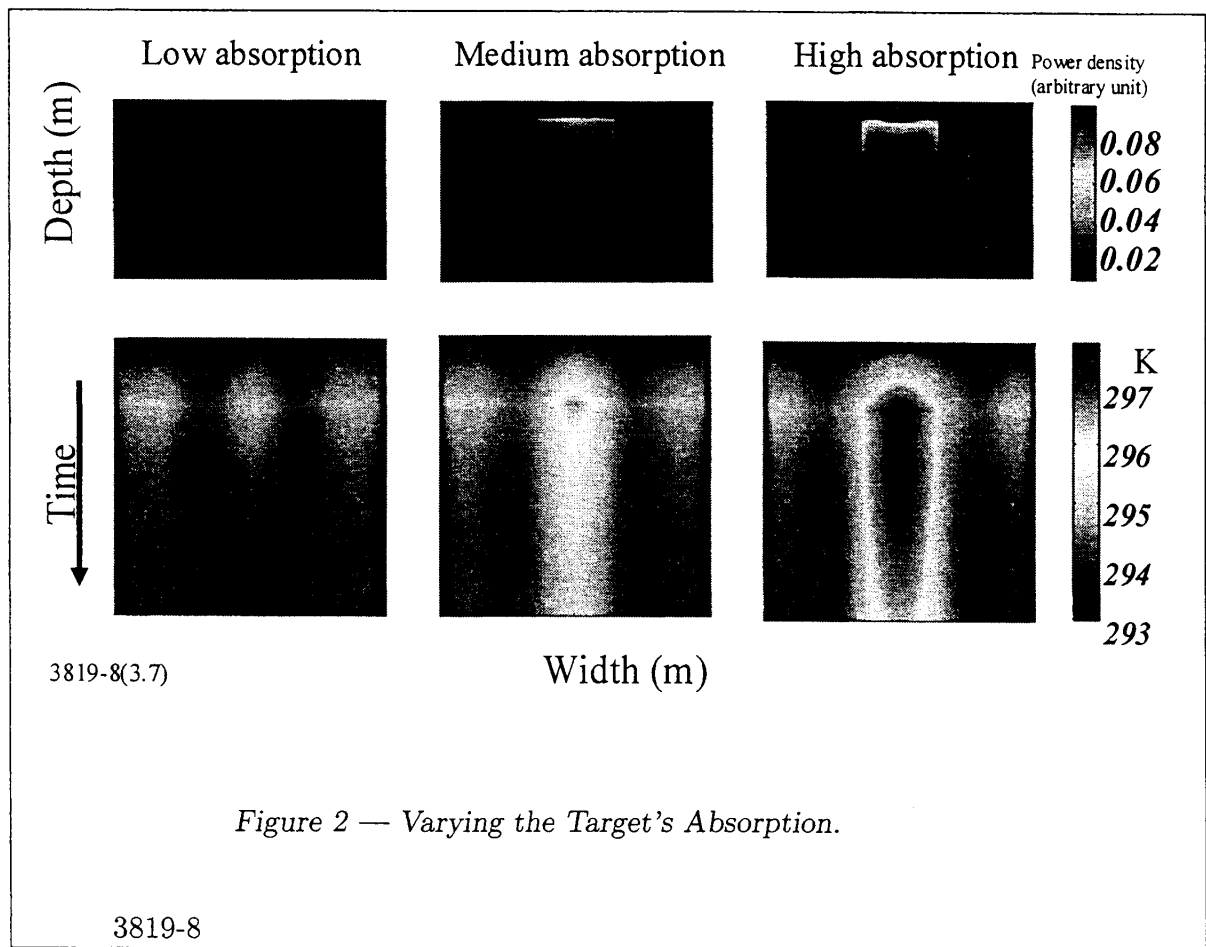
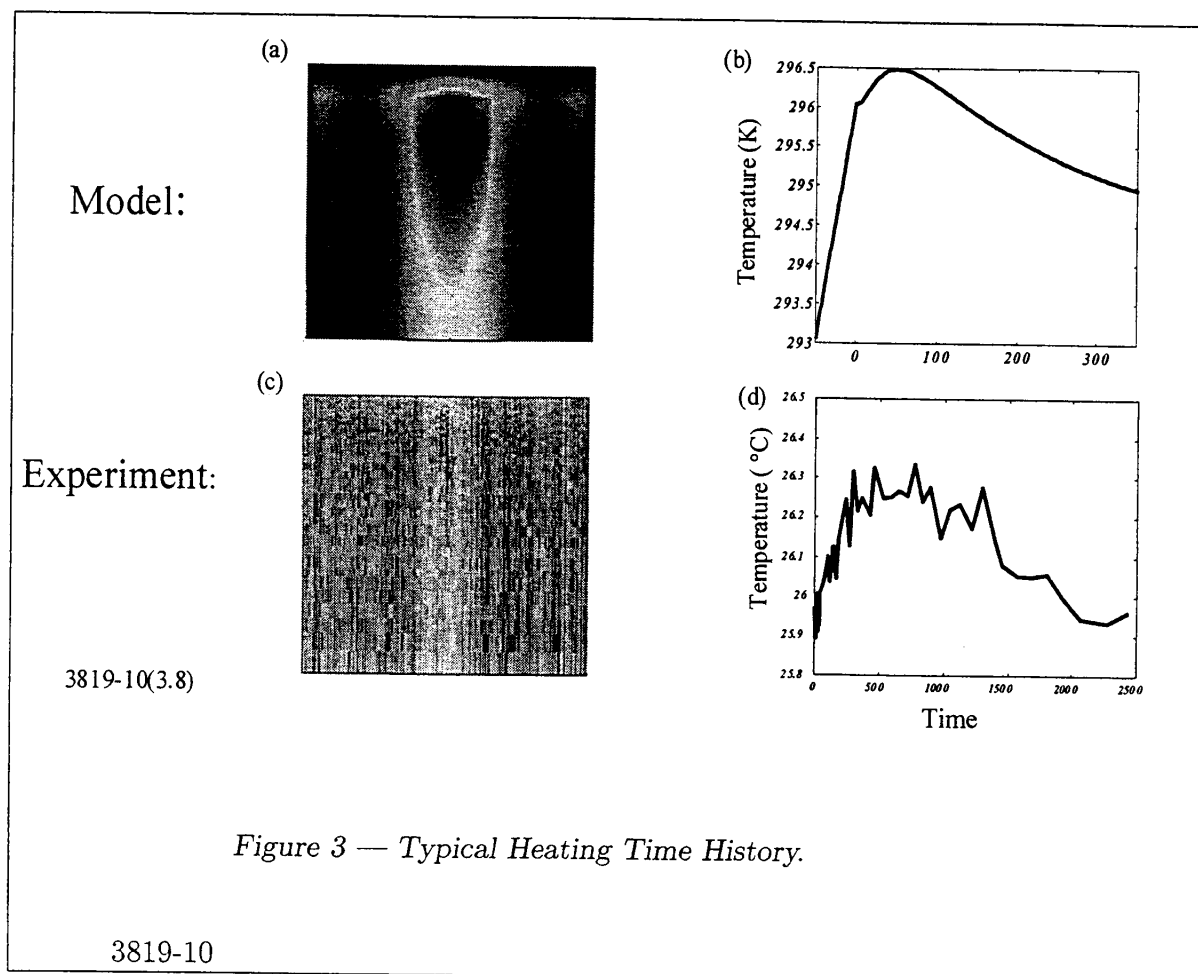
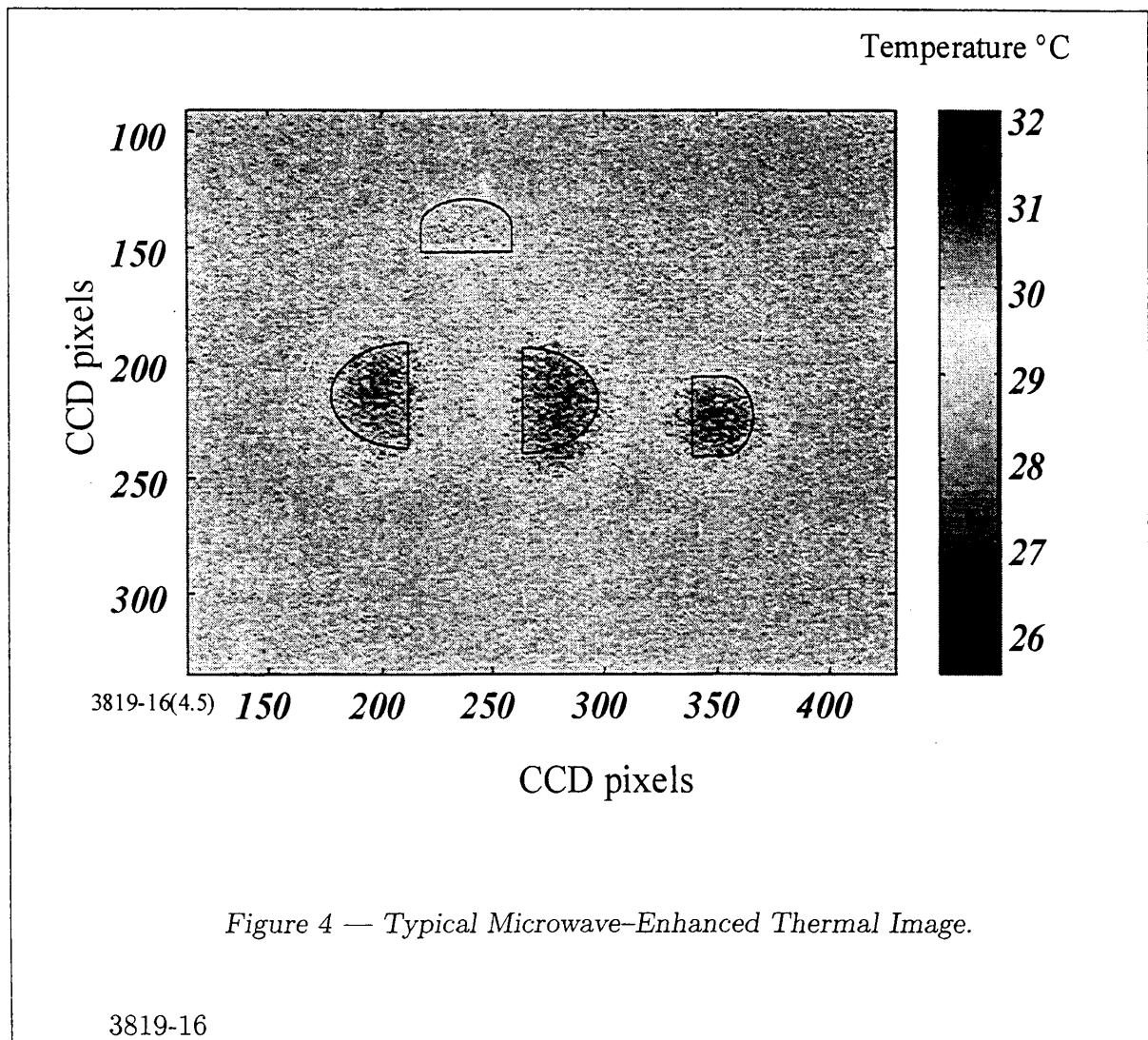


Figure 2 — Varying the Target's Absorption.



4. IMAGING THE TARGET SHAPE

Finally, we examine our ability to recognize the shape of buried objects. For this purpose, we cut hockey pucks into distinguishable shapes and buried them at a depth of about 1 centimeter in dry sand. Figure 4 shows a sample infrared image, overlaid with a line drawing of the objects. We note that the semi-circular shape of the half pucks is evident. One of the small segments shows a rather weak signature, and the other shows a stronger one. Location and aspect ratio are easily determined, although the subtle changes in shape are not readily distinguishable in this picture. We conclude that shape information can be obtained on objects with dimensions of a few centimeters. We anticipate that this information could help eliminate false alarms caused by rocks, roots and other natural buried objects.



5. SUMMARY

We conclude that microwave-enhanced infrared thermography produces useful signatures for the detection of buried mines and minelike objects. Signatures exist and change over a period of minutes after microwave illumination. Initial signatures show information about electromagnetic wave propagation in the vicinity of the object while later ones reveal more about its thermal properties. These images, particularly at the later times, provide useful shape information about objects which are more than a few centimeters in size.

6. ACKNOWLEDGEMENT

This work was sponsored by the OSD MURI Program under Army Research Office Grant number DAAG55-97-1-0013.

7. REFERENCES

DiMarzio, Charles A., Carey M. Rappaport, and Li Wen, "Microwave-Enhanced Infrared Thermography," *Detection and Remediation Technologies for Mines and Minelike Targets III, SPIE Aerosense Volume 3392*, September 1998. Pp. 1103-1110.

DiMarzio, Charles A., Carey M. Rappaport, and Gerhard O. Sauermann, "Microwave-Enhanced Infrared Thermography," *Industrial and Environmental Monitors and Biosensors, SPIE Volume 3534* and presented in Boston, MA, on 4 November 1998.

Laser Induced Acoustic Imaging of Underground Objects

Wen Li, C. A. DiMarzio, S. W. McKnight, G. O. Sauermann, E. L. Miller
Center for Electromagnetics Research, Northeastern University, Boston, MA 02115

ABSTRACT

This paper introduces a new demining technique based on the photo-acoustic interaction, together with results from photo-acoustic experiments. We have buried different types of targets (metal, rubber and plastic) in different media (sand, soil and water) and imaged them by measuring reflection of acoustic waves generated by irradiation with a CO₂ laser. Research has been focused on the signal acquisition and signal processing. A deconvolution method using Wiener filters is utilized in data processing. Using a uniform spatial distribution of laser pulses at the ground's surface, we obtained 3D images of buried objects. The images give us a clear representation of the shapes of the underground objects. The quality of the images depends on the mismatch of acoustic impedance of the buried objects, the bandwidth and center frequency of the acoustic sensors and the selection of filter functions.

Keywords: Photo-acoustic, Laser-induced, Acoustics, Buried object detection, landmine detection, pavement quality.

I. INTRODUCTION

This paper investigates the photo-acoustic process, in which the optical energy is converted into acoustic energy. This phenomenon has been observed as early as the 1960's when sound pulses were produced on absorption of pulsed ruby-laser radiation^[1].

When a laser pulse hits the surface of a medium (solid or liquid), several processes occur in a short time, including reflection, absorption (with side-effects such as optical to thermal energy conversion), heat conduction, medium expansion and contraction. Although we have not constructed an exact mathematical model for the photo-acoustic process, we can qualitatively understand the mechanisms: (1) Thermal coupling based on the absorption of laser radiation, (2) heating and expansion occurring in the medium. (3) excitation of sound caused by sudden changes in medium volume.

Hidden objects can be located by echo location. This is a well-known idea in everyday life, as in the example of tapping a wall or floor to find studs or holes. This idea could also be applied to the process of evaluation of pavement quality, and buried mine location.

Traditionally, acoustic waves can be generated by direct physical stimulation, for example using an acoustic transducer. Laser-induced acoustic waves have more advantages in detecting objects:

- (1) This is a non-contact method of sound generation. This is the most important advantages for detection of dangerous object like landmines.
- (2) Coupling is easy and efficient because over 90% of optical radiation at 10.6um is absorbed in soil. Thus, the sound wave is generated directly in the ground.
- (3) The process is flexible. The width and power of the laser pulses can be controlled, the size of the laser spot can be focused and the scanning of laser shots can be easily implemented and changed by an optical scanner.

II. ACOUSTIC ECHO AND SIGNAL PROCESSING

The solution of acoustic wave transport in a medium can be described by applying elastic wave theory. In the most simple case, we consider the monochromatic plane wave in a liquid or in an isotropic elastic solid. Combining the acoustic wave equations and appropriate boundary conditions, we obtain the reflection coefficient,

$$R = \frac{Z_{II} - Z_I}{Z_{II} + Z_I}$$

where $Z_I = \rho_1 c_1 / \cos \theta_I$ and $Z_{II} = \rho_2 c_2 / \cos \theta_{II}$ are the acoustic impedances in two different media. ρ , c and θ are the density of the medium, velocity of the acoustic wave and propagation angle, respectively.

Table 1: Density and Velocity of sound in various media^[2].

Substances	Density g/cm ³	V _l m/s	V _s m/s
Aluminum, rolled	2.7	6420	3040
Copper, annealed	8.93	4760	2325
Glass, pyrex	2.32	5640	3280
Lucite	1.18	2680	1100
Polyethylene	0.90	1950	540
Polystyrene	1.06	2350	1120
Rubber, butyl	1.07	1830	
Rubber, gum	0.95	1550	

V_l: Velocity of plane longitudinal wave in bulk material

V_s: Velocity of plane transverse (shear) wave

The large acoustic impedance mismatch between soil/sand and land mine is the prerequisite condition of using acoustic echo detection. Table 1 shows the densities and velocities of sound in different media. In our experiment, we use dry sand as an acoustic medium, with the density of sand being about 2.6g/cm³. The velocity of sand varies according to the density and moisture content of the sand. In our samples, we measured the speed of sound in dry sand as about 280m/s, 20% less than the speed of sound in air. Compared with the speed of sound in metal, plastic or rubber, the mismatch of acoustic impedance is large enough for echo detection. Another important point is that the paths of reflection and refraction obey Snell' law:

$$\frac{\sin \theta_I}{c_1} = \frac{\sin \theta_{II}}{c_2}$$

Thus, we can calculate the relations between the position of the mine and the signal track of the mine. Figure 1 illustrates this concept. As we move the laser spot toward a target (a piece of aluminum), the arrival time of the reflected signal should follow a straight line with a shape given by the speed of sound in the medium.

Water is a very good medium for acoustic waves. The soil or sand in which a mine is usually buried dampens sound waves, therefore, the echo from the target is usually very weak in soil/sand. It is hard to find the target directly in this case. Also, from the signal processing point of view, the data acquisition method by scanning a laser spot and burying a detector in the soil makes interpretation difficult. The small signal from a buried target is often lost in changes of the larger direct signal which results from varying distance from source to detector.

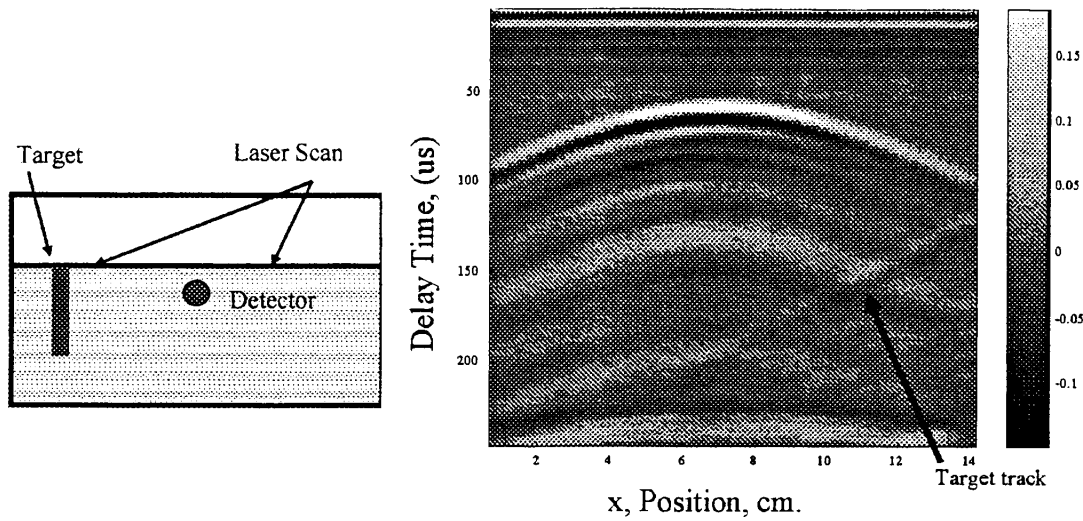


Figure 1. Acoustic echo in water

We considered another method for data acquisition that involves placing the detector over the surface of ground and making it relatively fixed respect to the laser spot, instead of burying the detector into the soil. This method lets us obtain a constant delay time, unless the surface is not perfectly flat.

Figure 3 shows some results acquired by using this method. In the cases where the mine is exposed to the surface of the ground, the laser pulse hits the surface of the mine directly, and the difference of acoustic excitation between the mine and the sand is easily detected. However, in the case of a buried mine, because of strong reflections at the boundary between sand and air, the sound echo from the buried mine is so weak that we can not find information about buried objects from the raw data directly. Thus, an appropriate method of digital signal processing is necessary.

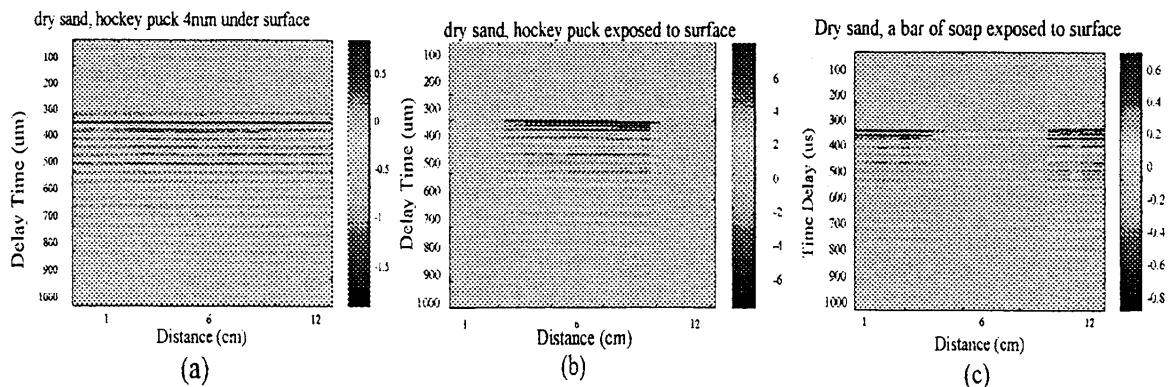


Figure 3. The sound echo. (a) Buried hockey puck 4mm under surface of sand. (b) A hockey puck exposed to the surface of sand. (c) A bar of soap exposed to the surface of sand.

Let us consider the procedure of acoustic echo detection again. After absorption and expansion, the spot of the laser pulse becomes a source of a pulsed sound wave. Echoes of the sound pulse arrive at the detector at different times after following different paths, as shown in Figure 4.

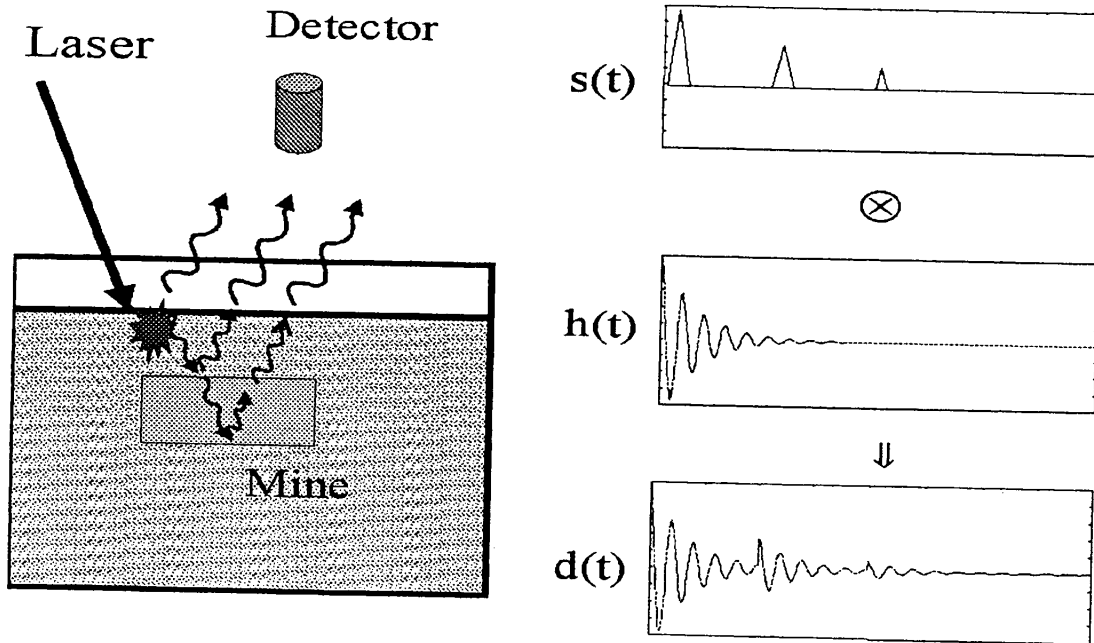


Figure 4. The signal analysis of reflection. S: The sound pulses from surface of ground, top surface of mine and bottom surface of mine. H: The impulse response of the detector. D: The resulting signal from detector.

The delay time contains information about the position of the ground surface and depth of the mine. The acoustic signal received by the detector is the convolution of the sound pulse and the impulse response of the acoustic detector. Because of the limited bandwidth of the detector and the attenuation of the echo signal, the reflection from the underground target is occluded by the strong signal from the surface of the ground. In order to pick up the sound echo from the buried mine, we can use an inverse processing technique in the frequency domain,

$$S = D / H$$

where D is the signal we got from the detector, H is the impulse response of the acoustic detector. Thus, $1/H$ is used as a filter function multiplied the received signal to obtain the original reflection distribution. At the points where the magnitude of H is near zero, the inverse process will produce a large amplification of noise. Thus, we use a slightly more complicated inverse method to limit the effects of noise:

$$Filter = \frac{H^*}{H \cdot H^* + \epsilon}$$

The new filter function avoids the problem of large amplification at small magnitude of H . The value of ε gives a general idea of the noise margin. The impulse response of the detector is usually a complex number in the frequency domain. We can use the real number $H \cdot H^*$, which is the intensity of the impulse response, and add it to ε . When the intensity of the impulse response is very large compared to ε , the filter function is nearly $1/H$, which is an inverse filter; when the intensity is small, the filter function is H^*/ε which represents a matched filter in the frequency domain.

From Figure 4 we can see that the filtering process is actually the inverse process from D (the detector signal) via H (impulse response of detector) to S (the sound echo) in the frequency domain. We call this inverse process deconvolution filtering.

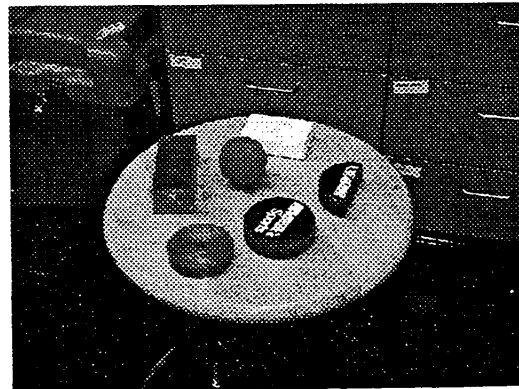
III. EXPERIMENTAL RESULTS

A Laser Science LS150 pulsed CO_2 laser was used in our experiment to generate an acoustic pulse. The wavelength of the laser is 10.6 μm and the pulse width is about 100ns, with the energy per pulse being 150mJ. We buried a mine into a 30cm by 60cm sand box, with the depth of the sand about 10cm. We believe that the acoustic pulse should be very narrow, meaning it has wide bandwidth in frequency domain. Unfortunately, wide-band audio frequency transducer are not readily available to collect all of the information contained in the sound pulses. We have tried several detectors in our experiment, include two identical seismic transducers and two audio microphones with different frequency response. We used an HP-54615b Oscilloscope for data acquisition, and a DAQ multifunctional I/O board for system controlling (control of optical scanner, laser triggering and synchronization). All of the signal processing work has been done on a computer running Matlab.

We concentrated our research on non-metallic mines. We used some simulants that compare favorably to the constituents of a non-metallic land mine, include plastic box, racquet ball, soap and hockey puck. Figure 5 shows the pictures of the experimental setup and the simulants of mine.



(a)



(b)

Figure 5. (a) The experiment setup. (b) Simulants of mine

Figure 6 shows some results after the deconvolution filtering we discussed before. We used a seismic transducer as the acoustic detector which has 29kHz resonance frequency with narrow bandwidth of about 5kHz. We made a scan on the surface of sand with 25 pulses and a track of 12.5cm long, which crosses the buried mine. When the hockey puck was buried vertically in the sand box, after filtering, we obtained the image shown in panel (a) of Figure 6, which shows the circular shape of the puck. We also buried a hockey puck horizontally with the surface parallel to the surface of sand, then we obtained the result in panel (b) which shows the surface of the puck. If the puck is tilted to the surface, the reflection track is also tilted (panel c). The other picture is a bar of soap buried in sand box.

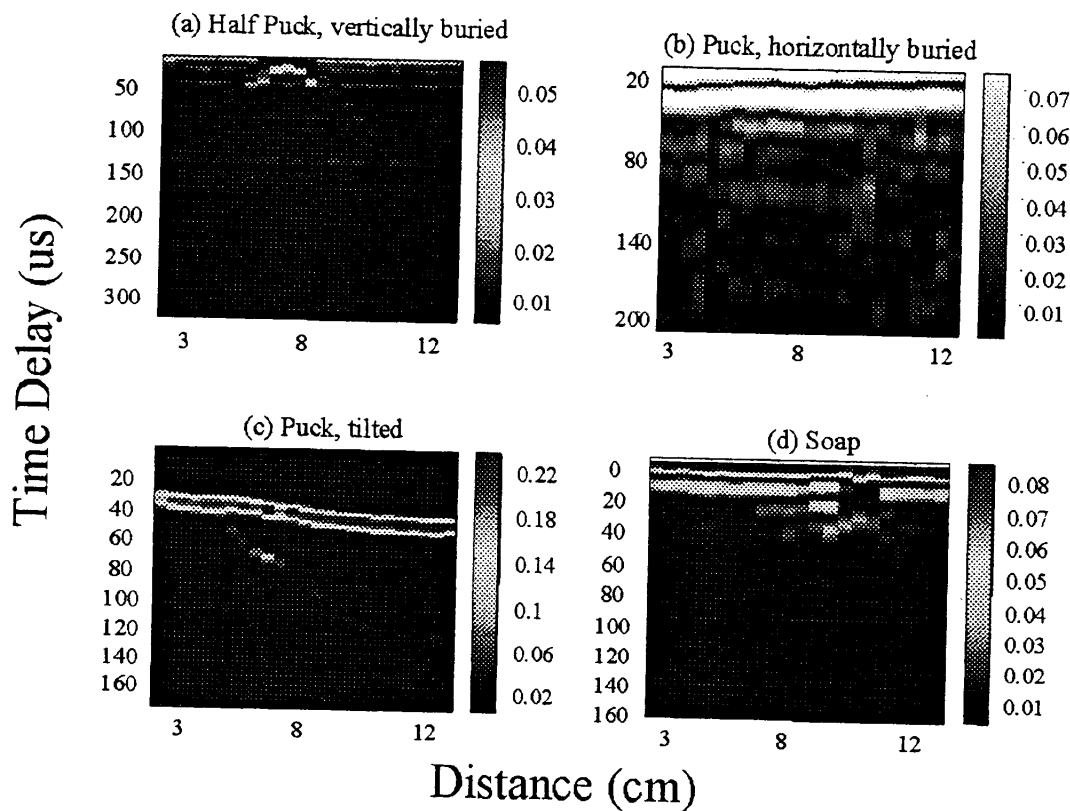


Figure 6. (a) Half hockey puck vertically buried in dry sand. (b) Hockey puck horizontally buried in sand with surface parallel to the surface of sand. (c) Hockey puck buried in sand with surface tilted compared with the surface of sand. (d) A bar of soap buried in sand.

Clearly, the results in Figure 6 provide the depth and shape information of underground objects. Furthermore, because the time delay of the sound echo is proportional to the depth, a surface scan of laser shots in 2 dimension, combined with the depth information, provides a 3-D picture of underground objects. Figure 7 shows the result of a surface scan for a half hockey puck. The detector is still the seismic transducer. We make slices, each with a different time delay, to show the strength of the reflection at that depth. We then combine these slices to produce a 2-D map as would be seen from above. Figure 7 shows several of these slices and the combined image.

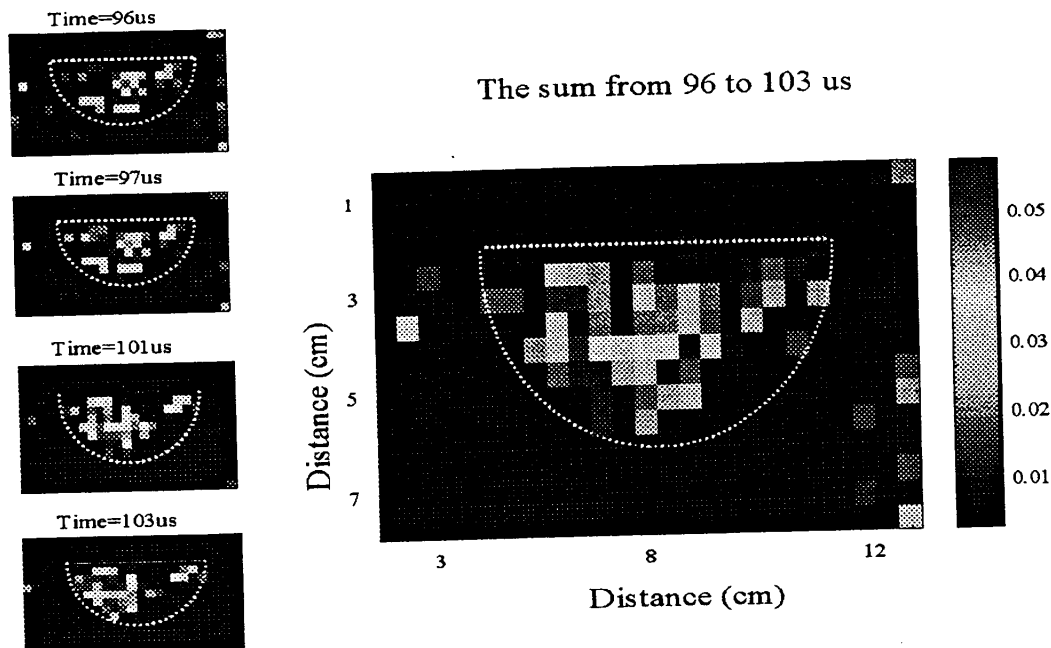


Figure 7. The 2-D scans for a buried half hockey puck: At different times, we can image different depth slices for underground imaging.

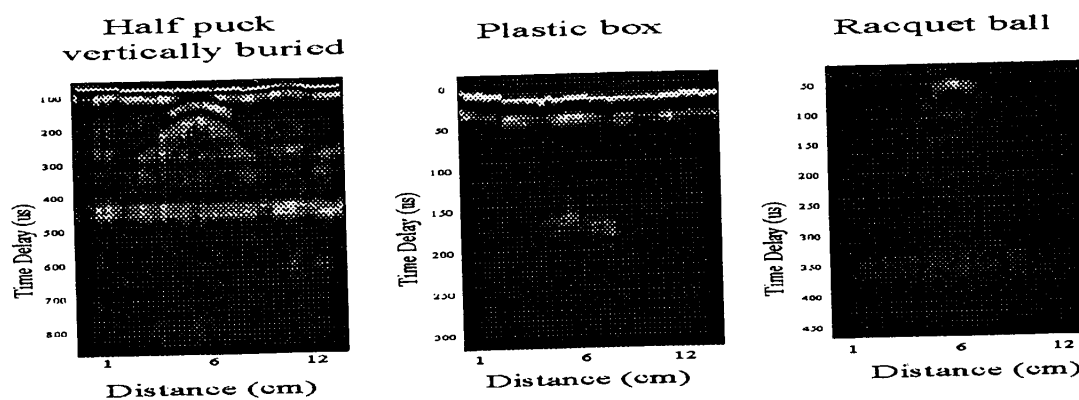


Figure 8. The experiment using an audio microphone. The simulants of mine are a vertically buried half hockey puck, an empty plastic box and a racquet ball respectively.

In the frequency domain, the acoustic detector only samples a part of the information contained in the sound pulse. Tests using different detectors are shown in Figure 8. We used an audio microphone which has wider bandwidth and lower cutoff frequency as acoustic detector. Different mine simulants including the half hockey puck, plastic box and racquet ball have been used in the experiment. The signal seems stronger than before because of the wider bandwidth. We also did a 2-d scan at the surface for a buried half hockey puck shown as Figure 9, and the result shows that the shape resolution is dimmer than it was with transducer. Thus we conclude that the resolution is degraded because the higher frequencies of sound (shorter wavelength) are not used.

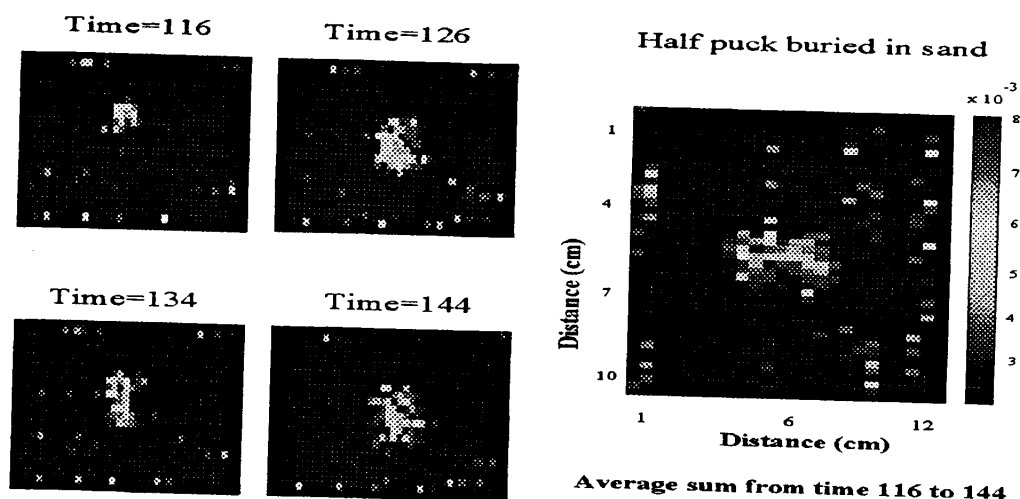


Figure 9. The 2-D scans for a buried half hockey puck using the microphone as an acoustic detector: At different times, we can image different depth slices for the underground imaging.

IV. CONCLUSION

The experiments shown in this paper demonstrate that laser induced acoustic imaging is feasible for mine detection. As one of the prospective demining techniques, laser induced acoustic wave has many advantages, such as non-contact, easy control and detection of non-metallic mines.

The selection of data acquisition and signal processing method is also very important. We found that fixed detector over the surface of ground with constant relative position to the laser spot is a good acquisition method, and the deconvolution filter works well for picking up the information of underground objects.

The echo of the sound pulse has a very wide bandwidth in the frequency domain, but we only obtained a small part of the whole information because of the limitation of our acoustic detector. As shown in the experiment, the higher the center frequency, the better the shape resolution; the wider the bandwidth, the stronger the echo signal. Therefore, the ideal detector should have white frequency response.

The large acoustic mismatch between soil and buried mine promises success of acoustic imaging. By detecting the reflected sound pulse from the surface of mine, we can get clear shape information of underground objects, and thereby reduce false alarm rates which are typical of all mine detection techniques.

A mathematical model of the laser acoustic interaction should be a high priorities in the future work. We need to know more about the energy distribution of the acoustic wave. In order to collect more information from the sound echo, a better detector should be used in future experiment. We think that a laser Doppler vibrometer is a good choice, because it is also a non-contact detector with very high sensitivity and almost white frequency response. Also, it can image multiple locations and be scanned with the pulse laser. An operational system of acoustic imaging might be laser induced and laser detected. In addition, different types of soil should be taken into account. We will continue to try more media and do some experiments outdoor in the real ground as well.

V. REFERENCE

- [1]. F.V. Bunkin, A.I. Kolomensky and V.G. Mikhalevich, "Laser in Acoustics", Laser Science and Technology, An International Handbook, Vol.12.
- [2] CRC Handbook of Chemistry and Physics, pp.14-32, 71st Edition, 1990-1991
- [3]. Malcolm J. Crocker, "Encyclopedia of Acoustics", John Wiley & Sons. Inc. 1997.
- [4]. C.Liu and S. R.Nagel, "Sound in sand", Physical Review Letters 68, p.2301, 1992.
- [5]. S.W. McKnight, C.DiMarzio, W.Li, etc. "Laser induced acoustic detection of buried objects" pp.841-847, Proceedings of SPIE, Detection and Remediation Technologies for Mines and Minelike Targets III, 13-17 April, 1998.

Effects of Surface Roughness on Microwave Heating of Soil for Detection of Buried Landmines

Taner R. Oktar, Carey M. Rappaport, and Charles A. Dimarzio

Center for Electromagnetic Research
Northeastern University, Boston MA, 02115

ABSTRACT

Two common techniques proposed for detection of landmines are ground-penetrating radar (GPR) and infrared imagery. Because of the wide diversity of mines, the clutter which is encountered in minefields, and variation caused by the ground surface, the task of interpreting GPR signals is daunting. Likewise, variations in thermal properties of soil, solar heating, clutter, and surface irregularities lead to limited performance for IR imaging systems.

Throughout the past two years, our research has emphasized using the microwave energy to heat the ground with the goal of enhancing the infrared signatures. Wavelength, polarization, angle of incidence, properties of soil, chemical and physical structures of buried objects, and amount of roughness are the parameters affecting the extent of enhancement. We have previously shown a two-dimensional model for flat surfaces. In the present work we included a rough surface, to see the effects of the amount of roughness on both temperature distributions in the soil and temporal and spatial variations of surface temperature. We present results which show that the quality of the infrared surface images is dependent upon surface roughness.

Keywords: Landmines, microwave heating, infrared imaging.

1. INTRODUCTION

Detection of landmines has been a challenging research topic through decades for both military and humanitarian reasons. One of the techniques in solving the landmine detection problem is thermal infrared image processing, which was widely used for the detection of defects in materials, the detection of thermal leaks of power plants, and environmental remote sensing as well [Li, *et. al.*]. This thermal infrared imaging technique has also been investigated by DiMarzio *et. al.* [Sept. 1998] with an efficient modification yielding high performance. The suggested modification was to use a microwave power source enhancing the normal diurnal thermal cycle, since the thermal effects of buried objects were too weak due to the dominant absorption of solar irradiance very near the surface.

Microwave heating of soil has many advantages for the detection of landmines or shallow-buried waste [DiMarzio, *et. al.*, April 1998]. First, it contributes an additional discriminant caused by the contrast between the dielectric properties of soil and target along with thermal properties. Second, this kind of enhancement of diurnal solar cycle accelerates the process by reducing the time necessary for observable temperature changes on the surface of the ground under which an object is buried. Finally, controllability of microwave power source allows us to focus on doubtful areas, decreasing the false alarm rates. Furthermore, DiMarzio, *et. al.* [Sept. 1999] have investigated the angle of incidence of microwave source and dielectric properties of the object due to the fact that enhancement is affected by many parameters, such as, wavelength, polarization type of the waves, angle of incidence, properties of the soil, and physical and chemical structure of the buried object.

Those investigations were performed under the assumption of smooth, almost flat, surfaces. As the roughness of the surface to be explored increases, the signal indicating the presence of a buried object weakens and it gets harder to detect that object. In the experiments carried out at the Army's Gold Regions Research and Engineering Laboratory Hanover New Hampshire, the presence of any buried material couldn't be noticed. In figure-1, one can see the surface clutter, and the indistinctness due to the disorderliness. The roughness is the consequence of the disorderliness on the surface caused by the structure of the ground and the clutter such as vegetation. Here, we examine the effect of roughness on surface temperature increase by introducing new 2-dimensional absorption and thermal models.

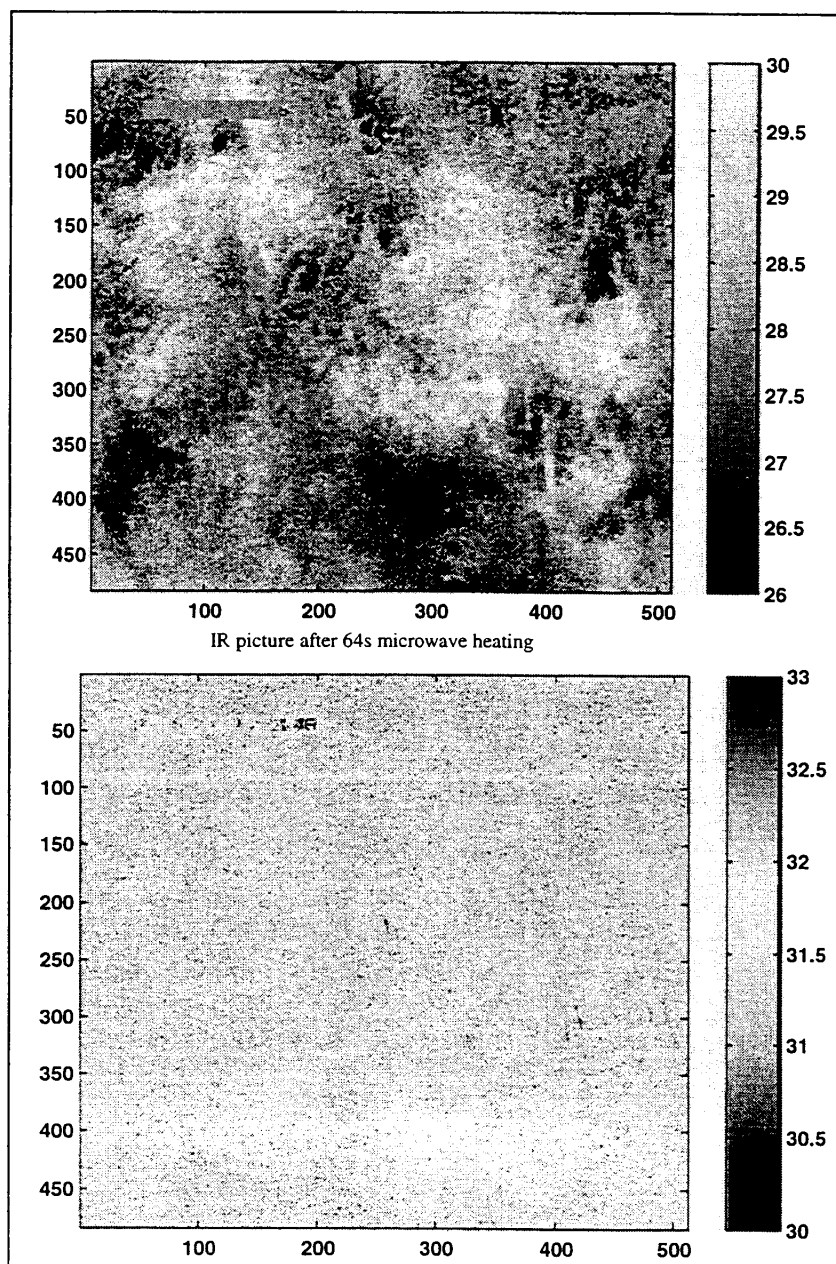


Figure 1. Surface clutter and detection difficulties due to this disorderliness.

2. TWO-DIMENSIONAL MODEL FOR ROUGH SURFACES

The new two-dimensional model for rough surfaces includes two main parts: An absorption model and a thermal model. These two models will be explained in detail in further paragraphs.

2.1. Absorption Model

The absorption model, based on the fundamentals of electromagnetic theory, uses the finite-difference, frequency-domain technique. In order to be able to avoid the reflections due to the computational limits, 8-layer PML boundaries surrounding the two-dimensional computational grid have been implemented into the model. These layers provide the electromagnetic grid to avoid reflections and have it propagate in an unbounded medium. Since roughness is introduced into the absorption model, the two-dimensional computational grid is composed of three different media: Air, soil, and target. A sample grid can be viewed in Figure-2. This model takes as inputs the horizontal and the vertical dimensions of the target to be buried under the soil and its location. Furthermore, the incident angle of excitation wave, the dielectric properties and the conductivity of each medium are described. The model forms the binary images of the rough surface and the target under the soil and calculates the absorbed power pattern in Watts/m³ resulting from both incident and scattered waves according to

$$P = \frac{1}{2} \bar{J} \cdot \bar{E} = \frac{1}{2} \sigma \bar{E} \cdot \bar{E} = \frac{1}{2} \sigma |\bar{E}|^2, \quad (1)$$

formulation. A sample absorbed power pattern has been demonstrated in Figure-3. The specifications of the material examined as target, ground, and air have been mentioned in Table-1.

It may be concluded that the surface just over the target would be cooler than elsewhere since the target heats up less than the ground surrounding it [DiMarzio, *et. al.*, Sept. 1998]. Nevertheless, due to the fact that electromagnetic part of the problem is much more complex, one must think over the other factors, such as the dielectric properties, location of the target, its size, *etc.* Several examples of the absorbed power at the surface for different depth have been examined. It has also been demonstrated that the power absorption pattern does not only depend on the chemical structure of the target, but also its physical characteristics and location [DiMarzio, *et. al.*, Sept. 1998].

Table 1. Specifications of the target, ground, and air.

Distance of target from the right edge of the computational grid.	8.85cm (≈3.48inches)
Depth of the target from the top of the computational grid.	3cm (≈1.18inches)
Size of the target	7.5cm×2.5cm (≈2.95inches×0.98inches)
Relative dielectric constant of the target	2.9
Relative permeability constant of the target	1
Conductivity of the target	1×10 ⁻³ S/m.
Relative dielectric constant of the ground	6.5
Relative permeability constant of the ground	1
Conductivity of the ground	7×10 ⁻² S/m.
Relative dielectric constant of the air	1
Relative permeability constant of the air	1
Conductivity of the air	≈0

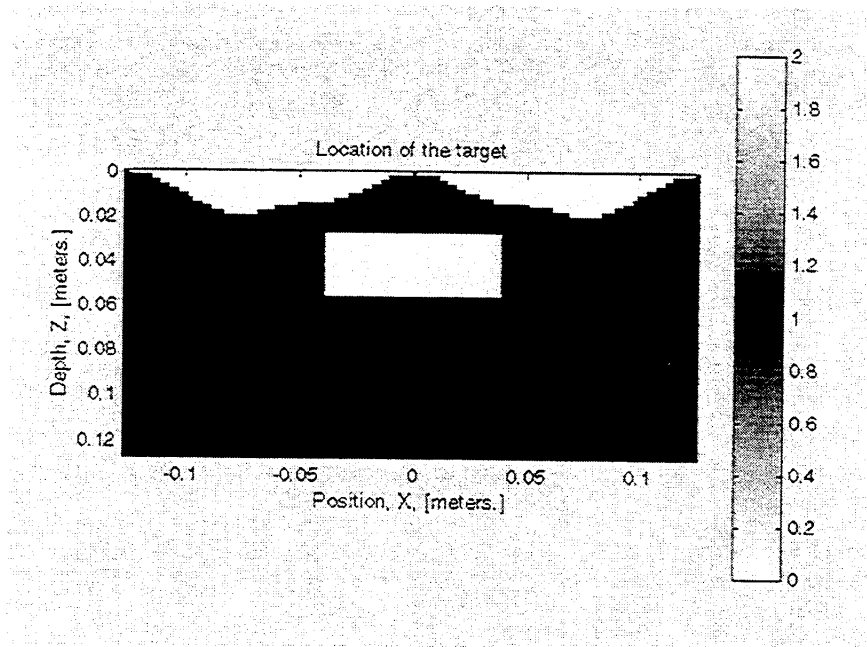


Figure 2. Sample grid for thermal model.

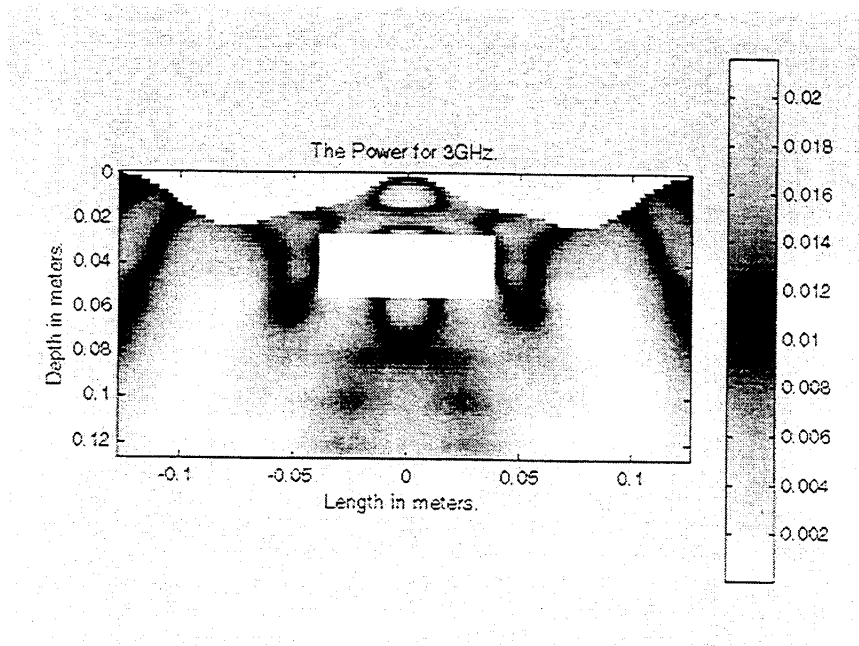


Figure 3. Absorbed power density pattern at 3GHz.

Here, how the shape of the rough surface affects the power absorption pattern has been examined. The surface, which is going to be investigated, could be either a periodic rough surface or a random rough surface. For the time being, periodic rough surfaces having sinusoidal harmonics with different ripple factors have been implemented. The function shaping the surface is given below:

$$f(x) = D - \delta \sin\left(\frac{\pi}{W}x\right) - \delta \sin^2\left(\frac{2\pi}{W}x\right) - \delta \sin^3\left(\frac{3\pi}{W}X\right), \quad (2)$$

where D , W , and δ determine the depth, the width of the grid, and the ripple factor respectively. Figure-4 shows the power absorption at the surface of a target with several different surface shapes.

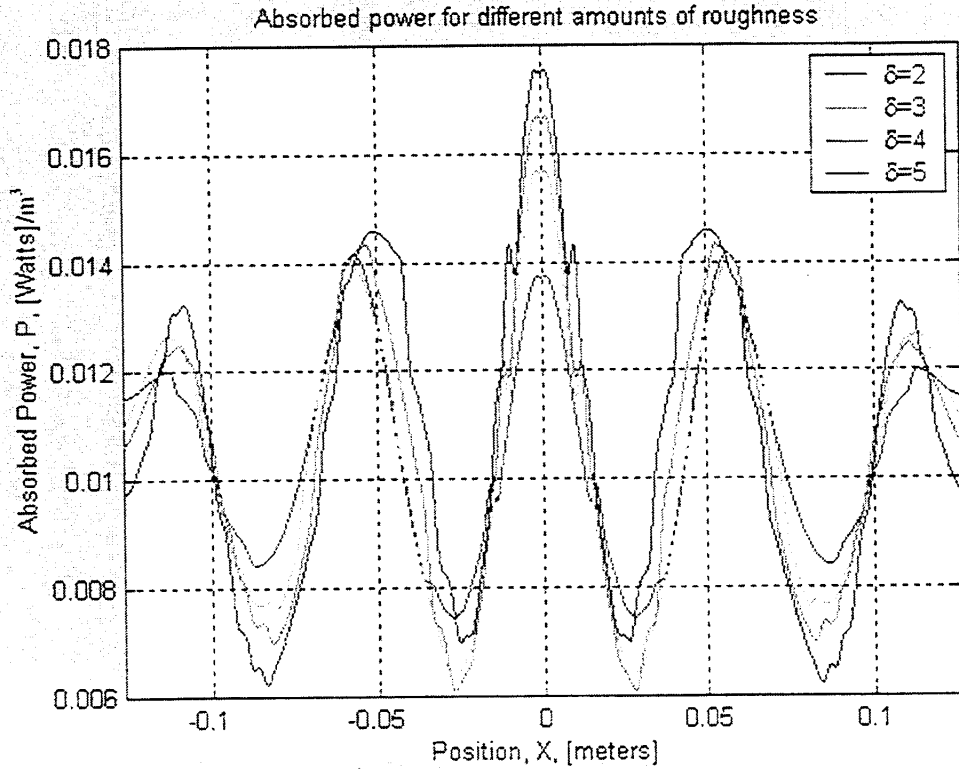


Figure 4. Absorbed Powers at 3GHz. with different roughness amounts.

2.2. Thermal Model

A two-dimensional thermal diffusion model was implemented under the excitation of a microwave power source. Actually, the microwave absorbed power density obtained as a result of the absorption model is used as a power source. Like in the absorption model, the ground and the target buried under the ground are described by their geometric properties. The thermal diffusion equation to be solved in the identified two-dimensional computational grid is shown below:

$$T_{ij}^{n+1} = R(T_{i+1,j}^n + T_{i-1,j}^n + T_{i,j+1}^n + T_{i,j-1}^n) + (1 - 4R)T_{ij}^n, \quad (3)$$

where R is

$$R = D \frac{\Delta t}{(\Delta x)^2} \quad (4)$$

D is

$$D = \frac{k}{\rho c} \quad (5)$$

where k , ρ , and c are thermal conductivity, density, and specific heat of medium respectively. Δt and Δx represent the time and grid spacing. The temperature quantity T_{ij}^n in equation (3) is computed from information at points $(i+1, j)$, $(i-1, j)$, $(i, j+1)$, $(i, j-1)$, and (i, j) at time n shown in Figure-5.

Constant-temperature boundary conditions are applied at the grid boundaries and a convective boundary condition reference is used at the interface between air and ground. The microwave source is kept on for a while and then shut down. All temperature increases and decreases, belonging to each specific time, are calculated during the heating and cooling cycles. The time history diagram, the temperature function of transverse distance and time, is built up.

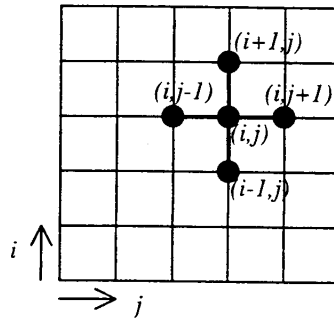


Figure 5. Computational grid for thermal model.

3. THE INVESTIGATION OF TIME HISTORY DIAGRAMS BELONGING TO ROUGH SURFACES WITH DIFFERENT RIPPLE FACTORS

As stated in section 2.2., rough surfaces with the target with the characteristics stated in Table-1 were implemented for the investigation. The investigation, which has been performed here, could be broken into two parts. In the first part, the amount of roughness, which was introduced into the surface shaping function mentioned in equation (2) by various ripple factors, 2, 3, 4, and 5 has been examined at only 3GHz. Furthermore, the effects of roughness on time history diagrams have been demonstrated in Figure-6. In the second group, the effects of frequency change have been investigated. The surface roughness applied here is unique; *i.e.* the study has been performed under the same rough surface. The ripple factor used in this simulation is 5. Frequencies are 0.5, 0.75, 1.0, 2.0, 2.5, and 3.0GHz. respectively. The results are presented in Figure-7. Both simulations were carried out under the assumption of zero angle incidences. The time history diagrams are also can be viewed from Figure-8 for frequency values of 0.5, 1, 2.5, and 3GHz.

4. CONCLUSIONS AND FUTURE WORKS

As can be seen from Figure-6, the time history diagrams look like very similar to each other. However, as inspected closer and in more detail, one can realize that the five distinct hot regions get narrow with increasing ripple factor δ . As a result of this, the gaps between these distinct hot regions can be distinguished more easily for the surfaces with higher roughness amounts than the surfaces that are relatively smooth.

From figure-7, one can draw several conclusions. It is useful to separate to have the images with two different groups: Low frequencies and high frequencies.

For low frequencies, one can easily observe that there is an obvious temperature rise in the middle of the location for the same surface shape, same soil and mine characteristics, and same mine location. As frequency increases, this temperature rise increases in the middle. On the contrary, the heating of the sides disappears. The temperature increases become concentrated in the middle portions.

For higher frequencies, the situation is a little bit different. As frequency increases, side-bands get more distinct. However, when the temperature rise in the middle is viewed, it is impossible to determine the presence of a mine. Overall, it is difficult to sort out temperature variations due to buried objects and those due to surface roughness. Temperature rise in the middle is sometimes high and sometimes low, due to the effects of electromagnetic interference of incident, transmitted, and scattered fields. In any case, the presence of the mine is evident.

We plan to investigate the frequency behaviors of these time history diagrams and to look at the changes in the heating pattern with frequency in the presence and absence of a mine. We believe this will yield new information, and reduce the effects of surface roughness.

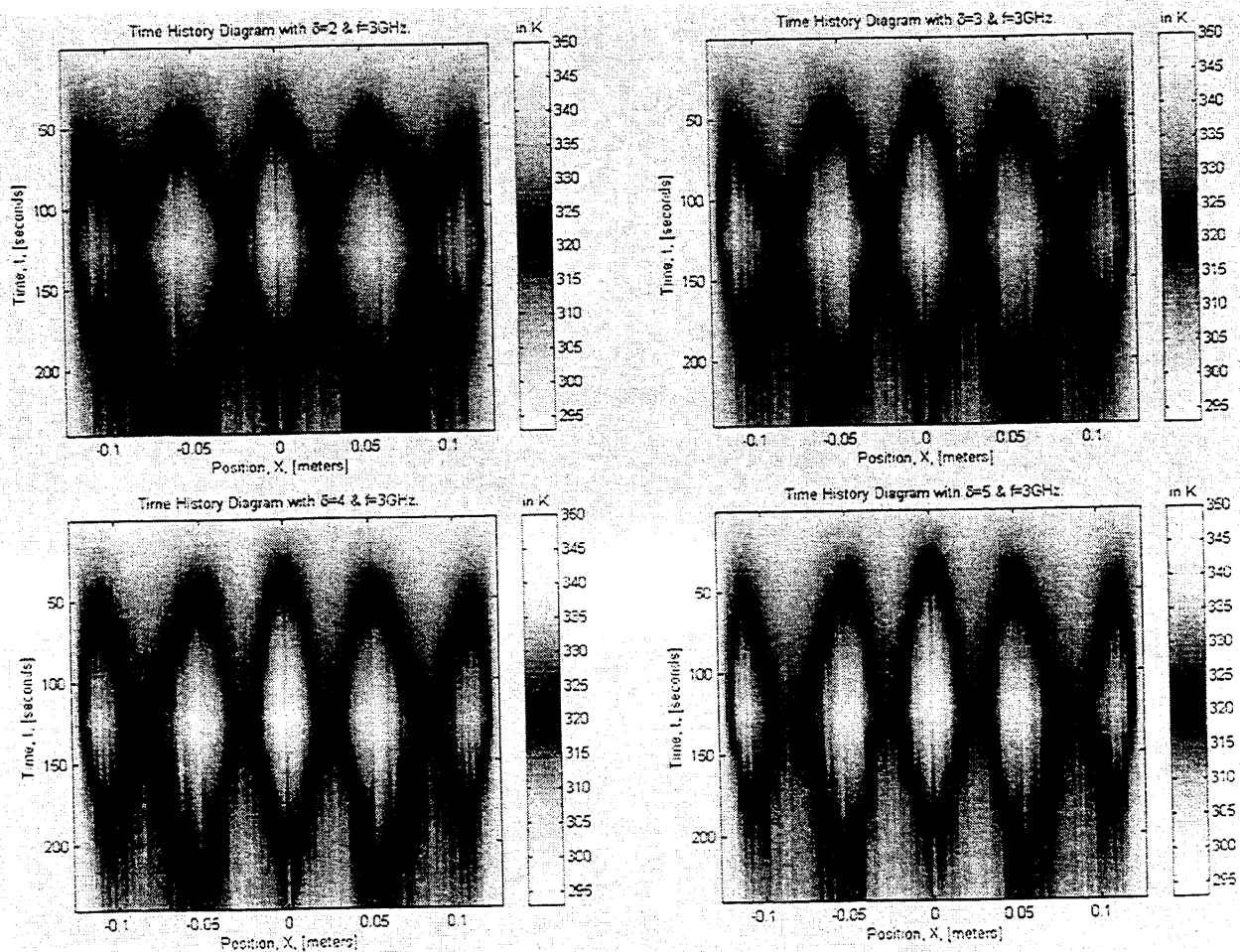


Figure 6. Time History Diagrams at 3GHz. with different roughness amounts.

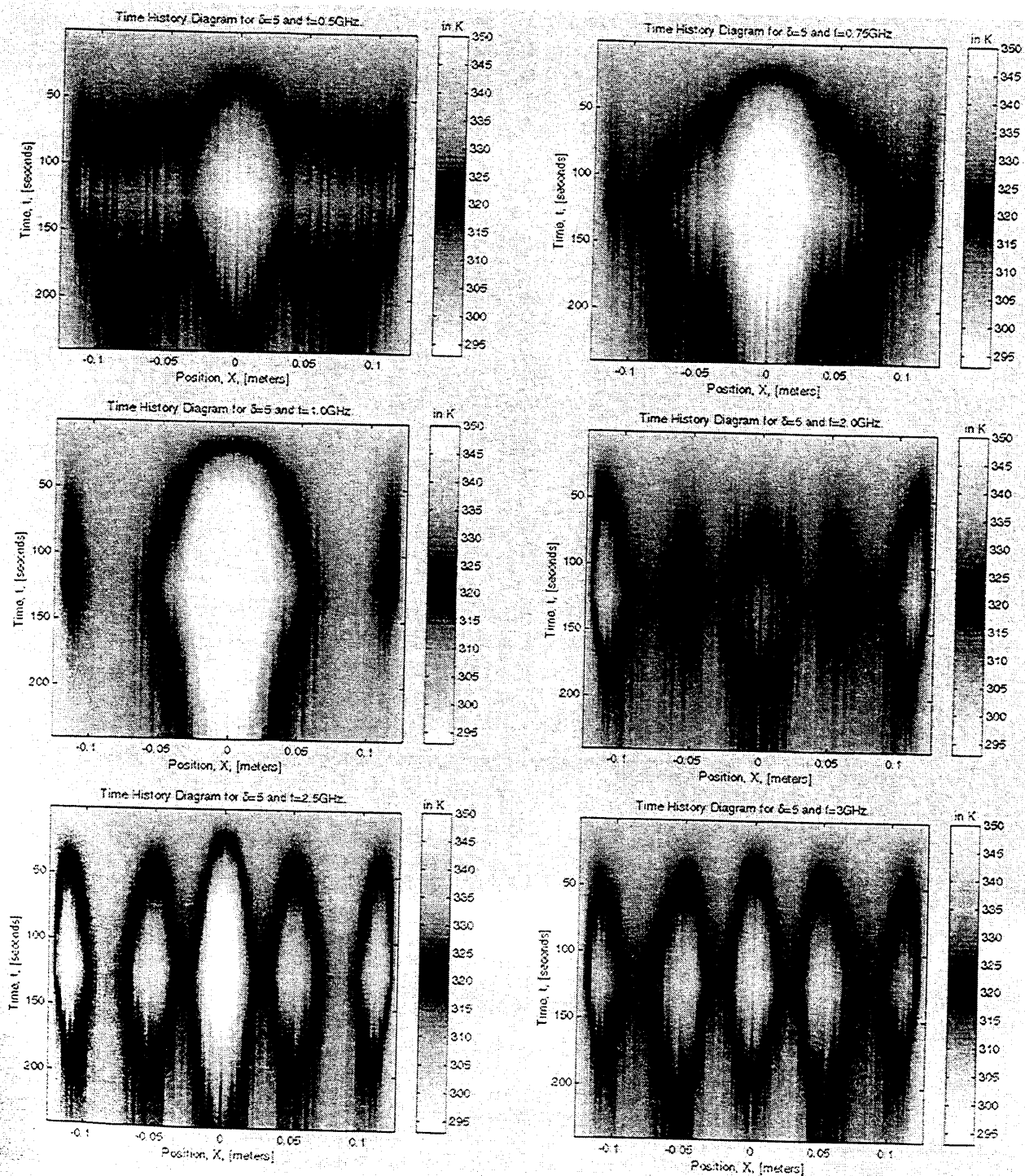


Figure 7. Time History Diagrams at different frequencies with same amount of roughness.

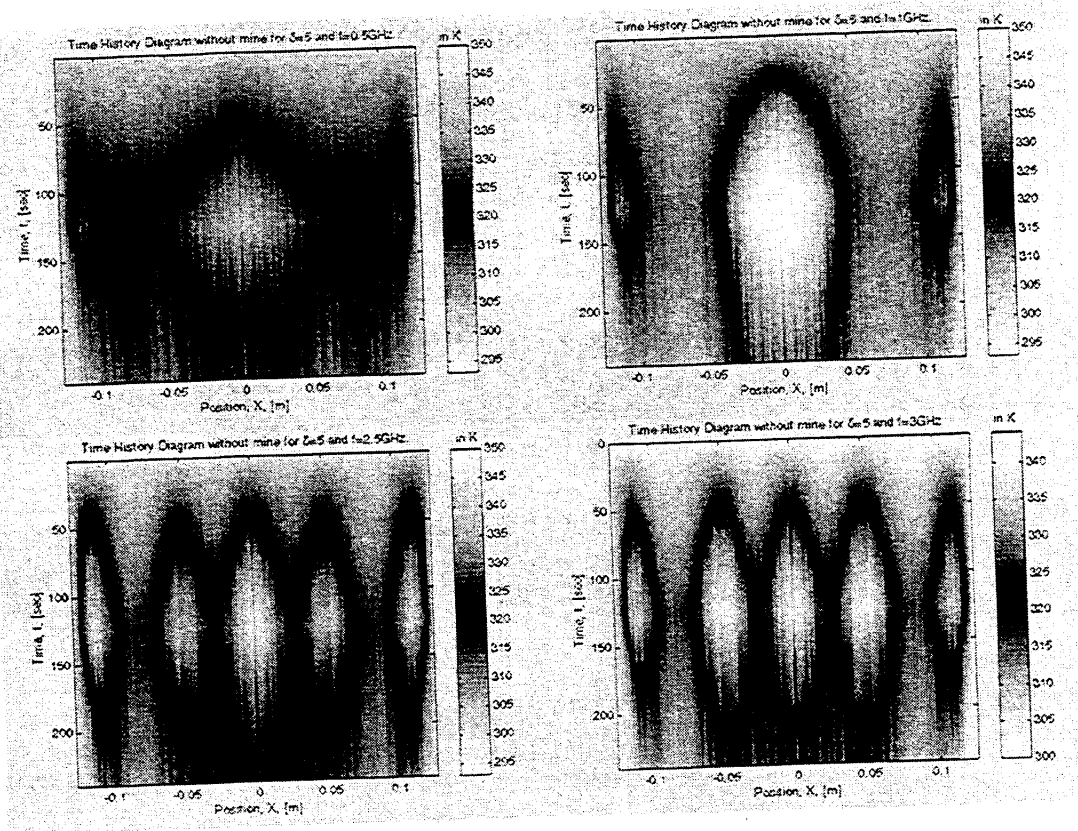


Figure 8. Time History Diagrams without mine at different frequencies with same amount of roughness.

ACKNOWLEDGMENTS

The OSD MURI Program under Army Research Office Grant number DAAG55-97-1-0013 sponsored this work.

REFERENCES

1. P. Li, A. Maad, F. Moshary, M.F. Arend, and S. Ahmed, "Infrared Imaging of Buried Objects by Thermal Step-Function Excitations," *Applied Optics* **34**, No. 25, pp. 5809-5816, 1 September 1995.
2. A. C. DiMarzio, D. O. Hogenboom, C. M. Rappaport, and G. O. Sauermann, "Microwave-Enhanced Infrared Thermography," *Detection and Remediation Technologies for Mines and Minelike Targets III*, *Proc. SPIE* **3392**, pp. 1103-1110, April 1998.
3. A. C. DiMarzio, C. M. Rappaport, W. Li, M. Kilmer, G. O. Sauermann, "Microwave-Enhanced Infrared Thermography," *Environmental Monitoring and Remediation Technologies*, *Proc. SPIE* **3534**, PP.337-342, November 1998.
4. A. C. DiMarzio, C. M. Rappaport, W. Li, G. O. Sauermann, "Microwave-Enhanced Infrared Thermography," *Detection and Remediation Technologies for Mines and Minelike Targets IV*, *Proc. SPIE* **3710**, pp. 173-179, April 1999.

Dual-frequency microwave-enhanced infrared thermography

Tianchen Shi^a, Gerhard O. Sauermann^a, Carey M. Rappaport^b, Charles A. DiMarzio^a

CENSSIS, the Center for Subsurface Sensing and Imaging Systems
Dept. of Electr. & Comp. Engineering, Northeastern University, Boston MA 02115, USA
^aOptical Science Laboratory; ^bCER, Center for Electromagnetic Research

Abstract

Humanitarian landmine detection and clearance is one of the most challenging, difficult and time-consuming tasks to be completed with existing technologies. Infrared (IR) Imagery has been used to find differences in heat transfer on the surface of the soil due to a buried object. In this paper, we will describe a method, *Dual Frequency Microwave Enhanced Infrared Thermography (MEIT)*. Heating with microwaves instead of natural sunlight leads to a number of advantages, such as more efficient heating to enhance the thermal signature, and the ability to sense electromagnetic as well as thermal properties of the buried object. However, like other IR techniques, it is limited by surface roughness. Thus, the two frequency technique is used to minimize the clutter introduced by the rough, irregular surface of the ground itself, and vegetation covering the ground. The dependence of scattered waves on frequency is weak enough to make this possible. A 2-D computational model of this method has been developed to simulate real-world landmine detection. Moreover, ROC (Receiver Operating Characteristic) curves are used to evaluate the performance of the system applying this method.

Keywords: Landmine detection, rough surface, infrared (IR) imagery, thermography, microwave enhanced, dual frequency, ROC curves.

I. INTRODUCTION

We have previously presented analytical and experimental results on Microwave-Enhanced Infrared Thermography. In [1] [2], we have shown that using microwave heating of the ground can produce strong enhancements of the infrared signature of the shallow-buried objects, while diurnal variation of the heating pattern, caused by natural sunlight, is much weaker. Moreover, we have described how different signatures impart electromagnetic and thermal information in [3].

In all the above work, analysis and results are based on an ideal uniform interface at air-ground boundary, say, a smooth surface. Furthermore, we have shown that the effects of surface roughness produce clutter, which masks the signals from buried objects in [4] [5]. As surface roughness increases, infrared signatures of buried objects become more unreadable and have an increasing probability of producing false alarms due to hot spots on the surface generated by local variation in surface orientation.

For detecting landmines and discriminating between landmines and other buried objects, this effect also increases the difficulty of the following image-processing procedure to extract information out of the obtained heating pattern.

In this paper, we present a modified MEIT technique, *Dual-Frequency MEIT*, which is developed based on previous work. In this new method, microwave sources with two different frequencies are used in two consecutive heating cycles on the ground. Then we obtain difference between two infrared signatures corresponding to the different frequencies, assuming that all the conditions of environment and of scatter have not been changed within the period between the two separated heating cycles.

In Section II & III, we will show that the signature of difference gives information about buried objects with reduced clutter, which mainly comes from rough surface and vegetation covering the surface. To validate the efficiency of this new method, at the end of this paper (Section IV), we have calculated ROC (Receiver Operating Characteristic) curves for both single frequency and dual-frequency technologies. The ROC curves indicate an overall improvement.

II. THE DUAL FREQUENCY TECHNIQUE V.S. THE SINGLE FREQUENCY TECHNIQUE

Surface roughness is the major factor limiting the MEIT technique in our recent research. As we discussed in [3], wavelength, polarization of the incident waves, as well as properties of the soil all have some influence on enhancement, while angle of incidence, physical structure and chemical composition of buried object are more important.

However, all the above discussion had assumed a flat, smooth surface. Based on some of our real world experiments, in comparison to those in the laboratory on a flat surface, we find that a rough surface contributes most to the distortion of infrared signature at the air-ground boundary, compared with all the other parameters. Thus, eliminating the clutter due to surface roughness becomes the first thing to be considered to ensure microwave enhancement.

We have noticed the fact that field irradiation under the surface depends most on Fresnel reflection and surface geometry and dielectric constants, and less on the wavelength of the incident wave. In our microwave model, computational simulation has also confirmed this idea.

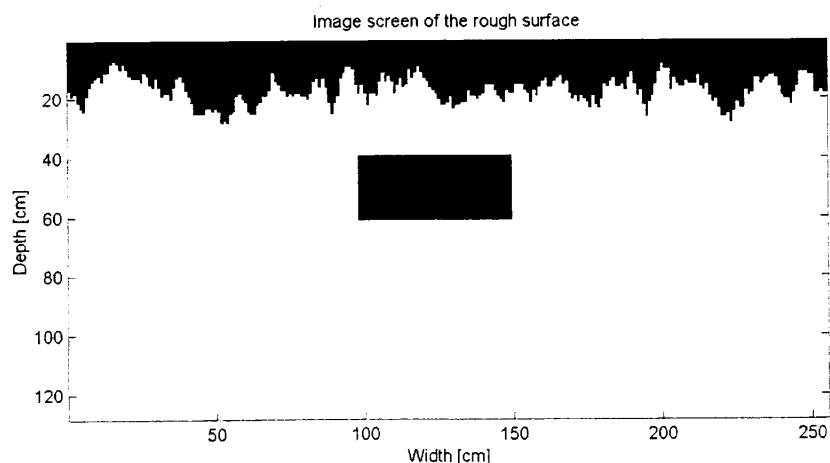


Figure 2.1, Geometry of computational model, rough surface and the buried object

Figure 2.1 shows the geometry of the computational model we used in this paper, and Figure 2.2 explicitly shows that incident microwave power with different wavelengths produce, predictably, almost the same power distribution, as well as electromagnetic field distribution on the same rough surface. We explain this phenomenon as follows: no matter what wavelength is used for incident waves, absorbed power density at the surface mainly focuses at tips of the rough surface, because a unit volume of soil at tips has more effective area for absorption than that of other region.

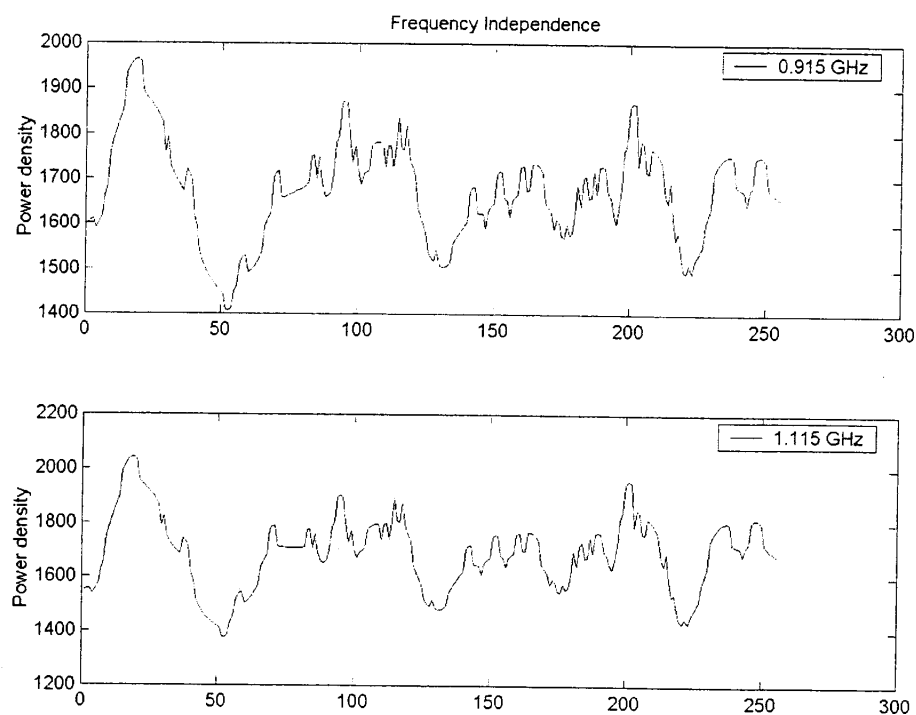


Figure 2.2, Frequency independence of scattered wave, power density distribution at the surface under two different frequencies

Next, we consider how this simulation reduces the surface clutter. We know that using microwave heating instead of natural sunlight has many advantages, one of which is that we can take microwave wavelength as an adjustable parameter in the MEIT technique. By subtracting the two infrared signatures obtained under two different frequencies, one can finally achieve a signature of the buried objects, with clutter introduced by rough surface greatly reduced. Then we can more easily extract information about the buried objects under the ground from the thermal signature data. In the following section, we present some results obtained by a 2-D computational model based on the idea above.

III. COMPUTATIONAL RESULTS

Now it is time for us to use the 2-D numerical model developed according to the above analysis to simulate what happens during the period of a microwave heating cycle, which includes heat-up and cooling down periods. As described in our previous paper [1][2], the numerical model includes two main parts, the microwave model and the thermal model. The microwave model calculates total electromagnetic field including undisturbed incident plain wave plus scattered waves to get the power density distribution all over the under ground region. Then, the thermal model applies a thermal diffusion equation to calculate the temperature distribution variation on the surface as a function of time to form a THD (Time History Diagram) [5]. As presented in [3], initial signatures of microwave heating cycle show information about electromagnetic wave propagation in the vicinity of the object, and the purpose of the dual-frequency technique is mainly focused on improving the infrared signature contrast in this early heat-up cycle. Therefore, we only use the microwave model here to obtain the absorbed power density distribution. Some results obtained from the 2-D FDFD numerical microwave model simulations are shown below.

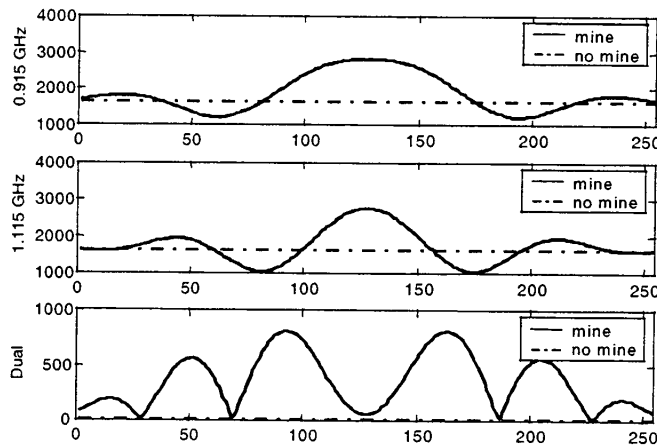


Figure 3.1 ($\sigma=0.00\text{mm}$)

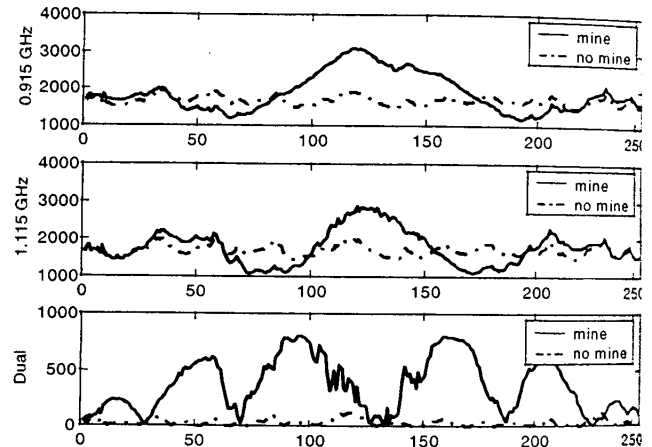


Figure 3.2 ($\sigma=2.21\text{mm}$)

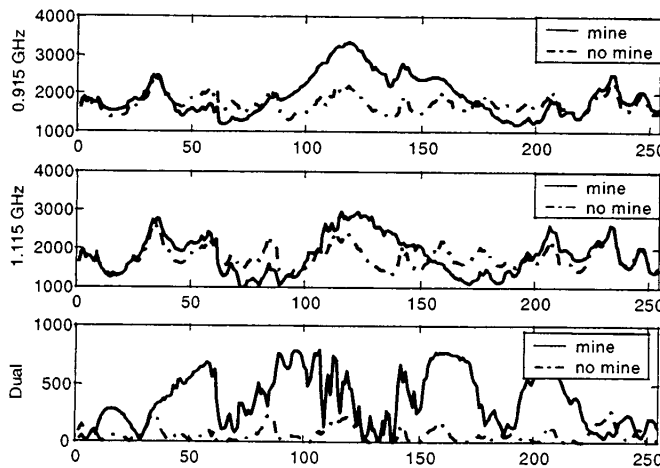


Figure 3.3 ($\sigma=4.42\text{mm}$)

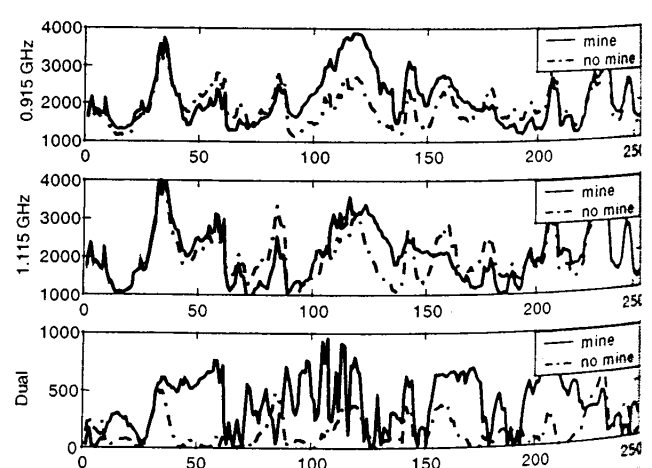


Figure 3.4 ($\sigma=8.84\text{mm}$)

In figures 3.1 ~ 3.4, power density distributions on the top of the surface in our 2-D geometry, under different surface roughness, are shown in cases of frequency 0.915Ghz, frequency 1.115Ghz, and the corresponding dual-frequency, from top to the bottom respectively.

All the figures are obtained with the same shape of rough surface, and different extent of roughness for the same surface profile is represented by factor k ($0 < k < 1$). For example, when k is equal to zero, surface becomes flat, and when k is equal to one, the standard deviation of the surface height is 1.15cm in the above case.

From these figures, we find out that using the Dual Frequency technique efficiently reduces the clutter of rough surface comparing with power density distributions for the single frequency case. The rougher the surface becomes, the worse the signature contrast becomes for single frequency, while for the Dual Frequency case power density distribution remains almost the same. Thus, the Dual Frequency technique successfully extracts useful information of buried objects out of the clutter caused by surface roughness. Similar results are also obtained with other different rough surface profile.

One more thing we want to point out is, however, we cannot totally eliminate the distortion of heating pattern. Basically, there always exists some extent of dependence on wavelength for irradiation field pattern, which increase will increase difference between two frequencies. On the other hand, it is obvious that too close frequency difference yields a small signal as well as low clutter. So we have to choose the proper frequency pair to get as much enhancement as possible. The frequency pair we have used in this simulation has been optimized. To get optimized frequency pair, we have searched in a large range of the frequency domain to compare maximum signature difference between mine-free and mine case.

Furthermore, in the following section, we have also used another means, ROC curves, to validate the dual-frequency technique's efficiency at removing or minimizing clutter caused by surface's irregularities.

IV. ROC CURVES

To evaluate the performance of a certain detection system, ROC curves provide much help to understand overall detection statistics of the system and compare systems when a priori probability and cost function are not available.

For a single measurement, four results are possible: a correct detection, missed detection, correct negative, or false alarm. The distribution of three results depends on the distributions of signals where an object is present and where it is not. We assume that both distributions are Gaussian. In the present case, the widths of both distributions are determined by a combination of a fixed number and the standard deviation of the signal in the absence of a mine.

For an ideal detection system, these two distributions are sufficiently separated to ensure that true signals are always detected and that the probability of false alarms is zero. However probability distributions of real system always have some extent of overlap, shown as Figure 4.1 below. For any given threshold value of detection, the areas under each Gaussian curve to the right side of threshold are the probabilities of detection and false alarm, respectively.

To get ROC curves, we scan the threshold value from negative infinite to positive infinite, and calculate the corresponding probability for detection and false alarm, i.e., P_d and P_f . In fact, ROC curves are curves for P_d versus P_f parametrically depending on threshold value of system. Typical ROC curve looks like Figure 4.2. The more area under the curve is, the better overall performance of the detection system will be. The ideal system ROC curve is a straight line from

origin (0,0) to point (0,1), then to (1,1), and the straight line from origin (0,0) to point (1,1), as shown in figure 4.2, is as bad as random choice.

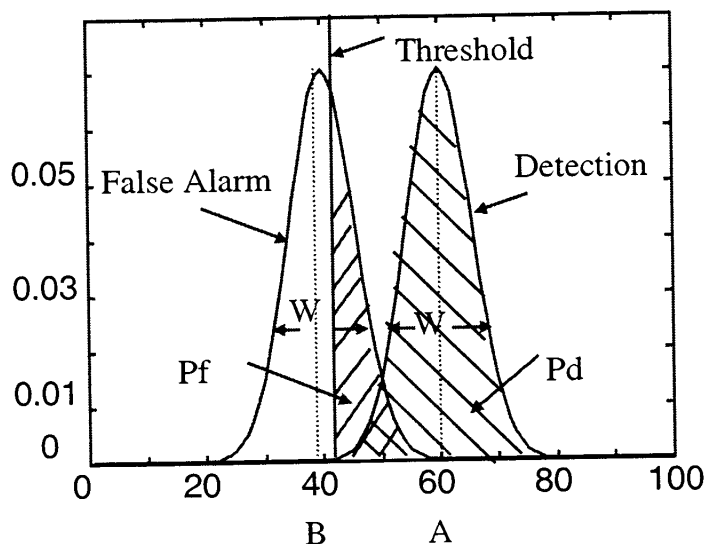


Figure 4.1 Preliminary of ROC Curves

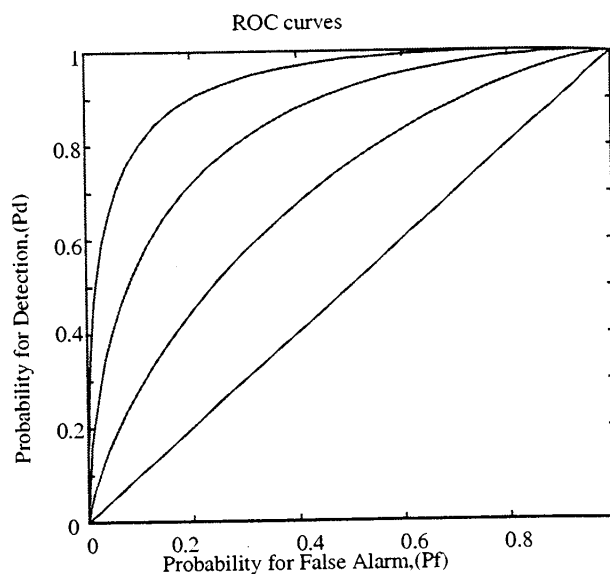


Figure 4.2 Typical ROC curves

In our analysis, signal and clutter, mine and no mine present respectively, are both assumed to be Gaussian. We know that the Gaussian curves can be totally determined by two parameters, mean and standard deviation, where the latter is related to noise. When these two parameters are obtained from surface power density pattern, we can easily get ROC curves according to above definition. Our strategy to get ROC curves is that, first, we find the difference between maximum and minimum values of power density along rough surface in mine and no mine case, then use them as the means of Gaussian probability density distributions for true signal and false alarm, respectively. From our previous experiment data, the unpredictable behavior of the ground temperature variation, which is due to sunlight, is around a few Kelvins. So we take the minimum power density variation corresponding to 1 Kelvin as the standard deviation of Gaussian probability density distribution for both true and false alarm case. Then we add the standard deviation of the signal in the absence of a mine to indicate the clutter level.

We show ROC curves from simulation data as described above. As can be seen from Figures 4.3, with increasing surface roughness, ROC curves become worse and worse for the single frequency detection technique. The ROC curves for the dual-frequency technique become worse, though at a pace much slower than those of the single frequency technique.

However, as shown in Figure 4.3, when the surface roughness factor k is large enough, the dual-frequency technique begins to show its advantage over single frequency ones. From this time on, the single frequency technique tends to fail and the ROC curve suggests that the single frequency technique, in this case, is no better than a coin toss.

By using ROC curves, we validate that the dual-frequency detection technique has better performance when a target is under ground surface with irregularity.

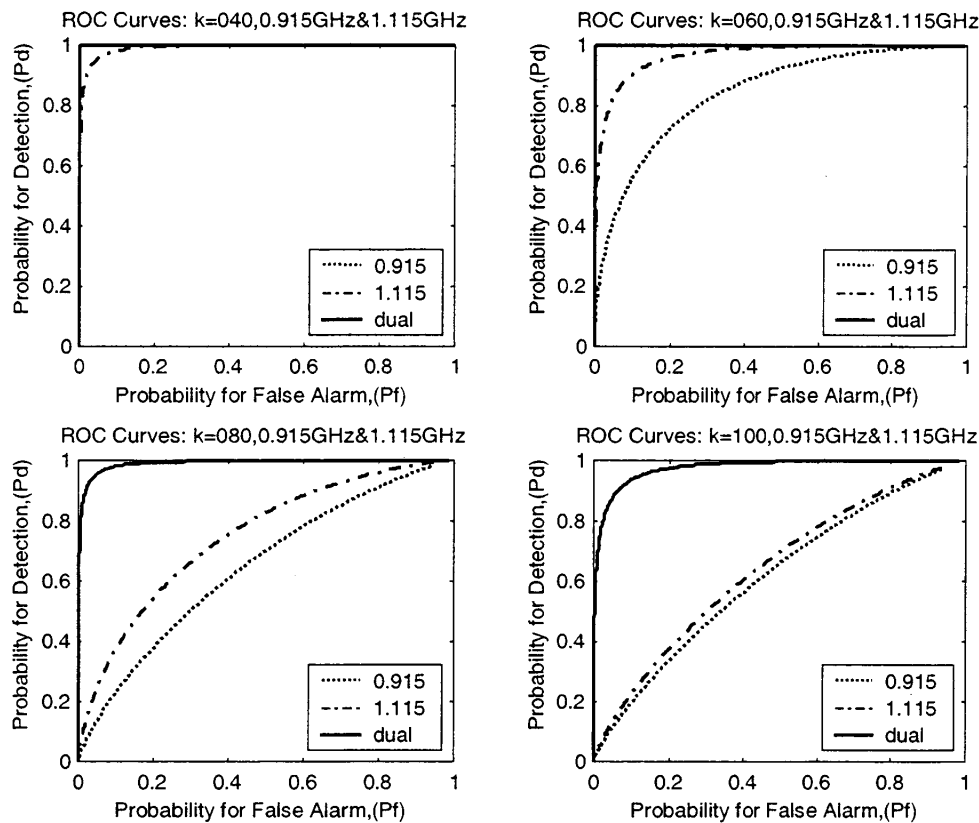


Figure 4.3 ROC curves for dual-frequency technique with increasing surface roughness factor k ($0 < k < 1$)

V. CONCLUSION AND FUTURE WORK

We have presented an improved MEIT technique for humanitarian landmine detection, *Dual-Frequency MEIT*. This technique mostly retains the advantages of original MEIT while it enhances us to observe an infrared signature of the shallow-buried object under rough surfaces. Then we have validated it by results obtained via numerical simulation and ROC curves. In the future work, we will account for two other essential factors, vegetation and variation in solar heating, both of which may mask the infrared signatures of the buried objects. By overcoming these difficulties, the MEIT technique will be a more valuable tool for detection of buried objects. Moreover, examining the thermal signature in later part of heating cycle will provide useful shape information about objects.

ACKNOWLEDGMENTS

The OSD MURI Program under Army Research Office Grant number DAAG55-97-1-0013 sponsors this work.

REFERENCES

1. Charles A. DiMarzio, Carey M. Rappaport, and Li Wen, "Microwave-Enhanced Infrared Thermography," Detection and Remediation Technologies for Mines and Minelike Targets III, SPIE Aerosense Volume 3392, pp. 1103-1110, April 1998.
2. Charles A. DiMarzio, Carey M. Rappaport, and Gerhard O. Sauermann, "Microwave-Enhanced Infrared Thermography," Industrial and Environmental Monitors and Biosensors, SPIE Volume 3534 and presented in Boston, MA, November 1998.
3. Charles A. DiMarzio, Carey M. Rappaport, L. Wen, and Gerhard O. Sauermann, "Microwave-Enhanced Infrared Thermography," Detection and Remediation Technologies, SPIE Volume 3710, pp. 173-179, April 1999.
4. Charles A. DiMarzio, T. Oktar, L. Wen, Carey M. Rappaport, "Detection of Objects Buried in Soil Using Microwave Heating," Environmental Monitoring and Remediation technologies II, Proceedings of SPIE, Volume 3853, pp. 311-320, September, 1999.
5. Taner Oktar, Carey M. Rappaport, Charles A. DiMarzio, "The Effects of Surface Roughness on Microwave Heating of Soil for Detection of Buried Landmines," Detection and Remediation Technologies for Mines and Minelike Targets V, Proceedings of SPIE, Volume 4038, April 2000.

Contact author: *tshi@ece.neu.edu; voice 1 617 373-8570, OSL, 334 Egan Research Center;
<http://www.censsis.neu.edu>; CENSSIS, 235 Forsyth, Northeastern University, Boston, MA 02115,

Toward a Laser-Based, Non-Contact Acoustic Landmine Imager

by

Charles A. DiMarzio, Wen Li

Center for Electromagnetics Research, Northeastern University
Boston, Massachusetts 02115

Lawrence J. Berg, Laser Science Inc.
8E Forge Parkway, Franklin MA 02038

James Sabatier

The National Center for Physical Acoustics
1 Coliseum Drive, University, MS 38677

ABSTRACT

Acoustic sensing shows promise for the detection of buried landmines. One of us has previously demonstrated successful imaging of mine simulants buried at depths from the surface to 15 centimeters, using speakers and a laser vibrometer, which collects spectral data at low frequencies. The strength of the method is in the contrast between the porous soil and the nonporous mine, while the limitations are the strong attenuation of the probing acoustic wave and coupling of the sound directly into the vibrometer.

Another member of our group has shown that shallow-buried objects can be detected by acoustic pulses generated by a high-power pulsed laser and a microphone, producing signals which are processed in the time domain to observe echos from buried objects. The strengths of this method are the lack of direct acoustic coupling into the vibrometer and the resolution implicit in the high frequencies. The limitations are in the depth of penetration, the large size of the laser, and the low acoustic energy.

We have recently performed an experiment which combines the best features of both techniques. A new, portable, pulsed laser and the laser vibrometer were used at an outdoor test site to show that (1) sufficient laser energy can be coupled into the ground to produce signals larger than the ambient seismic noise, and (2) these signals are modified by buried objects. We will show some examples and consider the configuration for an operational system.

Key Words: Photoacoustic imaging. Landmine detection.

1. INTRODUCTION

It is widely recognized that successful humanitarian demining will require many sensors to achieve the combination of certain detection required to return land to civilian use and low false alarm rate to allow the removal process to proceed at an acceptable rate. One interesting approach is the use of acoustic sensing. One of the authors of this work has previously demonstrated successful imaging of mine simulants buried at depths from the surface to 15 centimeters, using commercially available speakers and a laser Doppler vibrometer. The speakers insonify a large area while the vibrometer produces an image by scanning the region of interest, collecting spectral data in a bandwidth from hundreds to thousands of Hertz, at each point. The strength of the method is in the contrast between the porous soil and the nonporous mine, while the limitations are the strong attenuation of the probing acoustic wave and coupling of the sound directly into the vibrometer.

Another member of our group has shown that shallow-buried objects can be detected by acoustic pulses generated by a high-power pulsed laser. In this case, the pulsed laser source is scanned across the target area and a microphone records signals, which are processed in the time domain. The signals show an acoustic pulse from the surface and the echo of the pulse from the buried object. Ten microseconds corresponds to a depth of about 1.5 centimeters. The strengths of this method are the lack of direct acoustic coupling into the vibrometer and the resolution implicit in the high frequencies. The limitations are in the depth of penetration, the large size of the laser, and the low acoustic energy.

We have recently performed an experiment which combines the best features of both techniques. A new, portable, pulsed laser and the laser vibrometer were used at an outdoor test site to show that (1) sufficient laser energy can be coupled into the ground to produce signals larger than the ambient seismic noise, and (2) these signals are modified by buried objects. We will show some examples and consider the configuration for an operational system.

Here we explore an all-optical implementation of an acoustic landmine detection system. We will present a brief overview of the advantages of an all-optical implementation, followed by a study of the mechanisms by which light generates sound, and a summary of our experiments.

2. USE OF LIGHT IN ACOUSTIC IMAGING

The all-optical implementation of acoustic imaging has two parts. In the first, a pulse of light generates sound in the ground, and in the second, a continuous wave (CW) laser vibrometer measures surface motion to measure the reflected sound from the object. Here we compare this approach to conventional all-acoustic techniques.

If sound is generated in the air, and transmitted into the soil, it is subject to severe losses caused by reflection at the soil-air interface. These losses are related to the variations in the speed of sound, and density, as well as the angle of incidence of the sound. Typical reflections can exceed 98%, with the result that an incident power level of 200 Watts in air can lead to peak pressures of as little as 200 Pascals in the soil. In contrast, if a pulse of laser light is absorbed in the ground, the resulting sound wave does not need to travel through the interface, and is thus not subject to loss. For a 150 millijoule laser pulse, our simulations indicate that a pressure of 10^5 Pascals can be generated. Thus, laser generation of sound is attractive because the sound can be generated directly in the soil without incurring the usual reflection loss.

On the detection end of the process, a similar or worse situation arises on reflection from the interface. Specifically, if the sound reflected from the buried object is scattered over a wide angle, large portions will be reflected much more strongly, and thus the transmission can be as low as 0.1%, to a microphone above the ground. The possibility of detecting such small signals in the presence of the larger ones produced by the source, or even in the presence of noise, is limited. On the other hand, surface motions related to buried objects are of the order of fractions of a micrometer, which are easily detected by laser vibrometry.

Thus, an all-optical implementation seems desirable because of the opportunity to maximize signal strength and, perhaps more importantly, because of the ability to avoid coupling of the receiver to the transmitter directly through the air.

The laser vibrometer has been used successfully in work reported elsewhere in this volume. The issues that remain to be discussed are the generation of the sound by the laser pulse and the propagation of the waves which are generated.

3. MECHANISMS OF SOUND GENERATION

We consider four different mechanisms for sound generation, and we find that at least two of them are likely candidates in different situations. In the first and most obvious, the soil is heated, and expands, thereby producing a compressional wave. This was our first assumption and is consistent with our first observations, reported previously [Li, *et. al.*, McKnight, *et. al.*, 1999], for which imaging has been demonstrated [Witten *et. al.*].

The second potential mechanism is radiation pressure, in which photon momentum is transferred to the soil. Numerical calculations have shown that the resulting impulse is much less than that produced by heating, and this option has been dismissed.

The third option is that the air between the sand particles is heated, resulting in a very different form of acoustic wave, specifically, the so-called Biot Slow Wave. This appears to be somewhat consistent with the results reported in this volume McKnight *et.*

al., 2000, when the laser beam is tightly focused. The other alternative is that the laser beam produces a plasma in the air, the expansion of which produces the acoustic pulse. As reported in by McKnight (2000), this is the more likely choice; We have observed "white" light emitted from the region of the laser spot in the case of a tightly focused beam, and upon spectroscopic analysis, this light has distinct lines associated with atmospheric gasses, as would a plasma. If the process were heating alone, the spectrum would be the smooth spectrum of black-body radiation.

We thus conclude that sound is generated by heating of the sand in the case of a broad laser beam (150 mJ in a 1-cm spot), and by plasma in the case of a tightly focused beam (150 or 18 mJ in a 1-mm spot). We have developed a model which predicts the performance of the broad beam, and are working to develop one for the focused beam. Although the sound of the plasma pulse is more easily audible to observers in the laboratory, it is not immediately clear that it produces a larger sound pulse in the ground, nor that this pulse propagates in such a way as to detect buried objects. Indeed some of our experiments suggest the opposite. We repeated the experiment discussed in Witten's paper mentioned above, but using a focused beam. Results were significantly worse, and it was not possible to observe the shape of objects. This result is not understood yet, and is the subject of further research.

4. EXPERIMENTS

To determine the feasibility of the all-optical approach, we conducted two days (28–29 July 1999) of experiments at the University of Mississippi mine lanes, using a portable pulsed CO₂ laser (Laser Science, Inc., Franklin, MA) with a pulse energy of 18 mJ, and a pulse length of about 100 nsec, as a source, and a scanning HeNe laser vibrometer (Polytec, PI, Auburn, MA) as a detector. The work was conducted in a small patch of ground over an anti-tank mine casing buried two inches deep. We focused the laser with a short focal length lens, so that a flash of visible light was seen when the pulse occurred.

The source laser was focused to nine positions, spaced about 25 mm apart, and the vibrometer collected data at six points near each source spot. If the spots overlapped closely, the vibrometer signal was contaminated by ejection of soil particles, and if it was too far away, the signals were not strong enough to detect with the particular configuration of the hardware used. A typical set of results is shown in Figure 1.

The two panels on the left show Fourier transforms of the displacement signals calculated by the vibrometer. The top shows two curves, one collected with the laser on, and the other, for reference, collected without the laser, and thus indicating the ambient motion of the ground. The lower panel shows the ratio of the signal to the background. This is the data set normally recorded by the vibrometer. On the right, we have inverse Fourier transformed this data to produce displacement, and then differentiated numerically to obtain velocity, both as functions of time. We note that in the spectral plots, there is

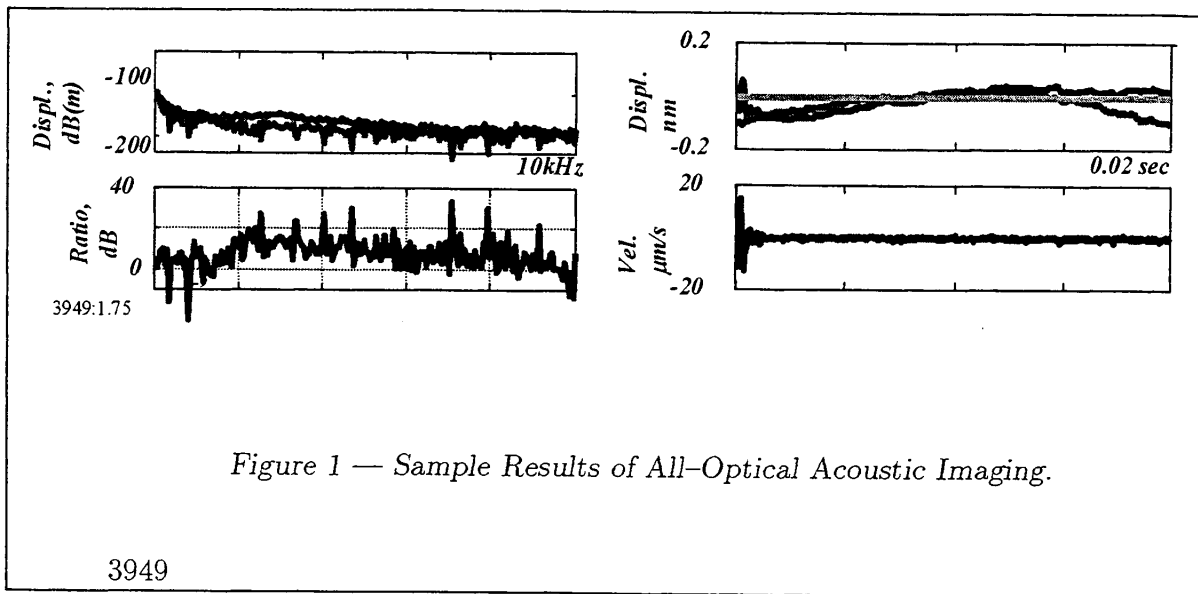


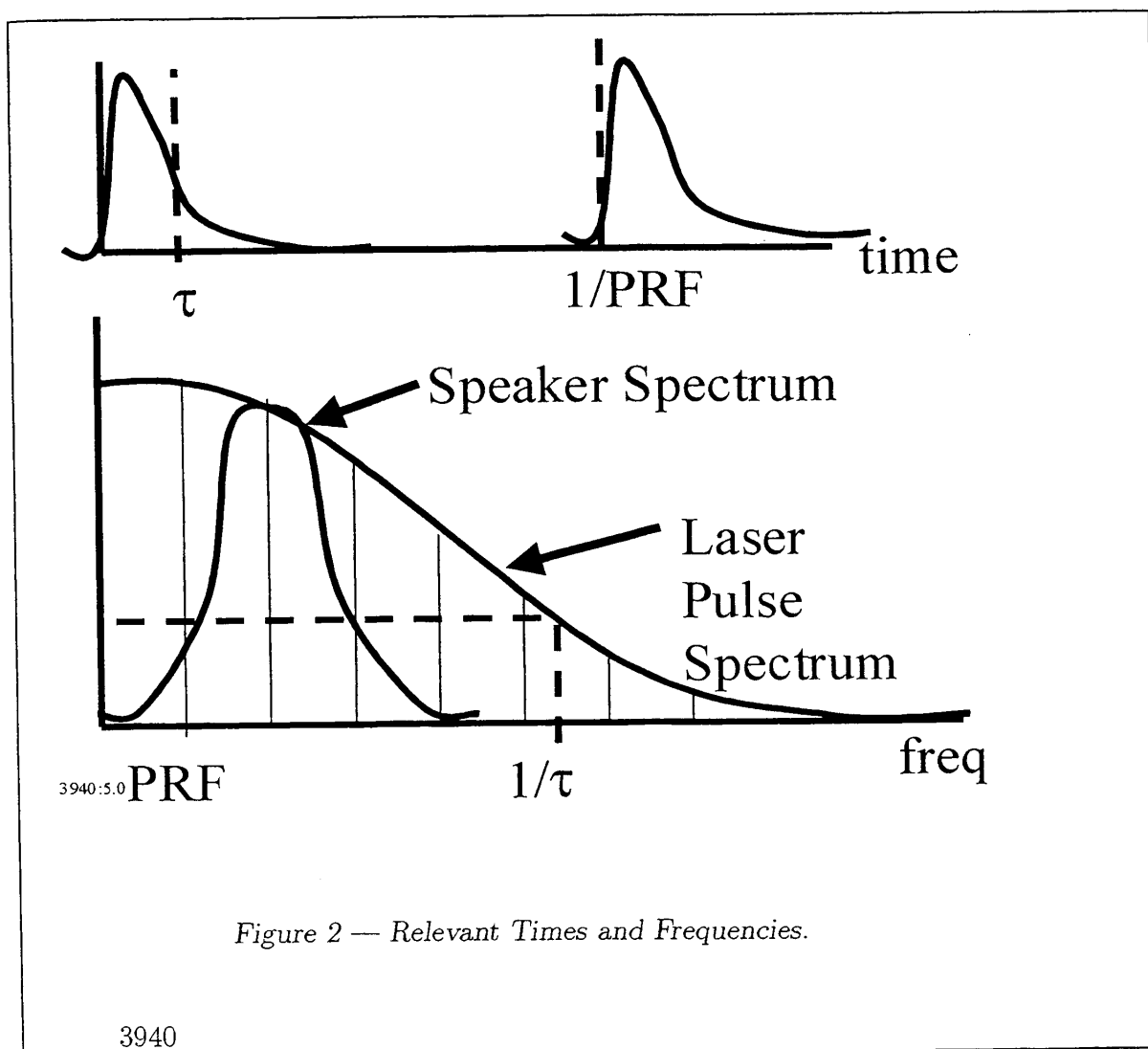
Figure 1 — Sample Results of All-Optical Acoustic Imaging.

a positive ratio at all frequencies, indicating that the ground surface motion resulting from the laser pulse exceeds the background acoustic activity, and does so by a substantial amount at most frequencies. In the time plots, the initial pulse is the most striking feature.

These plots illustrate the differences between the two types of acoustic imaging being performed by our individual groups. Using the speakers as sources, the most effective frequencies seem to be low, in the range of hundreds of Hertz, where the penetration of the ground is good. The successful mine detection work relies on spectral information in this range. The sound source is adjusted to produce mostly these low frequencies. In the laser-induced acoustic work, the energy is distributed across a much wider spectrum, as dictated by the interaction of the light and the soil. The high quality images reported by McKnight *et. al.* result from analysis of these signals in the time domain, and in particular, making use of frequencies as high as 30 kHz. However, it will be noted in the spectra that significant laser-induced acoustic energy exists at frequencies well below 2 kHz.

This discussion raises a number of interesting and complicated issues. Let us review the relevant times and frequencies. The pulse is typically a few microseconds in length, as shown on the top panel of Figure 2. This means that the spectrum has a width of something less than 1 MHz. For multiple pulses, this spectrum is multiplied by a comb function at the pulse repetition frequency, as shown in the lower panel. For a finite time sample, each comb tooth will be convolved with the Fourier transform of the time window. For the unfocused laser beam and higher frequencies, the footprint is large compared to the acoustic wavelengths, and the acoustic energy remains well collimated for some distance into the soil. Lower frequencies diverge faster, although they also travel longer distances in the soil. Although the laser-generated acoustic energy is spread over a wide bandwidth,

the narrow comb functions achievable with long dwell time may permit significant noise reduction by synchronous detection. Based on the work at the University of Mississippi reported elsewhere in this volume, these frequencies are particularly useful for detecting mines with a low false-alarm rate. The high frequencies can also be detected through synchronous detection, and based on McKnight, *et. al.* can provide exquisite imagery in three dimensions, for classification, particularly for smaller mines at shallow depths. Optimal use of the data will require further analytical and experimental effort.



5. CONCLUSIONS

We have shown that short pulses of laser light can produce easily measured acoustic signals in soil, which can be used for detection of buried objects. Many issues must be resolved to optimize this technique. In the most basic areas, the mechanism of sound generation is not well understood. Apparently different mechanisms exist for tightly and weakly focused spots. These evidently generate different types of waves in the soil, and the propagation of these waves is not yet understood. Next the information available at each range of frequencies must be investigated to determine the best use of the data. Finally, some choices for the laser pulse length, energy, and repetition frequency may have impact on improving the signal-to-noise ratio of the detected signal.

The ultimate result of this effort could be a non-contact acoustic sensor with the potential standoff distance limited only by the height of the vehicle on which it would be mounted. The optical source and receiver would eliminate problems of direct coupling of energy through the air, and the use of data over a wide frequency band could permit the detection of landmines of different sizes buried at various depths.

6. ACKNOWLEDGMENTS

The authors would like to thank the researchers who participated in these tests, including Julio Montenegro of Laser Science, Inc., and Xiang Ning, of the University of Mississippi.

This work was sponsored in part by the OSD MURI Program under Army Research Office Grant number DAAG55-97-1-0013.

7. REFERENCES

Li, Wen, Stephen W. McKnight, Charles A. DiMarzio, and Ron Roy, "Laser-Induced Acoustic Detection of Shallow-Buried Objects," *Subsurface Sensors and Applications, Proc. SPIE 3752*, October 1999. Pp. 285-291.

McKnight, Stephen W., Li Wen, and Charles A. DiMarzio, "Laser-Induced Acoustic Detection of Buried Objects," *Detection and Remediation Technologies for Mines and Minelike Targets IV, Proc. SPIE 3710*. Presented at AeroSense, Orlando, FL. April 1999.

Witten, Allen J., Stephen W. McKnight, Charles A. DiMarzio, and Li Wen, "Imaging and Detection of Mines from Acoustic Measurements" *Detection and Remediation Technologies for Mines and Minelike Targets IV, Proc. SPIE 3710*. Presented at AeroSense, Orlando, FL. April 1999.

McKnight, Stephen W., Charles A. DiMarzio, Wen Li and Ronald A. Roy", "Laser-Induced Acoustic Generation for Buried Object Detection," *Proc. SPIE 4038, Detection and Remediation Technologies for Mines and Minelike Targets, V*. Publication Pending.

Laser-induced acoustic generation for buried object detection

S. W. McKnight^a, C. A. DiMarzio^a, W. Li^a, R.A. Roy^b

^aCenter for Electromagnetics Research, Northeastern University, Boston, MA 02115

^bDepartment of Aerospace and Mechanical Engineering, Boston University, Boston, MA 02215

ABSTRACT

Mechanisms for the production of acoustic energy in soil by pulsed CO₂ laser excitation of the surface are reported. When the laser pulse is unfocused with a spot size about 1 cm in diameter, a single narrow acoustic pulse is observed with a spectral content near the detector limit of 100 kHz and a velocity of 255 m/s, close to the speed of sound in air. When the laser is focused to a spot size on the order of 1 mm diameter, the audible acoustic intensity is greatly increased and we observe a second broad acoustic feature. This feature has a much lower frequency (near 3 kHz) and velocity (75 m/s). We have tentatively identified the fast mode as a normal compressive mode and the slow mode as a Biot slow-wave. A study of visible light emission when the focused CO₂ laser beam strikes the sand surface indicates ionized nitrogen, oxygen, and silicon are present. This implies that the mechanism for sound production with the focused beam involves ionization by the optical electric field, expansion, and subsequent collapse of the air. The mechanism for sound production by the unfocused beam, which produces better imaging of underground objects, appears to be quite different.

Keywords: laser-induced, acoustic, landmine detection, Biot waves, porous media

1. INTRODUCTION

The use of high-frequency (~30kHz) acoustic waves produced by a pulsed CO₂ incident on the surface of dry sand for imaging of shallow buried objects such as anti-personnel landmines has been demonstrated in the laboratory^{1,2}. The mechanisms for the conversion of the optical pulses into sound, however, is poorly understood. Since the optimization of the process of optical to acoustic conversion in a porous media depends on understanding and modeling the physical processes, we have initiated an experimental investigation of the behavior of the acoustic modes created in soil under different laser pulse focusing conditions. For a broad, unfocused laser spot, modeling the soil as a uniform effective medium as in Figure 1a may be appropriate. If the laser is focused to a small spot comparable to the size of the sand grains as in Figure 1b, a more complex calculation may be necessary which takes into account the random position and orientation of the sand grains and the intervening air spaces. In addition, under focused beam conditions the optical field intensity can become very large, leading to different physical effects. We have observed

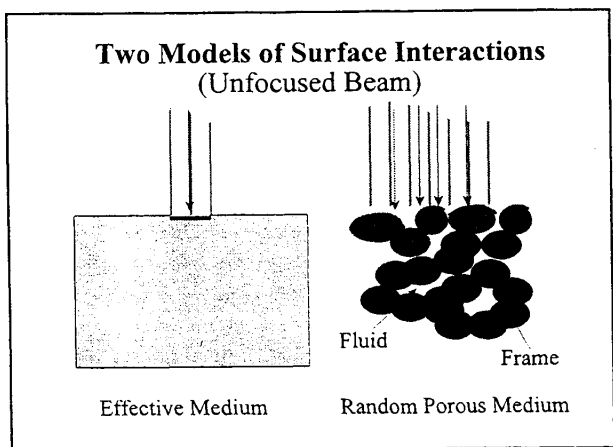


Figure 1 Conceptual models of optical-to-acoustic conversion in soil. a) An effective medium approximation may be adequate when the laser spot is much larger than the structure of the soil, b) The details of the interaction of the light with the discrete particles of the poro-elastic medium may be critical when the beam size is comparable to the particle size.

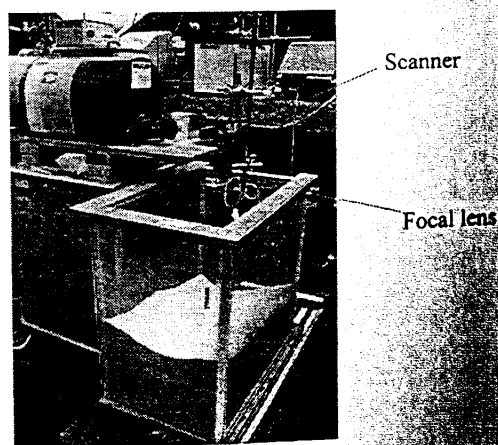


Figure 2 Photograph of experimental setup. The pulsed CO₂ is at top left and the experimental sand box is below. Objects to be detected or hydrophones to study the laser-to-acoustic coupling are buried in the sand.

that focusing the laser beam to a spot size on the order of 1mm in diameter greatly increases the audible acoustic sound associated with the laser pulse and, in addition, can cause the production of a flash of visible light at the point where the infrared CO_2 pulse impinges on the sand. We have studied the time-dependence of the acoustic pulse at various depths and angles in the soil for both focused and unfocused laser pulses. For unfocused pulses we find a single acoustic mode with a velocity slightly less than the speed of sound in air with a directionality perpendicular to the soil surface. With a focused beam we detect two distinct acoustic features in the time trace: a sharp pulse with a velocity and spectral content similar to that achieved with the unfocused beam, and, in addition, a second broad pulse with a much lower frequency content which propagates at a slower velocity. We suggest that these two modes may be related to the two Biot modes allowed in poro-elastic media³.

2. EXPERIMENT

A photograph of the experimental setup is shown in Figure 2. The source laser is a LSI pulsed 10.6μ CO_2 TEA laser with a pulse length of 100ns, a pulse energy of 150mJ, and a repetition rate less than 20Hz. The laser is incident on a 1 m^2 surface of dry sand about 60 cm deep in which we would bury detectors or various types of subsurface targets. In the present studies of the optical-acoustic conversion mechanism, the acoustic signal was measured with a wide-band hydrophone (0-100kHz) which was buried at various depths and positions in the sand. This configuration contrasts with the experimental setup that we used to image underground objects. For imaging, our best results were obtained with a narrow-band (30kHz) tuned PZT detector which was suspended in the air above the position of the laser pulse on the surface of the sand.

The laser was used in two configuration: unfocused or with a focusing lense to reduce the spot size. The unfocused laser spot is an oval approximately $1.1 \times 0.7\text{ cm}$ in size. A germanium lens with a focal length of 25cm was used to focus the beam to a spot size less than 1mm in diameter, comparable to the size of the sand grains. 10.6μ light is very effectively absorbed in soil with a reflection less than 10% and skin depth of approximately 5μ . The sound produced by both the focused and the unfocused beams was audible across the room. With the focused beam the sound was considerably louder and there were flashes of visible light emitted and some movement of sand grains at the spot where the laser hit the surface.

The acoustic signal detected by the buried hydrophone for the focused and unfocused beams is shown in Figure 3 and Figure 4.

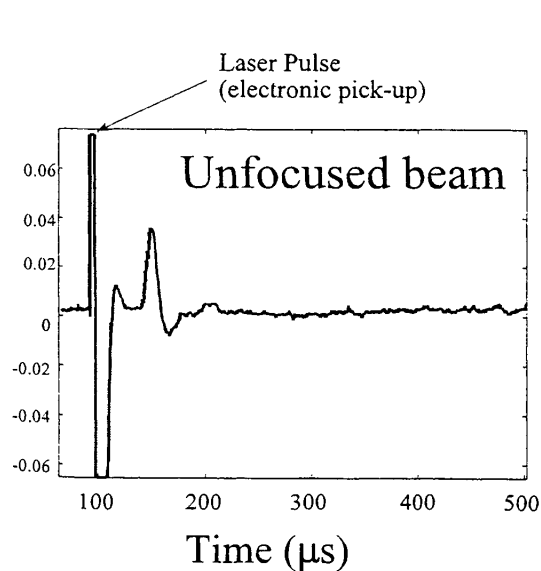


Figure 3 The time trace of the acoustic signal measured by a buried hydrophone as a result of a pulse from the unfocused laser on the sand surface.

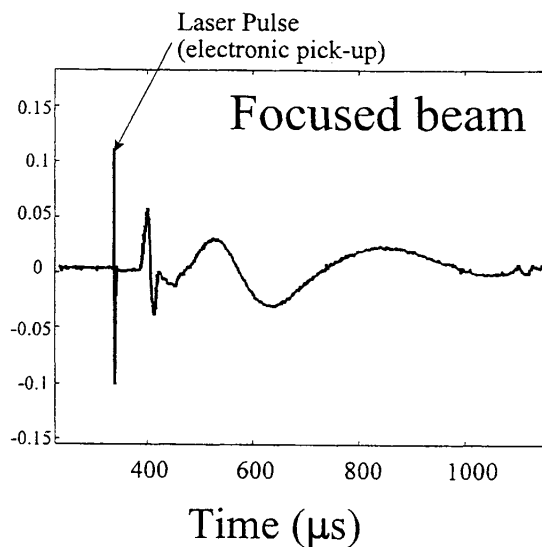


Figure 4 The time trace of the acoustic signal measured by a buried hydrophone as a result of a pulse from the focused laser on the sand surface.

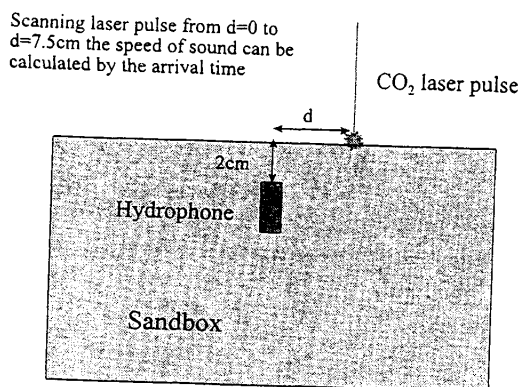


Figure 5 Experiment to measure the velocity in sand of the acoustic features in Figure 3 and 4.

The initial spike is electrical noise caused by the laser discharge. The acoustic pulse is dominated by a sharp peak with a width (about $10\mu\text{s}$) that is probably limited by the bandwidth of the 100kHz detector. For the focused beam only, there is a second, much broader feature with a peak-to-peak width of about $300\mu\text{s}$.

To determine the propagation speed of these two different acoustic features observed with a focused laser beam, we measured the change in the acoustic signal as a function of distance from the laser spot to the detector using the configuration in Figure 5. The time trace of the acoustic signal measured by a buried hydrophone as a result of a pulse from the focused laser at two different positions on the sand surface are shown on the left in Figure 6. Identifying the sharp feature by the initial spike, and the broad feature by the dip between the two broader peaks, in Figure 6 we plot the time delay after the laser fires of these two features as a function of distance from the laser spot to the detector. By taking the slope of these two curves, we find the velocities of the two features are dramatically different. The sharp feature propagates with a velocity around 255 m/s, while the broad feature propagates with a velocity of about 75 m/s.

The observation of acoustic modes with two different velocities in porous media is a well-known phenomena. In the theory of Biot, the solution for acoustic propagation in a random porous elastic media gives rise to two modes: a mode in which the frame and the fluid move in phase (compressional wave) and an addition mode in which the motion of the frame is out of phase with the motion of the fluid. Typically the in-phase mode has a velocity which is close to the velocity of sound in the fluid, while the out-of-phase mode, the "Biot slow wave" has a lower velocity. If we tentatively make an identification of the features we observe with the Biot modes, the sharp feature which travels at 255 m/s, close to the velocity of sound in air, could be the Biot in-phase mode, while it is natural to identify the broader feature, low-velocity feature with the Biot slow wave. Note that we only observe this second, slower wave under excitation with a focused laser beam.

To measure the angular dependence of the magnitude of the acoustic signal from the focused beam, we positioned hydrophones at different angles along an arch buried in the sand centered on the position of the laser spot, as shown in Figure 7. In Figure 8 and 9, we plot the peak amplitude of the sharp "fast-wave" signal and the

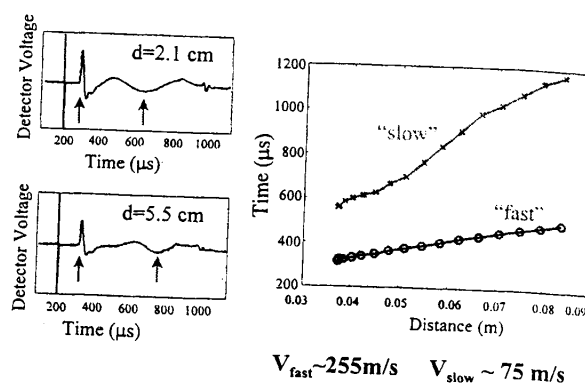


Figure 6 The arrival time vs. distance for the features indicated by the arrows in the inset. The propagation velocity in sand of the two features is determined as indicated from the slope of the lines.

Angular Distribution of Acoustic Intensity

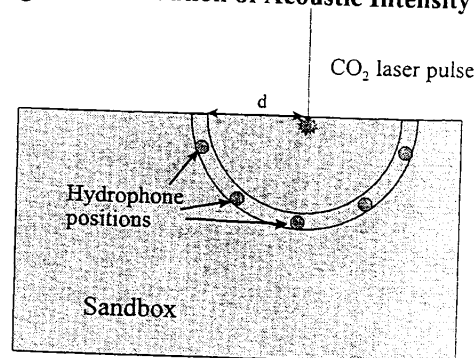
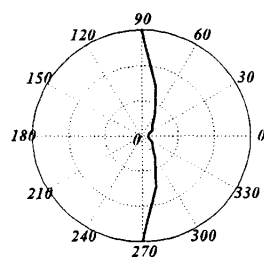


Figure 7 Experimental setup to measure the angular directivity of the acoustic features in Figure 3 and 4.

**Fast Wave Angular Distribution
(Focused laser beam)**



$$\text{Rayleigh distance} = D^2/\lambda \leq 0.01 \text{ cm}$$

Figure 8 Angular dependence the amplitude of the “fast wave” feature produced with a focused laser beam.

**Slow Wave Angular Distribution
(Focused laser beam)**

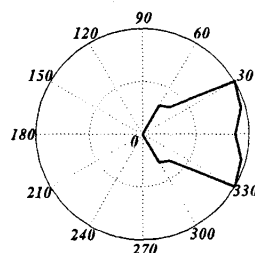


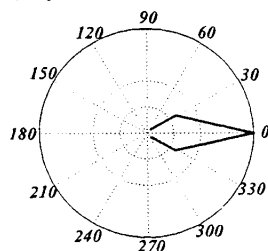
Figure 9 Angular dependence the amplitude of the “slow wave” feature produced with a focused laser beam.

broad “slow-wave” signal as a function of angle from the normal. The two modes show a dramatic difference in angular distribution, with the slow wave signal propagation peaked at angles perpendicular to the surface, while the fast-wave signal strength is peaked along the surface of the sand. The angular dependence of the fast-wave mode is hard to understand, since when the beam is unfocused we observe only a single mode with a shape and velocity similar to the fast-wave signal, but with an angular distribution sharply peaked at a direction normal to the surface, as shown in Figure 10.

The unfocused beam has a Rayleigh length d^2/λ on the order of 1 cm, so for distances into the sand on the order of 1 cm the unfocused beam looks like an extended source. The focused beam, on the other hand, has a Rayleigh length less than a millimeter and is a point source to a good approximation. Since the radius of the detector arch was about 10 cm, it is unlikely, however, the difference between the fast-mode angular dependence seen in Figure 9 and 10 can be accounted for by such physical optics effects. It is more likely that the different acoustic excitation physics between the broad laser beam and the focused beam, such as illustrated in Figure 1, may account for these differences. We will return to this question later.

Table I summarizes the differences between the two acoustic modes that are observed with the focused laser beam. We note that these differences may impact on the utility of the two modes for underground detection and imaging. The fast mode, having a

**Angular distribution
(Unfocused Laser Beam)**



$$\text{Rayleigh distance} = D^2/\lambda \geq 1 \text{ cm}$$

Figure 10 Angular dependence of the amplitude of the fast-wave feature produced by an unfocused laser beam.

Table I: Comparison of Fast-Wave and Slow-Wave laser-induced acoustic features produced with a focused laser beam.

Comparison of Fast and Slow Acoustic Waves

	Fast Wave	Slow Wave
Unfocused Beam	Yes	No
Focused Beam	Yes	Yes
Speed	250 m/s	75 m/s
Frequency peak	100 kHz*	3 kHz
Directivity	Along Surface	Perpendicular to Surface
$\alpha(\omega)$	Unknown	Unknown

* Band-width of detector=100kHz ($1/\Delta t_{\text{detector}} = 10 \text{ MHz}$)

higher spectral frequency content may be more useful for high-resolution imaging where a small wavelength is desired. On the other hand, if we have correctly identified the slow wave as a Biot out-of-phase wave, it will see an exceptionally large contrast between the porous soil where the slow wave is allowed and any subsurface non-porous object where, since there is no longer any differentiation between frame and fluid, its propagation is forbidden. There may also be implications about detection strategies, since to the extent that the energy of the slow-wave is carried in the fluid, techniques that measure the motion of the solid surface, such as laser vibrometry, will be less effective than with an ordinary compressional wave. On the other hand, for the same reason the mismatch between the poro-elastic wave and the wave in the air may be less for the slow-wave than the fast wave, so that detection by means of a microphone above the soil may be more favorable for the slow wave.

It is noteworthy that our attempts so far at subsurface imaging with the focused beam have been less successful than our previous work with the unfocused beam. These imaging studies were carried out with an acoustic detector suspended in the air above the surface. It is not possible to say at present whether the lack of success was due to an inherent effect such as the longer wavelength of the slow wave, less efficient acoustic coupling to or from the soil and air, or some interference effects between the fast and slow acoustic waves, or more incidental effects such as noise from the movement of sand grains and drifting of the laser focus due to heating of the air and movement of the surface.

One clue about the mechanism for acoustic excitation is that a flash of visible light is observed which the CO₂ pulse hit the surface of the sand when the beam is focused, but not for the unfocused beam. Origins of this visible light that we considered were black-body incandescence of the sand surface, photoexcitation and luminescence of the sand material, and electric field ionization and recombination. The energy density of the focused beam is sufficient to heat the sand over 10⁶ °C if the effects of radiant emission and conduction are neglected, but this does not seem likely on the relatively long (0.1 μs) timescale of the pulse. Luminescence of SiO₂-based materials is usually very broad-band and can extend into the ultraviolet.

To gather more information about the light emitted when the focused CO₂ laser pulse strikes the sand surface, we measured the spectrum of the emitted light using a multispectral filter and a photomultiplier detector as in Figure 11. The measured spectrum is reproduced in Figure 12 and Figure 13. The emission peaks are negative, because the electron current of the photomultiplier

Spectral Signature of Laser Flash (Focused beam)

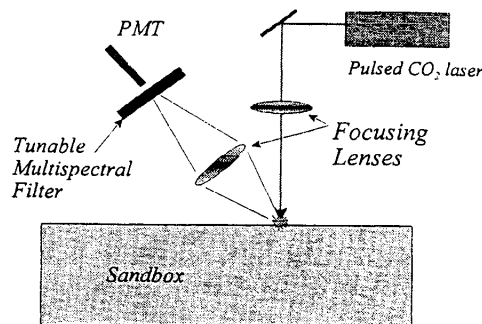


Figure 11 Experimental configuration for measuring the spectrum of visible light produced when a focused laser beam is incident on sand.

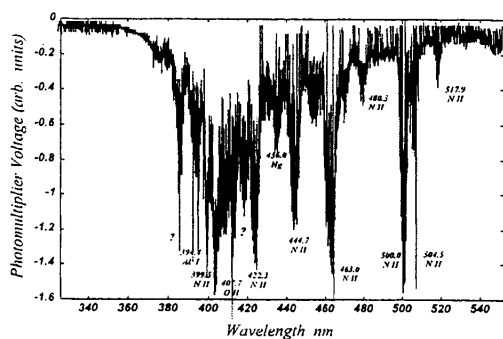


Figure 12 Spectrum of visible light produced by focused laser beam incident on surface of sand (short wavelength). Emission lines are identified from standard handbooks.

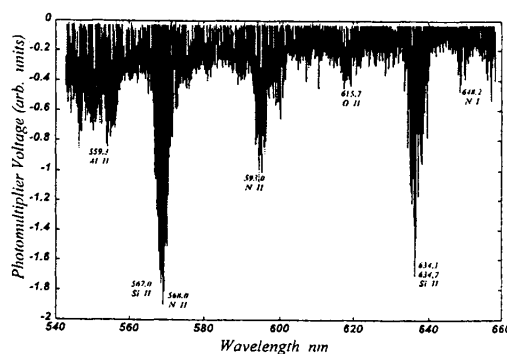
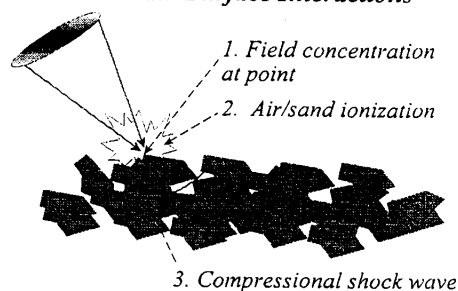


Figure 13 Spectrum of visible light produced by focused laser beam incident on surface of sand (long wavelength). Emission lines are identified from standard handbooks.

is converted to a negative voltage by a voltage-to-current amplifier. Electron photocurrent resulting from the detected light yields a negative voltage.

We have identified emission lines in the spectrum by comparison with the CRC Handbook of Physics and Chemistry. We note that large emission lines associated with ionized nitrogen, oxygen, and silicon are present in the spectrum which implies that the model for optical field ionization is the best model for the production of the light. The concentration of the electric field of the optical signal near the sharp points of the sand grains is the most likely mechanism for the breakdown and ionization of the air and sand.

Focused Beam Surface Interactions



The acoustic production for the focused beam can then be best explained by the collapse of the air following the rapid expansion associated with the ionization and plasma formation as shown in

Figure 14. The intense audible noise associated with the focused beam is analogous to the crack associated with an electrical arc or with thunder. One implication of this model is that the sound production for the focused beam is probably primarily as a compressional wave in the air above the soil. The reduced effectiveness of imaging with the focused beam may be a result of the poor coupling of this acoustic compressional wave in air into the soil.

Figure 14 Model for production of light and sound by focused laser beam impinging on soil surface.

The unusual angular dependence of the compressional fast wave for the focused beam seen in Figure 9 may represent the more effective transport of acoustic energy in the air along the surface of the sand than propagation in the sand. In effect, the sound produced by this ionization and collapse travels through the air to a point close to the high-angle, near-surface sensor and then propagates only a short distance into the soil to the detector. Since the velocity of the fast wave is very near the speed of sound in air, it is very difficult to distinguish this propagation channel from propagation through the soil. The slow wave may not have the same channel available to it because a Biot wave will not propagate in a single component fluid such as the air.

3. CONCLUSIONS

The production of acoustic energy in soil by a pulsed CO₂ laser has been studied. When the laser beam is focused to a spot size comparable to the soil grains, a much louder acoustic sound is observed and two different acoustic modes are detected in the soil. One mode, which we identify with a normal compressional acoustic wave has a velocity (255 m/s) close to the speed of sound in air, and a bandwidth of at least 100kHz. The second mode, which we have tentatively identified as a Biot slow-wave phenomena, has a velocity of only 75 m/s and a spectral peak near 3 kHz. By analyzing the visible light emitted by the focused laser beam, we conclude that for the focused laser pulse the acoustic energy is produced by ionization of the air and soil components, with a rapid expansion and subsequent collapse of the air column. Preliminary investigations have indicated that imaging of subsurface objects with the focused beam laser is not as successful as with a broader unfocused beam. This may be a result of the poor coupling of sound produced in air into the soil. Studies to measure the attenuation and dispersion of the acoustic signals produced by the focused and unfocused laser beams are underway.

4. ACKNOWLEDGEMENT

The research described here was supported by the DARPA and ARO under the Multidisciplinary University Research Initiative.

5. REFERENCES CITED

1. S. W. McKnight, C. DiMarzio, W. Li, D. O. Hogenboom, G. Sauermann, "Laser-Induced Acoustic Detection of Buried Objects," SPIE Transactions on Detection of Mines and Mine-like Objects Vol. 3392, p. 841-847 (1998).
2. S. W. McKnight, W. Li, and C. DiMarzio, "Imaging of Buried Objects by Laser-Induced Acoustic Detection," SPIE Transactions on Detection of Mines and Mine-like Objects, Vol. 3710, 231-238 (1999).
3. M. A. Biot, "Generalized Theory of Acoustic Propagation in Porous Dissipative Media," J. Acoust. Soc. Am. 34, 179-191 (1962).

PROFILING THE PERFECTLY MATCHED LAYER TO IMPROVE LARGE ANGLE PERFORMANCE

B127.6

Scott C. Winton and Carey M. Rappaport
Center for Electromagnetics Research
Northeastern University
Boston, MA 02115

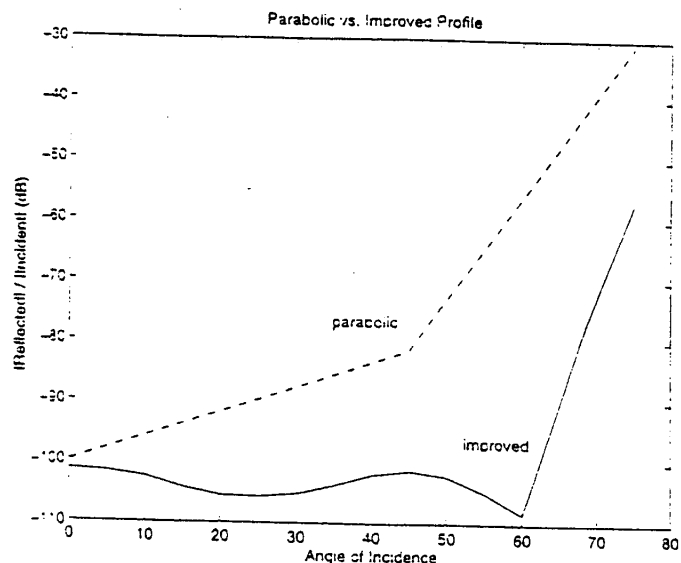
Absorbing Boundary Conditions (ABCs) are used in Finite Difference Time Domain (FDTD) scattering problems in order to prevent reflections off the lattice edges. The introduction of the Perfectly Matched Layer (PML), (Berenger, *J. Comp. Phys.*, Oct. 94), revolutionized ABCs. In continuous time and space, the PML completely absorbs waves, independent of frequency and angle of incidence and can attenuate these waves at an arbitrary rate. Despite the obvious utility of the PML, its numerical implementation is problematic. The amount of attenuation experienced by waves incident on a PML drops off sharply as the angle of incidence, θ , increases. Despite considerable effort in interpreting, adapting and improving it, little success has been achieved in modifying the PML in such a way as to achieve consistent performance over a wide range of angles, until now. Presented are the constitutive values of conductivity σ that will allow the PML to achieve a total reflection of -100 dB (0.001%) through angles beyond 60°. Also presented is the background and methodology required to determine these values.

In order to understand the problems in discretizing the PML, consider that it has been established that the decay rate of the PML in continuous time and space is $\alpha = \sigma \eta \cos \theta$. When discretized, the total two way loss for the an N-layer PML is $L = \text{Exp}(-2 \sum_{n=0}^{N-1} \sigma_i \eta \cos \theta \Delta)$, assuming perfectly conducting termination. The value of σ at a spatial point within the PML is σ_i , Δ is the spatial increment and i is the spatial index. Because the decay rate per PML layer, defined by $S_i = \sigma_i \eta \Delta$ is a source of discretization error, the values of S_i must be chosen carefully (Rappaport, *IEEE Trans. Magn.*, May 96) and clearly any S_i may not be arbitrarily large. The minimization of computational overhead requires that N be as small as possible. However, since L is inversely proportional to $\cos \theta$, if N and the values of σ_i are too small, L may become unacceptably large. Berenger found that varying the conductivity parabolically $\sigma_i = \sigma_f (i/N)^2$, where σ_f is a constant, optimized absorption. Still, Berenger's published results indicate that reflections from the discretized PML increase rapidly as θ increases.

To find a set of σ_i that will allow the PML to achieve consistent results over a wide range of angles requires one assumption and two observations. Consider an 8 layer PML, this allows for profiling of 16 half-layers, since at every spatial point, there is a σ_i associated with both \vec{E} and \vec{H} fields.

Based on Berenger's results, it is assumed that there exists a profile of the form $\sigma_j = \sigma_f (j/16)^p$, $j = 1, 2, \dots, 16$ ($i = [j/2]$) where σ_f and p must be found to optimize performance. The first observation is that reflections off the PML at small angles are dominated by discretization error. Finding the best set of σ_j 's for small angles is simply balancing the discretization error against L . An increase in σ_f will increase the discretization error and decrease L . The value of p effects the change in one value of σ_j to the next. If this "contrast" is large, reflections from adjacent spatial points are increased. The overall effect of changes in p are difficult to predict. The second observation is that reflections from the PML at large angles are dominated by L . It will be shown that discretization error goes down with an increase in θ . The key then, is to have σ_f large enough to provide a suitable attenuation at large angles and to adjust p to minimize both the discretization error and L at low angles.

One-dimensional FDTD simulations (Winton, Rappaport, *97 ACES Symp. Dig.*, to be published) are used for an automated search for σ_f and p . The gaussian pulse width and FDTD parameters are those used by Berenger to facilitate comparison. Reflections were analyzed for angles from normal to 60°. When an appropriate set of parameters was found, several angles were tested with the 1-D simulations and verified with 2-D simulations.



The results above clearly indicate that significant improvement over the parabolic profile has been achieved for angles between 0° and 75°. In some cases this improvement is as much as 40 dB.

DETECTING ANTIPERSONNEL MINES WITH A HANDHELD PARABOLIC REFLECTOR TRANSMITTER / MULTISTATIC RECEIVER IMPULSE GPR

Carey Rappaport, and Bo Yang
Northeastern University, Boston, MA
(617) 373-2043 (v) (617)373-8627 (f)
rappaport@neu.edu

Stephen Azevedo, Tom Rosenbury, and Jamie Gough
Lawrence Livermore National Laboratory, Livermore, CA
(925) 422-8538

Arnold Dean
Geo-Centers, Inc. Newton, MA
(671) 964-7070

Category: Seesion 9, Mine Detection

Abstract

A novel handheld time-domain array GPR antipersonnel mine detection system using an offset paraboloidal reflector antenna is described. The reflector collimates rays from an ultra-wideband transmitting feed, directing the microwave impulse forward, in front of the antenna structure. As such, much of the ground reflected wave is directed further forward, away from the operator, the reflector, and the receiving antennas, and thereby reducing the major source of clutter. The wave transmitted into the ground that interacts with the target, generating significant backscatter returning toward the receiving antennas. These receiving antennas are configured in a 2 by 2 array to provide spatial focusing in both the along- and cross-track directions.

This system has been built and tested at both Lawrence Livermore National Laboratory, and GeoCenters, Inc. In both cases, custom-built wideband antenna elements generate narrow pulse shapes, which allow for resolving small non-metallic targets buried at shallow depths. The LLNL's Micro-Power Impulse Radar (MIR) operates in the 1.5 to 5 GHz range a very narrow pulse shape. The Geo-Centers wideband TEMR antenna elements have higher power, though lower frequency range (850 to 1700 MHz), and generate less residual ringing in the time signal.

Preliminary measured data from both systems indicate that the surface clutter is indeed reduced relative to the target signal, and that small non-metallic anti-personnel mines can be reliably detected at burial depths as shallow as 1 inch in both dry sand and dry vegetative clay loam soil.

Introduction

The Northeastern University Multidisciplinary University Research Initiative (MURI) demining effort, sponsored by the Army Research Office, has been investigating novel sensing systems and processing algorithms to detect small, shallow buried, low metal content antipersonnel mines. Finding buried plastic mines with conventional electromagnetic induction metal detectors is problematic, due to the relatively low amount of metal (just in the firing pin) relative to surrounding metallic clutter [1]. Ground penetrating radar (GPR) has been shown to be effective in detecting shape anomalies characteristic of buried mines [2]. However, identifying buried target signals amid rough ground clutter is particularly difficult for small mines buried close to the ground surface. To address this problem, it is essential that the GPR sensor minimize the strongest scattering contribution: the ground surface reflection.

A new GPR system designed at Northeastern University, and fabricated at both Lawrence Livermore National Labs (LLNL) and Geo-Centers, Inc. reduces this ground clutter by illuminating the sample ground surface with a forward propagating, quasi-planar wave, and receiving the scattered signals with a two-dimensional multistatic array. Since the scattering by a small target is relatively isotropic, while scattering by the ground is primarily specular, a planar transmitted signal is well suited for shallow GPR detection. Plane wave illumination has another advantage beside clutter reduction compared to point source excitations: for a given target burial depth, the wave incident on a target from a plane wave source will always scatter the same way. For a point source, the incident wave on a given target will be illuminated from the side for one transmitter position and directly above for another. This constant exposure angle for the planar wave makes processing the returned signals more straightforward.

In both the LLNL and Geo-Centers systems, the excitation signal is sufficiently short in time duration to resolve small targets and discriminate the ground surface from a shallow buried target. The multistatic array concept provides for additional clutter rejection and time-domain focusing [3]. This focusing is accomplished by measuring, comparing, and summing the backscattered signals at each receiver in the narrow time window between the times when the residual ground reflected wave passes the receiver and before this wave re-reflects from the reflector components.

The finite-difference time-domain (FDTD) [4] method has been used to electromagnetically model the novel GPR configuration. We implemented the 2-D FDTD code to simulate the generation of the non-uniform plane wave, the scattering by the modeled dispersive soil ground surface, the scattering by the target, and the retransmission back into the air, confirming the clutter minimizing characteristics of this mine detector.

Parabolic Reflector Transmitter GPR

The quasi-planar transmitted wave is generated using an offset paraboloidal reflector antenna. The resulting wave is incident at 45 deg. to normal, and is fairly uniform over the portion of ground being investigated. Because the transmitted wave diverges very little from reflector to the ground, most of the power incident from the illuminating feed is transferred to the ground. Much of the ground reflected wave is directed further

forward, away from the operator, the reflector, and the receiving antennas, reducing clutter. In addition, the wave transmitted into the ground is incident on the target in the same manner for any antenna position: always as a plane wave with constant soil path length and incident angle. Since the scattering from an electrically small buried target is primarily isotropic, there will be a significant backscattered signal, propagating oppositely to the surface clutter signal, returning toward the receiving antennas.

Figure 1 shows the geometry of the parabolic reflector and the way it directs rays from the transmitting feed to the ground. Diverging rays leaving the transmitter reflect from the paraboloidal surface, emerging as parallel rays, in such manner as to keep the path length from the feed to an inclined wavefront constant. The inclined wavefront is perpendicular to -- and propagates along -- the axis of revolution of the parent paraboloid, which includes the parabola focus and vertex. Also, the reflector produces a beam of microwave energy with an abrupt drop in power outside the ray tube bounded by the perimeter of the reflector. As long as the distance from reflector to ground S is comparable to the projected reflector diameter D , the rays representing the transmitted wave will be parallel, and the wave will be planar. The governing equation for the nearfield of the reflector is: $S \ll 2 D^2 / \lambda$, so when the reflector is positioned close to the ground, the radiated wave is in the nearfield, and concepts of antenna gain and radiation pattern are irrelevant.

An offset section of the paraboloid is selected to avoid blockage of rays by the feed structure. In contrast to offset reflectors used in communications applications, this offset section is particularly deep, extending from the vertex past the focal point by twice the focal length, giving an F/D ratio of the parent paraboloid of about 0.15. The best offset section extends from 45 deg. to about 115 deg. from the symmetry axis, which ensures that the front and rear edges are at the same height above ground. For a parabolic focal length of 20 cm, the projected aperture diameter of the reflector is about 47 cm, which nominally illuminates an elliptical spot of ground with axes 47 and 67 cm. For a reflector positioned 33.5 cm above the ground, the center of this elliptical region is immediately below the front edge of the reflector, and all of the collimated rays from the reflector would reflect from a flat ground just missing the front of the reflector.

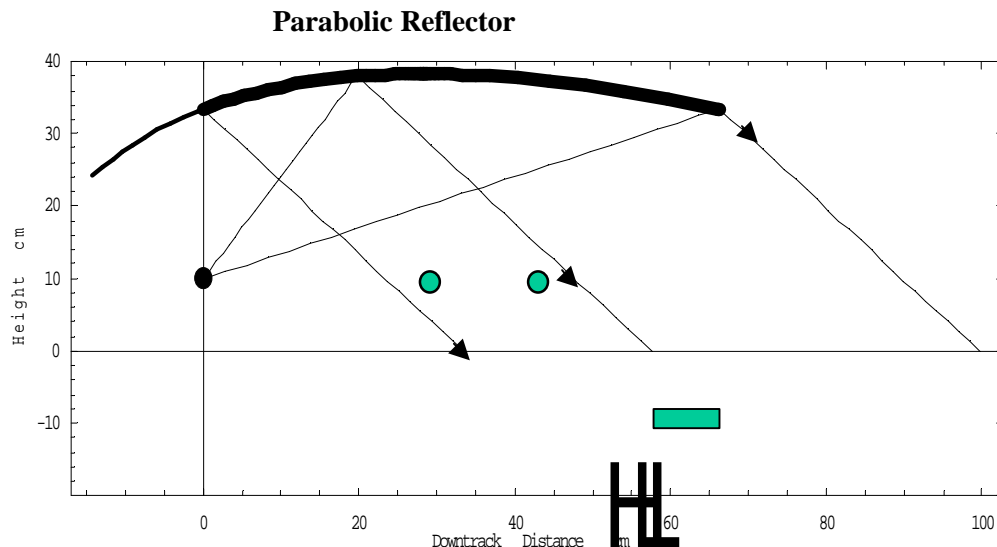


Figure 1. Geometry of the offset parabolic reflector

The receivers are positioned under the reflector, but behind the point on the ground at the center of the illuminated spot. The 4 receivers are arranged in a rectangular 2 by 2 array, with the forward pair separated by about 40 cm, and the backward pair by the same distance, 20 cm behind the forward pair. Since the receivers are displaced from the centerline, they do not appreciably block any of the wave from the reflector to the ground.

For FDTD modeling the dispersive soil, several modeling methods have been proposed to avoid convolution in the time domain [6,7]. Our method makes use of the Z-transformation. We approximate the frequency dependent conductivity using the (2-2) Pade' approximant

$$s(Z) = \frac{b_0 + b_1 \cdot Z^{-1} + b_2 \cdot Z^{-2}}{1 + a_1 \cdot Z^{-1}} \quad (1)$$

and assume a constant average dielectric constant. With the a_1 , b_0 , b_1 , b_2 parameters in (1) chosen to match the experimental data, such as Puerto Rican clay loam [8] with density 1.4 g/cc and moisture 10%, we have $a_1 = -0.88$, $b_0 = 0.9162$, $b_1 = -1.6766$, $b_2 = 0.7611$, and $\epsilon = 4.2$. The perfectly matched layer (PML) absorbing boundary condition is used in the FDTD model to terminate the computational grid [9,10]. Computational models of the parabolic reflector system indicate that the wave reflected by the offset parabola does indeed remain planar over the illuminated region of ground, and that the scattering by rough soil surface is primarily specular. In addition, the mine scattered signal appears to be circular, showing that the electrically small target scatters almost isotropically.

Lawrence Livermore National Laboratory System and Results

The Lawrence Livermore National Laboratory (LLNL) Micro-Power Impulse Radar (MIR) was used as the transmitter source for one of fabricated systems [5]. This radar source generates an impulse with pulse width of about 300 ps and frequency range from about 1.5 to 5 GHz (see Figure 2), and has the particular advantage of being small and extremely low cost: both important features for mine detectors used in developing countries. This source was assembled with a custom-built metallic offset paraboloidal reflector. Figure 3 shows the full mine detector prototype; Figure 4 shows the device performing measurements at the test site at LLNL. In this test, a non-metallic antipersonnel mine simulant was buried in dry sand 1 in. below a very rough surface. This is a particularly challenging detection problem, because the dielectric constants of the plastic body TNT filled mine and the surrounding soil are very close. In addition, the random rough surface height variation is of the order of the height of the mine, and its burial depth. Thus, the anomaly detection is frustrated by low signal to clutter both in terms of size and contrast.

The result of processing the measured signals is shown in Figure 5, with bright areas signifying anomalies. Although there is still appreciable clutter from the rough ground, the target is still visible in the center of the image. The extent of the rough ground variation precludes clutter suppression using purely signal processing means. However,

by ensuring that the ground scattered signal specularly reflects away from the receivers, the target signal can be discriminated from the clutter.

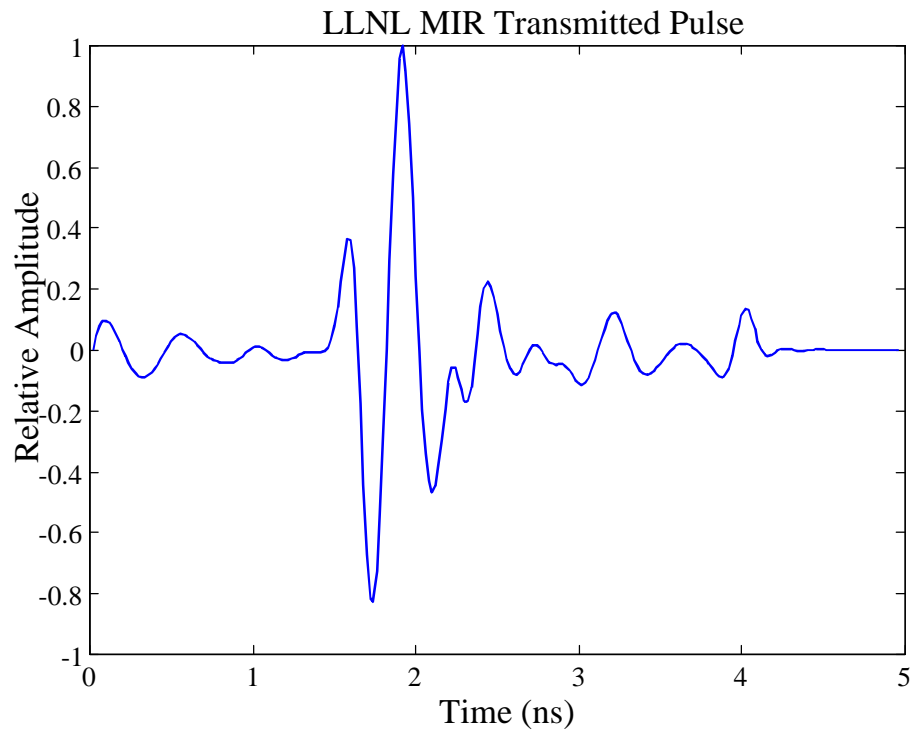
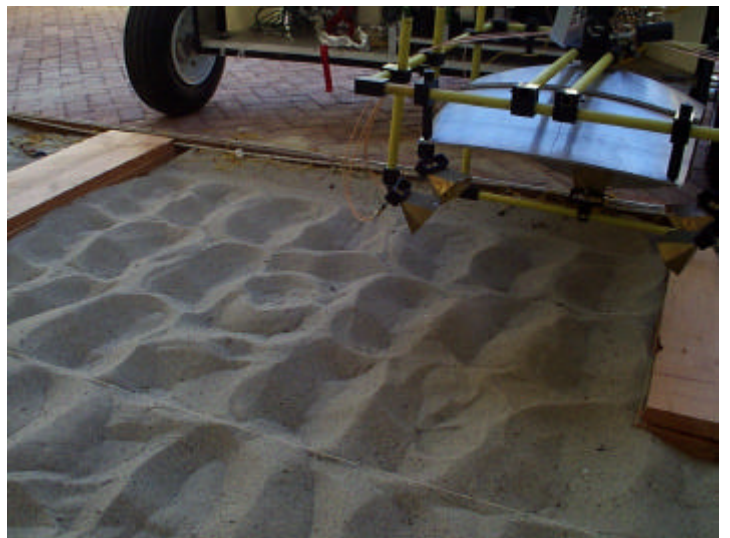
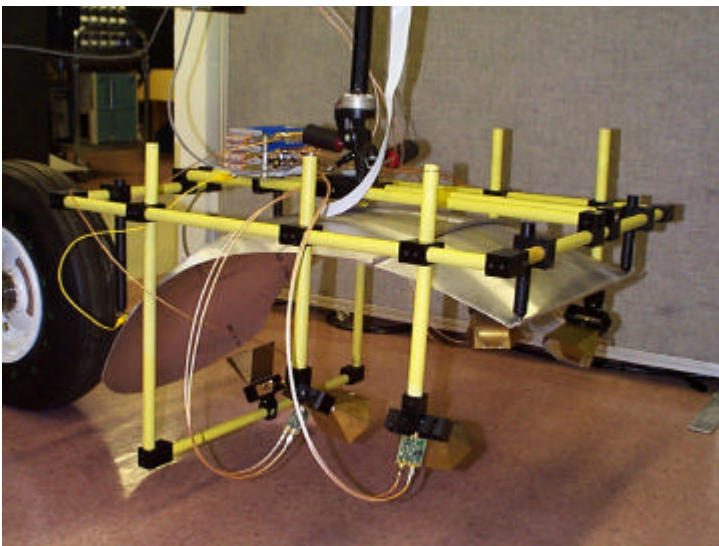


Figure 2 LLNL MIR pulse shape



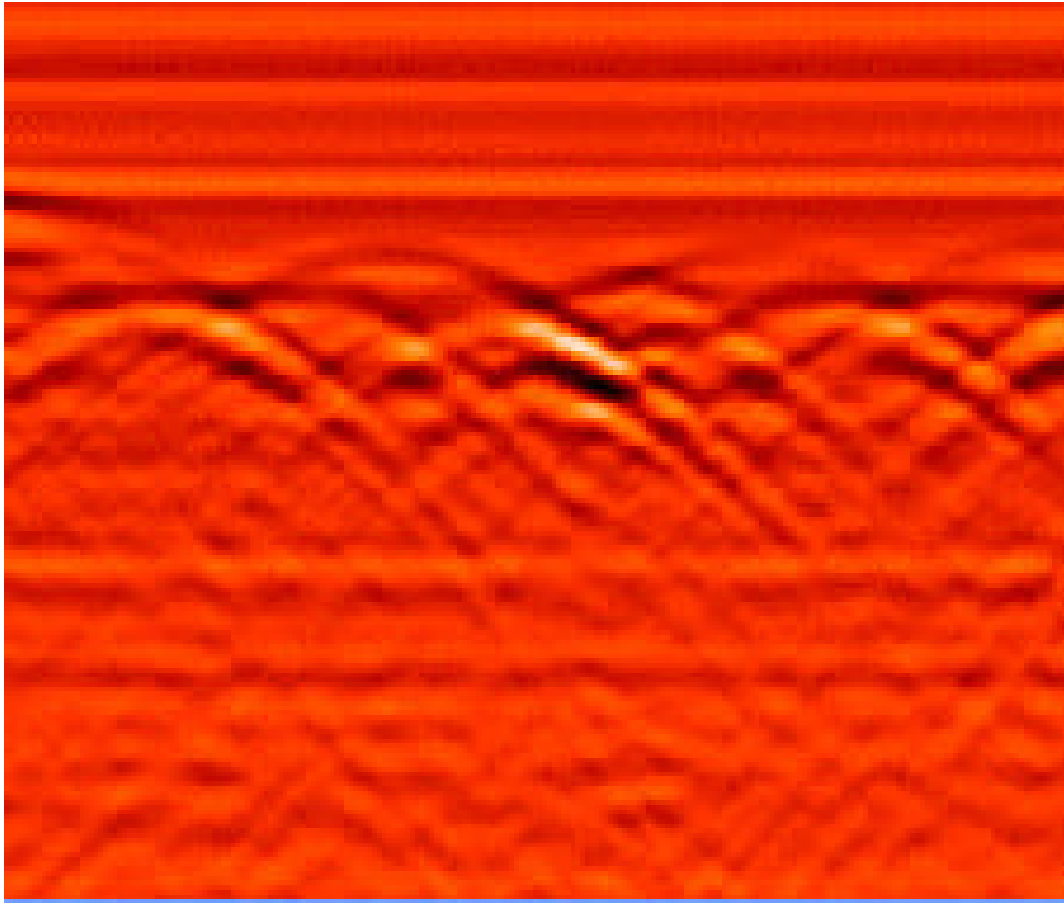


Figure 5 Detection results for the LLNL MIR system for rough sand with AP mine buried 1 in.

Geo-Centers System and Results

Another mine detection system based on the offset parabolic reflector transmitter was fabricated by GeoCenters, Inc. The reflector consists of metallized fiberglass, and the antenna elements are proprietary Transverse Electromagnetic Rhombus (TEMR), fed by a 1 ns impulse shown in Figure 6.

The system is shown in Figure 7 as it is configured for measuring signals on the Northeastern University test track. The targets in this test were seven non-metallic antipersonnel mine simulants buried 1 in. in moist loam, with naturally occurring vegetation. While this soil surface was not as rough as in the LLNL test, there is significant realistic clutter from long grass. The seven targets were spaced roughly 60 in. apart.

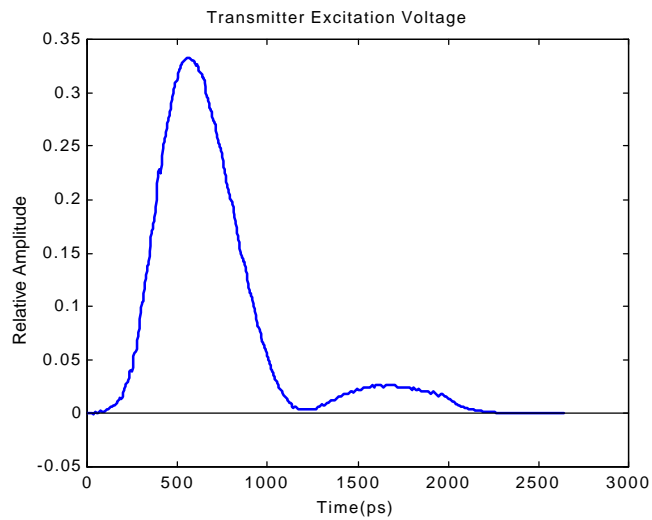


Figure 8 shows the detection results by combining the registered signals from the four receivers each with the moving average background signal removed. It is apparent that six of the seven targets are detected (note the proximity to the indicated true positions), and only one false alarm is generated.

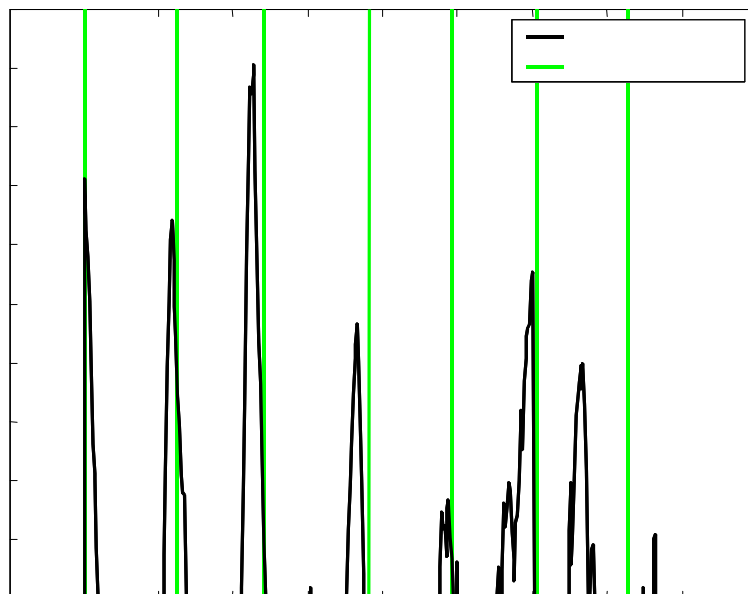


Figure 8 Detection results for the Geo-Centers TEMR system for vegetated moist loam with AP mine buried 1 in.

Conclusions

A novel GPR mine detection system that reduces ground surface clutter has been developed and tested. The detector uses an offset parabolic reflector to generate a forward propagating plane wave to illuminate the ground and a 2-dimensional multistatic array to focus and enhance the received backscattered signal. Numerical simulations support the concept of specular ground reflection with more isotropic target scattering.

Measured signals using both the LLNL MIR and the Geo-Centers TEMR radars for rough dry sand and vegetated moist loam indicate that small non-metallic targets can be detected and discriminated from the cluttered background, even when shallow buried to a depth of 1 in.

Acknowledgement

This work was supported by The Army Research Office Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013.

References

- [1] Witten, A., Won, I.J., Norton, S., "Imaging Underground Structures Using Broadband Electromagnetic Induction," *J. Environ. Eng. Geophysics*, vol. 2, no. 2, Sept 1997, pp. 105-114
- [2] Peters, L., and Young, J., "Applications of subsurface transient radars," I *Time Domain Measurements in Electromagnetics*, Miller, E. Ed., New York: Van Nostrand Reinhold, 1986.
- [3] Rappaport, C. and Reidy, D., "Focused Array Radar for Real Time Imaging and Detection," *1996 SPIE Aerosense Symposium*, Orlando, FL, April 1996, pp. 202--213.
- [4] Yee, K., "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media," *IEEE T. Ant. Prop.*, vol.14, 1966, pp. 302-307.
- [5] Warrick, A., Azevedo, S., and Mast, J., "Prediction of Buried Mine-like Target Radar Signatures using Wideband Electromagnetic Modeling." *Proc. of SPIE*, vol. 3392, pp. 776-783, 1998.
- [6] Rappaport, C., and Weedon, W., "A general method for FDTD modeling of wave propagation in arbitrary frequency dispersive media", *IEEE Trans.Ant. Prop.*, March, 1997, pp. 401-410.
- [7] Rappaport, C., Wu, S., and Winton, S., "FDTD Wave Propagation Modeling in Dispersive Soil Using a Single Pole Conductivity Model," *IEEE Trans. Magnetics*, vol. 35, May 1999, pp. 1542--1545
- [8] Hipp, J., "Soil Electromagnetic Parameters as Functions of Frequency, Soil Density, and Soil Moisture," *Proc. of the IEEE*, vol. 62, no. 1, January 1974, pp. 98-103.
- [9] Berenger, J., "A Perfectly Matched Layer for the Absorption of Electromagnetic Waves," *J. of Computational Physics*, vol. 114, no 1, October 1994, pp. 185-200.
- [10] Rappaport, C., and Winton, S., "Using the PML ABC for Air/Soil Wave Interaction Modeling in the Time and Frequency Domains," accepted for publication in *International Journal of Subsurface Sensors and Applications*, February 2000.

Detection of Objects Buried in Soil Using Microwave Heating

by

Charles A. DiMarzio

Taner Oktar

Wen Li

Carey M. Rappaport

Center for Electromagnetics Research

Northeastern University

Boston, Massachusetts 02115

ABSTRACT

Microwave heating of soil offers the potential to enhance infrared signatures of buried objects such as landmines. In uniform soil, with no vegetation and a flat surface, images can be obtained showing the shape of the objects, to aid in their identification. Combined with other subsurface imaging modalities, this promises a reduced false alarm rate, leading to more effective demining operations. However, in the presence of rough ground, non-uniform soil, vegetation, and solar heating, the signatures become much more complicated. In this work, we examine some of these issues, based on outdoor experiments and a two-dimensional model.

Key Words: Microwave heating, infrared imaging, landmine detection.

1. INTRODUCTION

We have previously reported analytical and experimental results showing that microwave heating of the ground can produce enhancements of the infrared signature of buried objects [DiMarzio, et. al., Sept. 1998], and we have described two time periods in which different signatures impart information about the electromagnetic properties and thermal properties, respectively [DiMarzio, et. al., 1999]. In those works, analysis was performed assuming a flat ground surface, and experiments were conducted in controlled laboratory settings, with dry sand or soil in the absence of vegetation. Here we report on preliminary measurements in an outdoor environment and a model which accounts for surface roughness.

Previous laboratory work had been hampered by the requirement to enclose the test area in a screen box to prevent exposure of the IR camera and the personnel to microwave

energy. In the outdoor experiments, we used an elliptical reflector antenna to direct the energy to the test area. We observed that this antenna was very successful in keeping the energy concentrated on the target, with the result that a short distance away, microwave power density was at a safe level. This permitted us to operate the IR camera during heating.

However, we also determined that spatial and temporal variations in heating by solar energy exceeded the temperature fluctuations we were trying to observe to detect buried objects. Furthermore, the surface signatures were severely contaminated by the presence of surface irregularities and vegetation. These issues led us to modify the model reported in the previous work to accommodate surface roughness. Future work with the model will also accommodate solar heating and vegetation.

2. EXPERIMENTS

The outdoor experiments were performed at the Army's Cold Regions Research and Engineering Laboratory (CRREL), in Hanover, NH, on 29-30 June 1999. Two cameras were in use; a modified Mitsubishi IR focal-plane array (320X240 pixels), and a Boeing microbolometer. A key element of the project was the use of an elliptical reflector for the microwave energy. In previous research, the microwave source, a 2.45-GHz. magnetron from an industrial microwave oven, attached to a homemade copper feed horn, was enclosed in a copper-screen shield which prevented microwave energy from leaking into the surroundings. Infrared images taken through the screen were not satisfactory, and it was necessary to stop heating and open the enclosure in order to take pictures. The upper left portion Figure 1 shows the test site with the reflector and feed horn mounted on a gantry crane. The feed horn is at one focus of the ellipsoid, 0.5 meters from the vertex, while the other focus, 4.5 meters from the vertex, is placed approximately at the ground surface. Because the ellipsoid is large (1.5 meters across), the radiation pattern from the feed horn is contained within the reflector, the power at the edges of the reflector is almost zero, and sidelobes are insignificant. The center right portion of the figure shows the two infrared cameras mounted on tripods. Survey measurements taken continuously during the heating experiments revealed an exposure level at and beyond the rail seen at the base of the tripods of less than the detection limit of our power meter, a fraction of a milliwatt per square centimeter. During experiments, personnel were kept outside this rail although measurements of power were still below the safety standard at locations considerably closer to the heated area.

Three mine simulants, shown in the right-hand side of Figure 1, were buried just below the surface near the focus of the ellipsoid. Because the antenna performed better than expected, we only heated the area containing one of the simulants. The grass in the area where the mines were buried was very dense, as shown in the left panel of the figure.

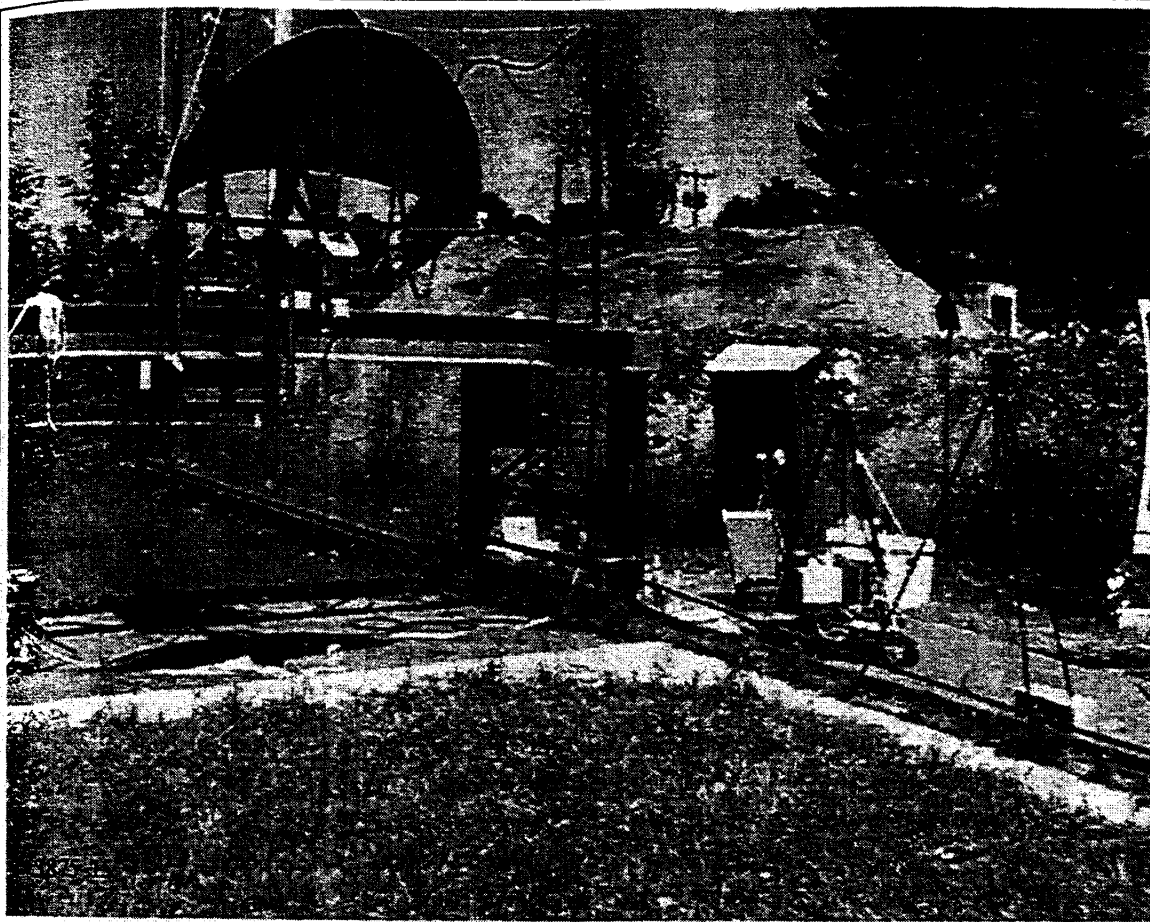


Figure 1 — Outdoor Test Site at Cold Regions Research and Engineering Laboratory, with Equipment in Place for Microwave-Enhanced Infrared Thermography.

3875-15

Solar heating of about 1000 Watts per square meter is often mentioned as sufficient for infrared detection of buried objects such as landmines. Typically, one waits until the diurnal heating or cooling reaches a maximum, and then examines an infrared image of the scene. In addition, the presence of clouds can dramatically alter the temporal distribution of solar heating. Figure 3 shows an example, taken from the two days of our test. On the first day, solar irradiance was only about half the normal level, and, although there were some short-term fluctuations associated with clouds, and there were some showers during the day, the overall heating was less than would occur on a sunny day. The temperature

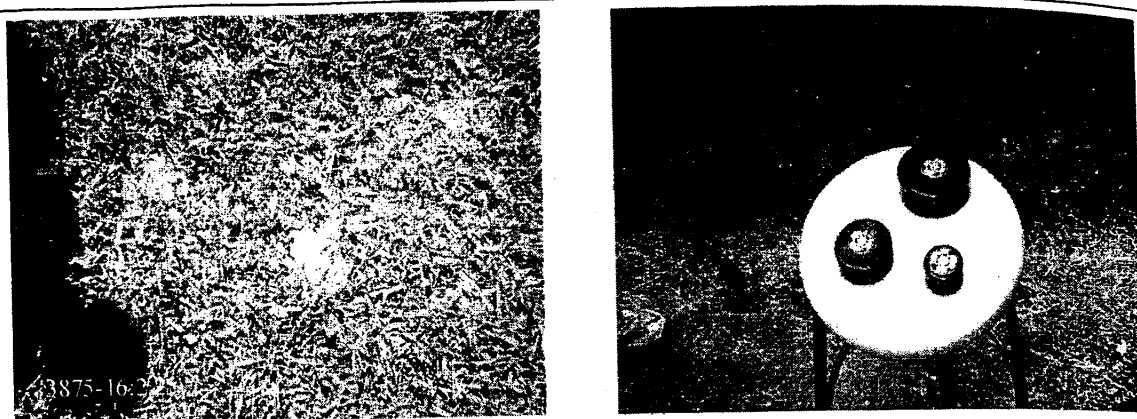


Figure 2 — Area Where Mine Simulants Were Buried (Left), and the Simulants (Right).

3875-16

of the ground was relatively stable at the time of our experiment, as shown by the top curve of Figure 4, which indicates the surface temperature as a function of time. In both figures, the time axis is hours from midnight local time (0400Z) on 29 June, extending through two days, until the end of 30 June. Figure 5 shows an infrared image from the first experiment. The mine location clearly shows a level of heating greater than that of the surrounding environment, but there are a number of clutter signals in the surrounding area with temperatures equally high. Furthermore, the mine signature is not particularly well defined, probably because of the clutter caused by the vegetation. This can affect the microwave heating directly because of the absorption of the microwave energy, and indirectly, by modifying the water content of the soil by varying the exposure to sunlight. More importantly, it can affect the temperature reported by the infrared camera. In some cases, the camera will see through gaps in the vegetation, while in other pixels it will see leaves of grass. Some pixels may include some of each. Clearly further work in processing is required.

The second test occurred at a time of intense increase in solar irradiance in the visible, as morning clouds began to clear. Note the increasing surface temperature following the slight flattening during the cloudy period. Note that the high rate of increase in temperature is seen at depths of greater than 10 centimeters. This presents the greatest challenge for observing microwave-enhanced infrared signatures, because the ground temperature is changing rapidly, and the rate of change is also changing rapidly. Thermocouple readings on the ground showed unpredictable behavior over several Kelvins, depending on

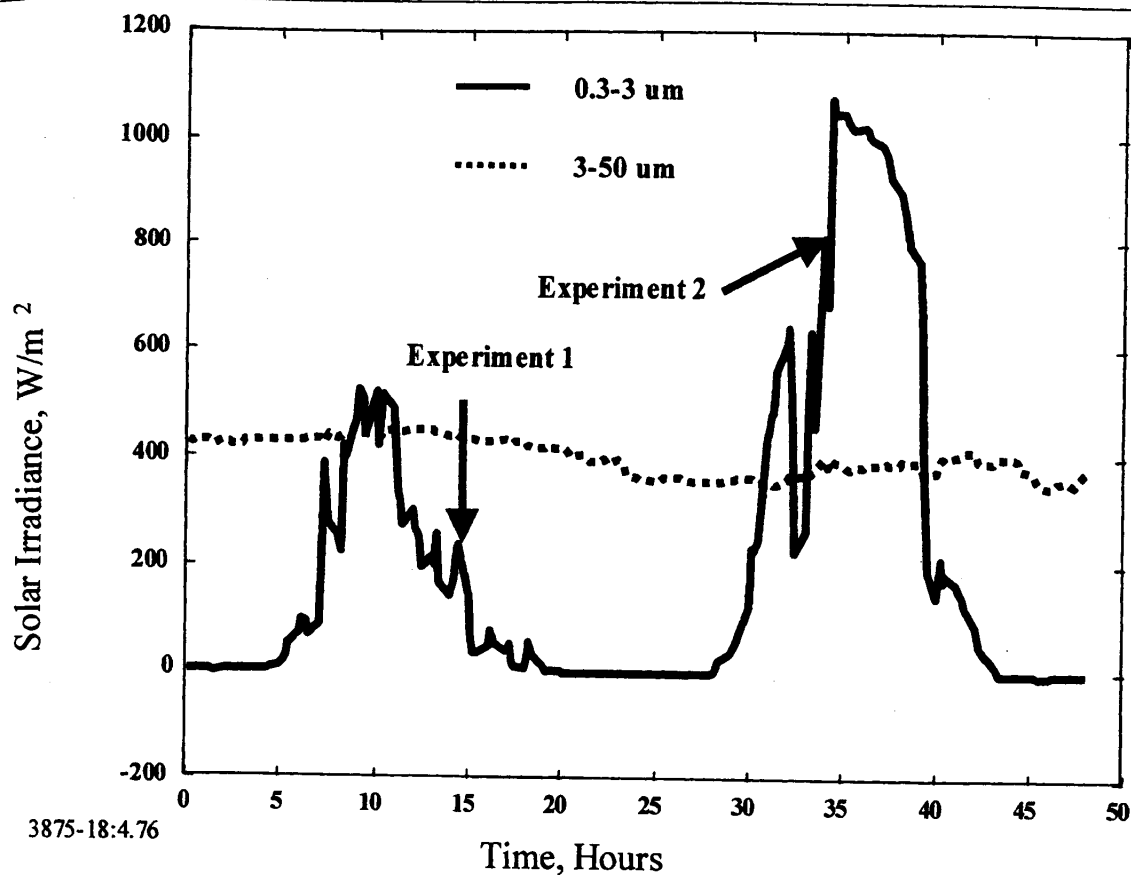


Figure 3 — Time History of Solar Irradiance in 2π Steradians, at the Test Site.

3875-18

probe placement, shading by grass, and changes in sunlight and soil composition. In general these changes were larger than those we expected to observe from microwave heating. Figure 6 shows an image taken during this second experiment. Although there is a suggestion of increased heating in the area near the mine, it is evident that there are much higher temperatures in other regions. Any attempt to make measurements based on microwave heating will be require extensive processing to remove the effects of such clutter. It should be pointed out that these problems arise for both passive solar and microwave-assisted heating scenarios.

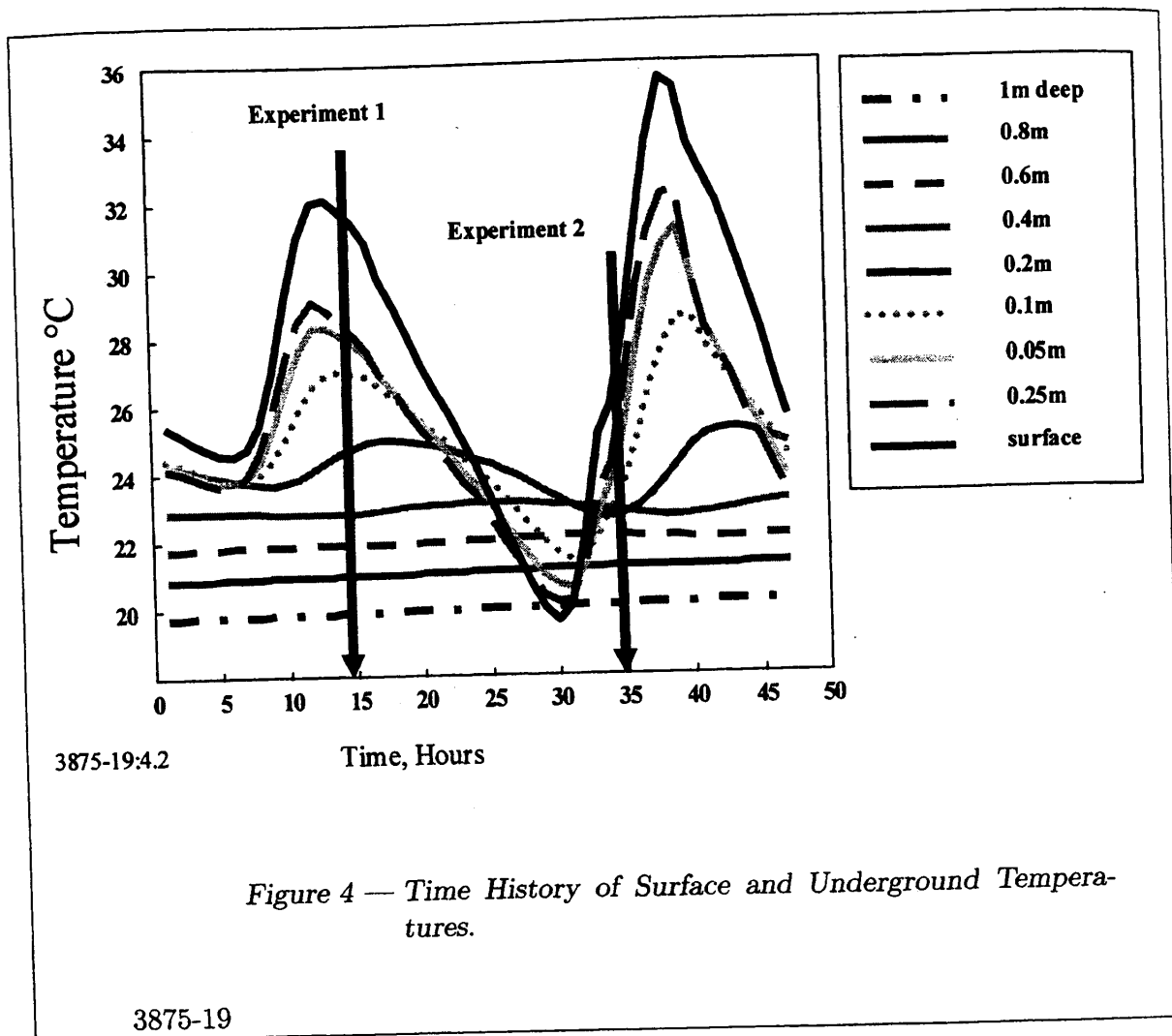
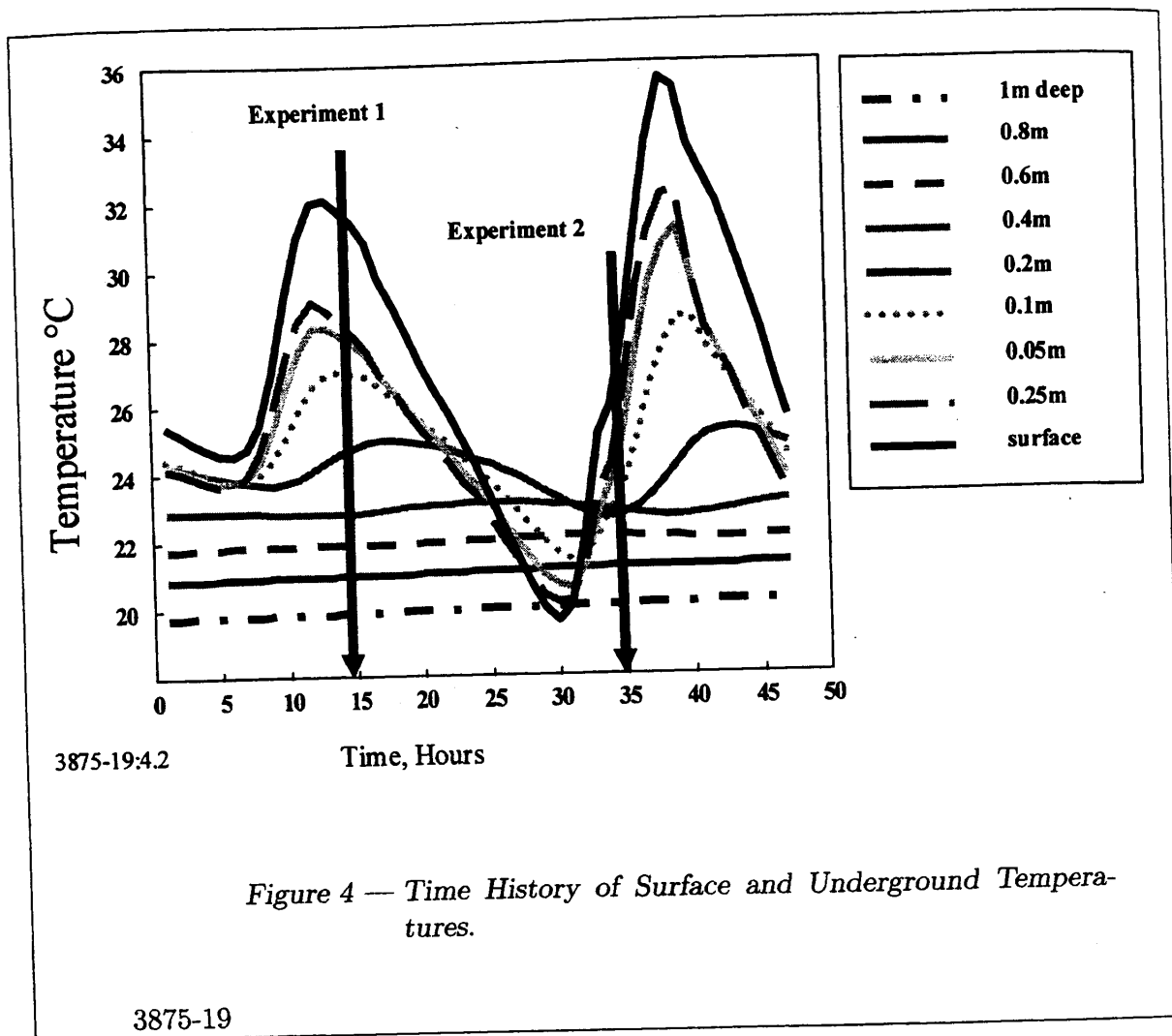


Figure 4 — Time History of Surface and Underground Temperatures.

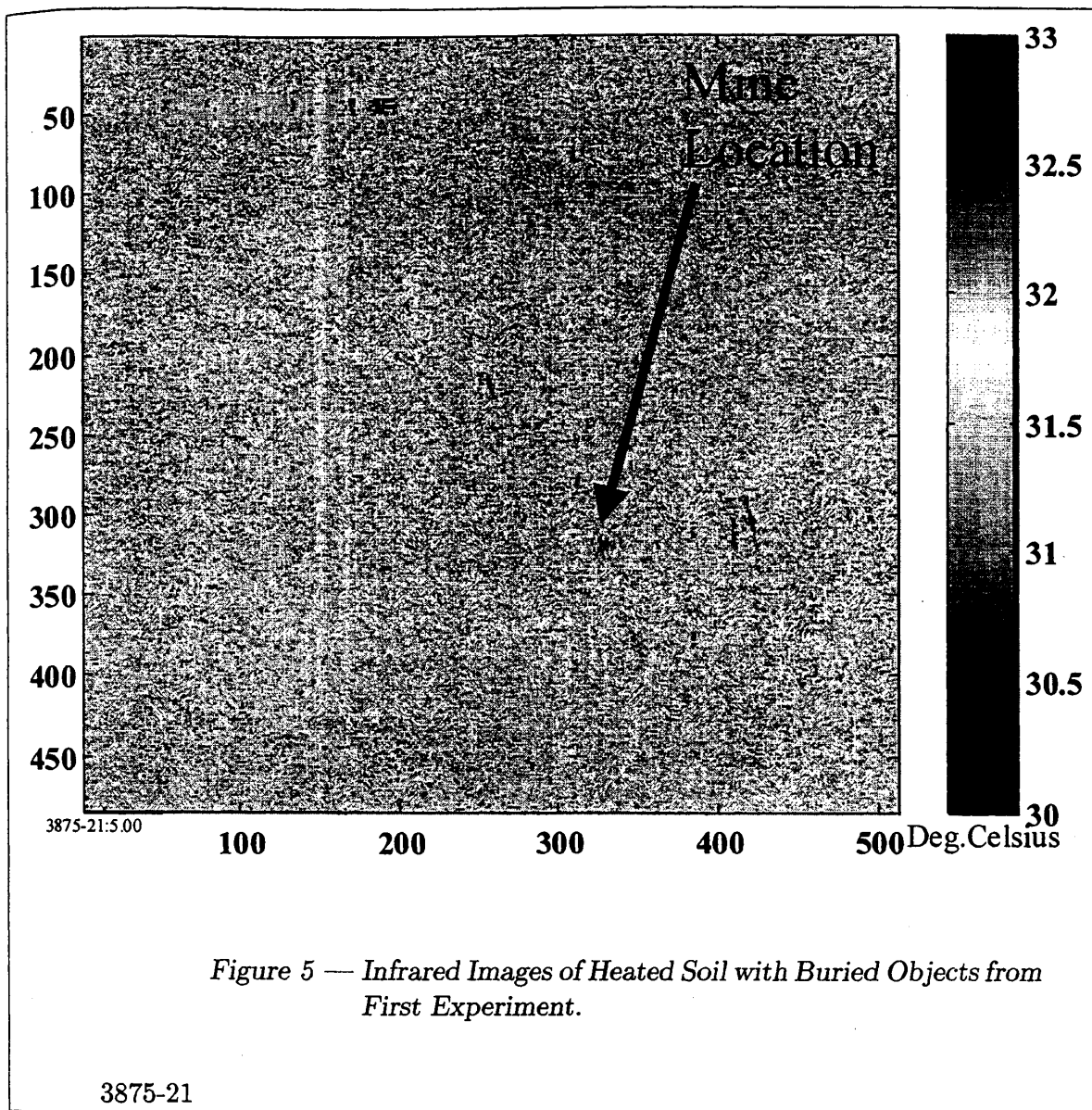
2. MODEL DEVELOPMENT

In the previous work, we reported a model for microwave heating of landmines buried in soil, based on a finite-difference, frequency-domain solution of the wave equation and a finite-difference solution of the thermal diffusion equation. In light of the results reported in the above section, we see the need to expand our model to address rough surfaces, vegetation, and solar energy. The first step is the inclusion of a rough surface. Figure 7 shows a result for thermal diffusion in the presence of a landmine (TNT) under a rough surface in Puerto Rican clay loam, for a microwave heating at 2.45 GHz. Because the mine is less conductive than the surrounding soil at this frequency, it heats less. At this point, the microwave model has not been implemented, so the flat-surface model is used, with data set to zero in the air above the actual rough surface. Thus the electromagnetic distribution of energy is not correct but, given that distribution, the thermal behavior is

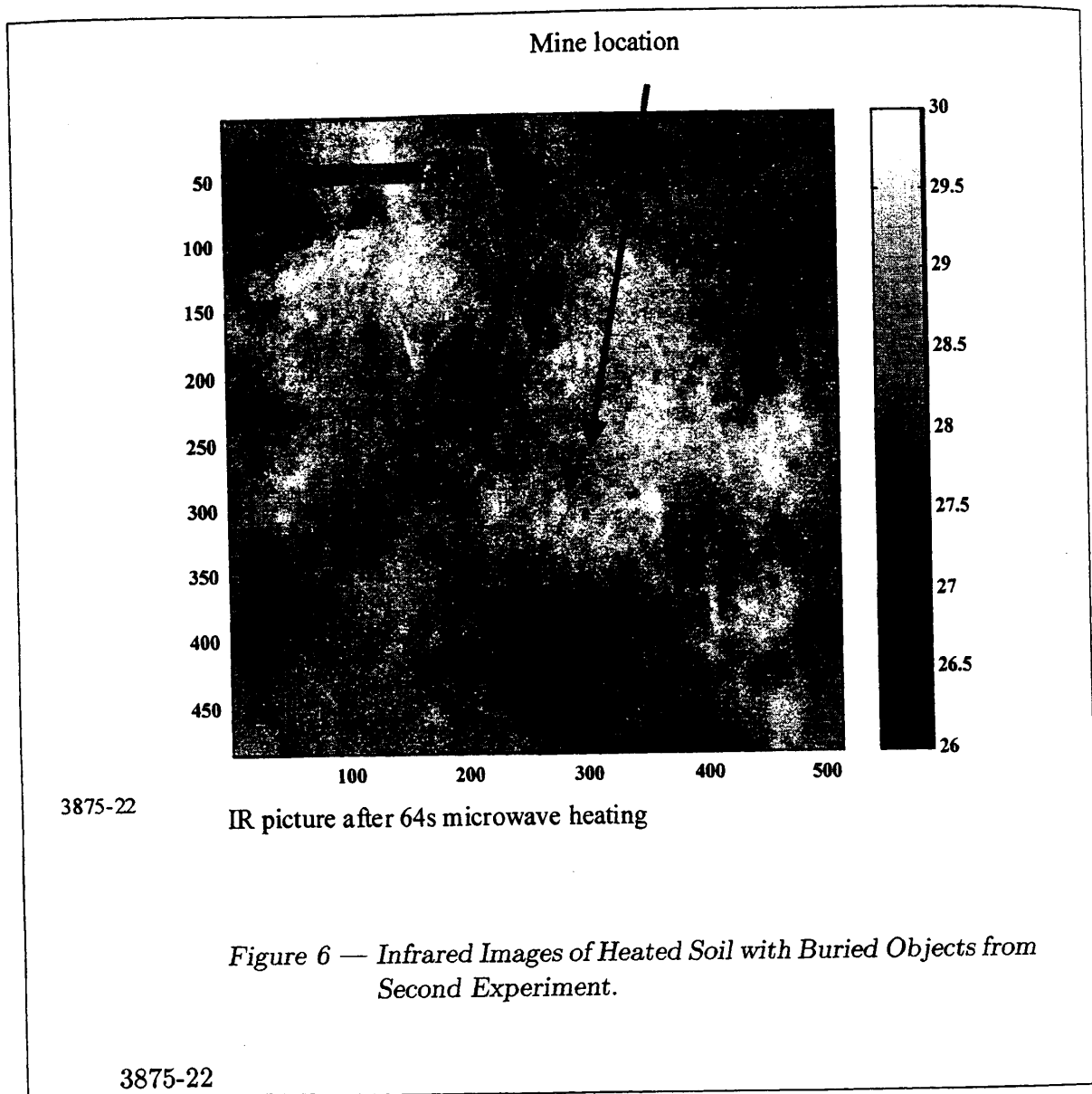


2. MODEL DEVELOPMENT

In the previous work, we reported a model for microwave heating of landmines buried in soil, based on a finite-difference, frequency-domain solution of the wave equation and a finite-difference solution of the thermal diffusion equation. In light of the results reported in the above section, we see the need to expand our model to address rough surfaces, vegetation, and solar energy. The first step is the inclusion of a rough surface. Figure 7 shows a result for thermal diffusion in the presence of a landmine (TNT) under a rough surface in Puerto Rican clay loam, for a microwave heating at 2.45 GHz. Because the mine is less conductive than the surrounding soil at this frequency, it heats less. At this point, the microwave model has not been implemented, so the flat-surface model is used, with data set to zero in the air above the actual rough surface. Thus the electromagnetic distribution of energy is not correct but, given that distribution, the thermal behavior is



correct. It is evident that the thermal diffusion process is influenced by the rough surface, but that the mine signature is still visible at the surface as it was in the case of a smooth surface. The two panels on the left show color plots of temperature for the cross-section during heating, while those on the right show the same plots during cooling. In the lower left plot, the mine has begun to heat through conduction from the surrounding hotter soil. As the mine heats up, a hot spot is observed in the local high point in the surface above the mine. Later this hot spot disappears through conduction and convection, and the shape of the mine is evident in a cool spot, although the mine continues to increase in temperature.



4. CONCLUSIONS AND FUTURE PLANS

We have shown that microwave heating of soil can be used to enhance infrared signatures of buried objects. Both solar and microwave-enhanced signatures are influenced by surface irregularities, vegetation, and variations in solar heating. At times of high solar heating, the microwave enhancement can be masked by clutter. Vegetation masks the transfer of energy in and out of the soil. Nevertheless, bare spots can be observed through the vegetation, and imaging algorithms may be developed to capitalize on these bare spots. The model will be completed to account for vegetation and solar heating. Experiments

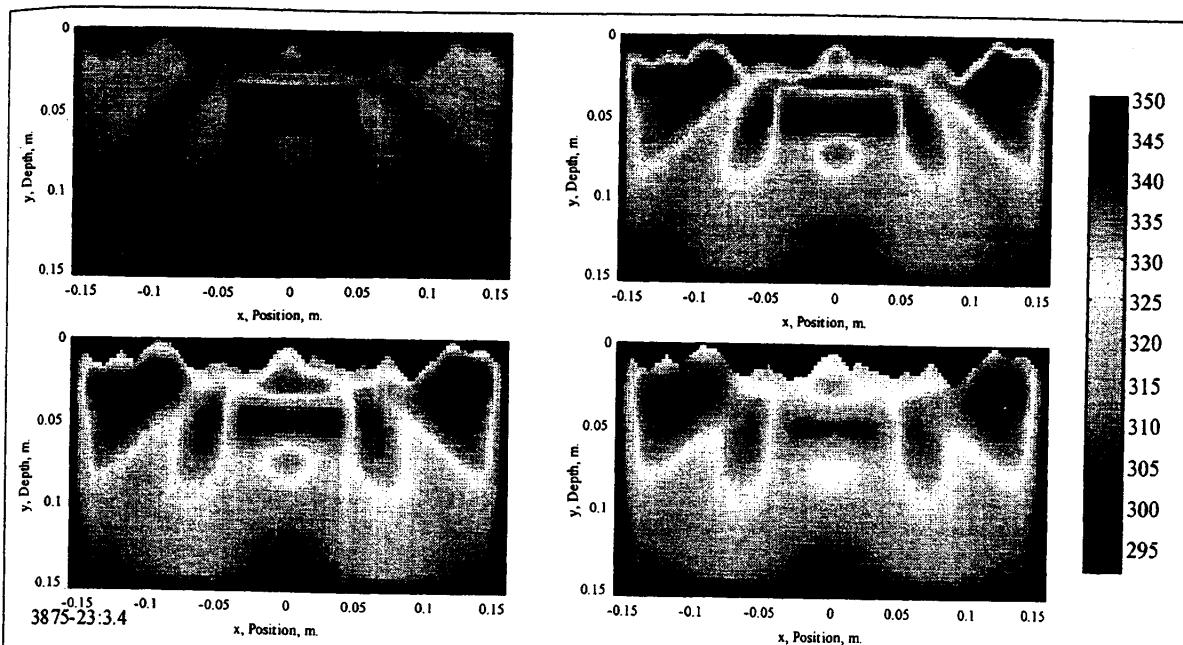


Figure 7 — Preliminary Results from Model for Microwave Heating of Ground with a Rough Surface.

3875-23

and analysis will begin to address whether the use of different wavelengths and fusion with other sensing modalities can provide help in reducing the effects of clutter. If these issues can be addressed, microwave-enhanced infrared thermography can provide a valuable tool for detection of buried objects.

5. ACKNOWLEDGMENT

We wish to thank Dr. Gary Koh of the U.S. Army Cold Regions Research and Engineering Laboratory, in Hanover, New Hampshire, for the use of his test site, supporting instrumentation, and for his help during the test program. We also thank Arnold Dean of GeoCenters for providing the elliptical reflector. This work was sponsored by the OSD MURI Program under Army Research Office Grant number DAAG55-97-1-0013.

6. REFERENCES

DiMarzio, Charles A., Carey M. Rappaport, and Li Wen, "Microwave-Enhanced Infrared Thermography," *Detection and Remediation Technologies for Mines and Minelike Targets III*, SPIE Aerosense Volume 3392, September 1998. Pp. 1103-1110.

DiMarzio, Charles A., Li Wen, Carey M. Rappaport, Gerhard O. Sauermann, and Herman E. Scott, "Microwave-Enhanced Infrared Thermography," *Detection and Remediation Technologies for Mines and Minelike Targets IV*, Proc. SPIE 3710. Presented at AeroSense, Orlando, FL. April 1999.

Some Approaches to Infrared Spectroscopy for Detection of Buried Objects

by

Charles A. DiMarzio
Center for Electromagnetics Research
Northeastern University
Boston, Massachusetts 02115

Tuan Vo-Dinh
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, Tennessee 37831-6101

and
Herman E. Scott
Aerodyne Research, Inc.
45 Manning Road
Billerica, MA 01821

ABSTRACT

Detection of buried objects presents a formidable challenge which requires many different approaches. Infrared imaging has proven its versatility in a number of applications. Recent advances in technology have opened the door for spectroscopic imaging systems which can produce images of reflectivity or emissivity as a function of two spatial dimensions and wavelength. These imagers have been largely unexploited for detection of buried and surface-laid landmines.

Several promising opportunities exist for this application in different parts of the infrared spectrum. Variations in soil moisture content, vegetation condition, and soil composition may well be related to the presence of shallow-buried objects. In addition, polarimetric signatures appear useful in detecting man-made objects on the surface and may even help in detecting buried objects.

This paper will explore both the feasibility of using infrared spectral imagery in the 1-to-2.5 and 8-to-12 micrometer infrared bands to detect surface-laid and buried objects.

1. INTRODUCTION

The detection of landmines is widely recognized as an extremely difficult challenge, requiring multiple sensing strategies with different operating scenarios, sensor modalities, and algorithms. In this work, we explore the use of hyperspectral imaging in two bands of the infrared spectrum. We consider the short-wavelength infrared band from 1 to 2.5 micrometers, in which water and soil contamination are detectable using scattered sunlight, and the long-wavelength IR band from 8 to 12 micrometers, in which material properties and surface characteristics can be determined from emissions from the surface. In the latter case, we also consider polarimetric measurements.

Recent advances in technology allow nearly simultaneous imaging and spectral characterization in what is called hyperspectral imaging. A dispersive element is placed in front of a camera and multiple images are collected. In one configuration, the dispersive element acts as a filter, allowing the camera to capture an image at only one wavelength. Successive pictures are collected at different wavelengths. In another, the optical system causes the camera to image a single line in space, for example, in the horizontal direction on the camera, while the dispersive element distributes the contributions from different wavelengths across the camera in the vertical direction. If the camera is scanned or mounted on a vehicle, successive pictures build a two-dimensional image in a "pushbroom" mode. In either case, a three-dimensional array of data is obtained, in which the coordinates correspond to two spatial dimensions and wavelength. Conventional spectroscopy can be performed on each spatial pixel, or more sophisticated processing can be performed on the full three-dimensional array. Furthermore, data can be collected in polarimetric channels to extract further information from the scene.

In preparation for using hyperspectral imaging in the detection of landmines and similar objects, we explore the spectra of relevant materials to determine what signatures may be expected in conditions related to demining. Here we report on some studies in two wavelength bands.

2. SHORT-WAVELENGTH-IR MEASUREMENTS

In the short-wavelength infrared band, self-emission is low, and the spectrum of an object is determined by reflected sunlight. The reflection spectrum can be used to determine composition of a material. Frequently soil composition varies strongly with location. Disturbance of the soil in laying a buried mine will result in deeper soil being brought to the surface and the spectral difference between the disturbed region and the surroundings will remain for a long time. Given the many natural variations which can be detected in soil spectra, this technique is not sufficient to locate landmines by itself, but is likely to be useful in conjunction with others, particular to address the issue of false

A laboratory spectrometer was modified to hold soil samples in a reflective mode. Soil samples were placed in the holder and measured from 1.1 to 2.5 micrometers. Figure 1 shows two soil spectra with different water content. Soil 1 is a loam with 10%

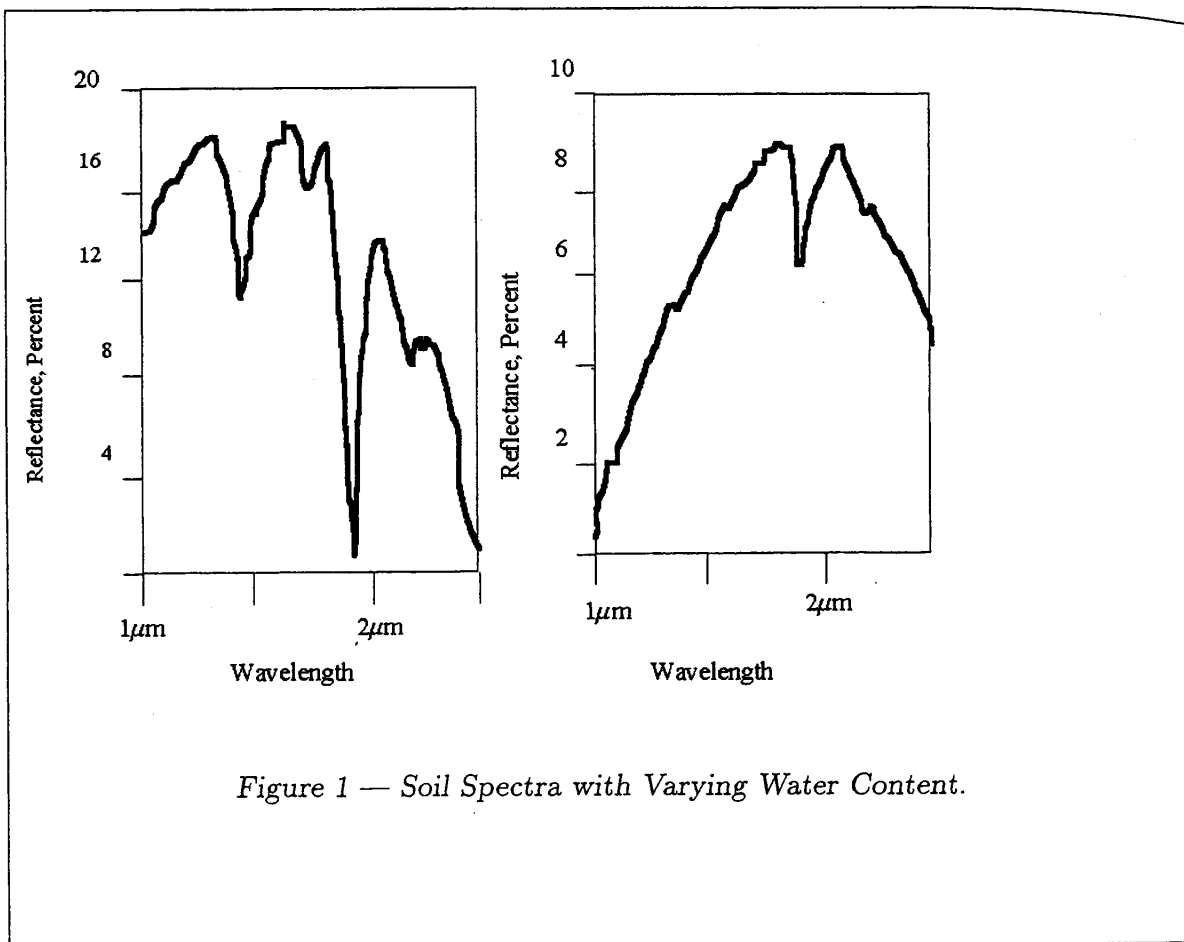


Figure 1 — Soil Spectra with Varying Water Content.

clay, 50% sand and 40% silt, while Soil 2 has 10% clay, 65% sand and 25% silt. There are significant differences in the content of several elements as well. The difference in absorption around 1.9 micrometers is a measure of the water content. In soil 1, several other water absorption lines become apparent. Soil 1 has an overall higher reflection. The overall reflectivity is probably not a useful parameter in detecting burial of objects, as wide ranges of variation are to be expected, and in operational scenarios, these results would be affected by lighting conditions. What may prove useful is the comparison of adjacent regions. Thus, hyperspectral imaging will permit simultaneous detection of the spatial variations in overall reflectivity and spectral features.

The emerging technology of acousto-optical tunable filters may permit the development of hand-held hyperspectral cameras to detect soil moisture, spatial changes in soil composition, and perhaps changes in the condition of vegetation, which can indicate the presence of objects below the surface. Used in conjunction with other, more established demining detectors, this technology could help to reduce the false-alarm rate and thus improve the speed of the overall demining process.

3. LONG-WAVELENGTH-IR MEASUREMENTS

In the long-wavelength infrared band, the contribution of scattered sunlight is low and the spectrum of an object is determined by its temperature and emissivity. Infrared thermography can detect changes in these parameters and, combined with polarimetry, can detect different surface characteristics. The emissivity of a surface depends on its composition, shape, and orientation. In the case of soil, transmission does not occur, so conservation of energy dictates that the sum of reflectivity and absorption equal unity.

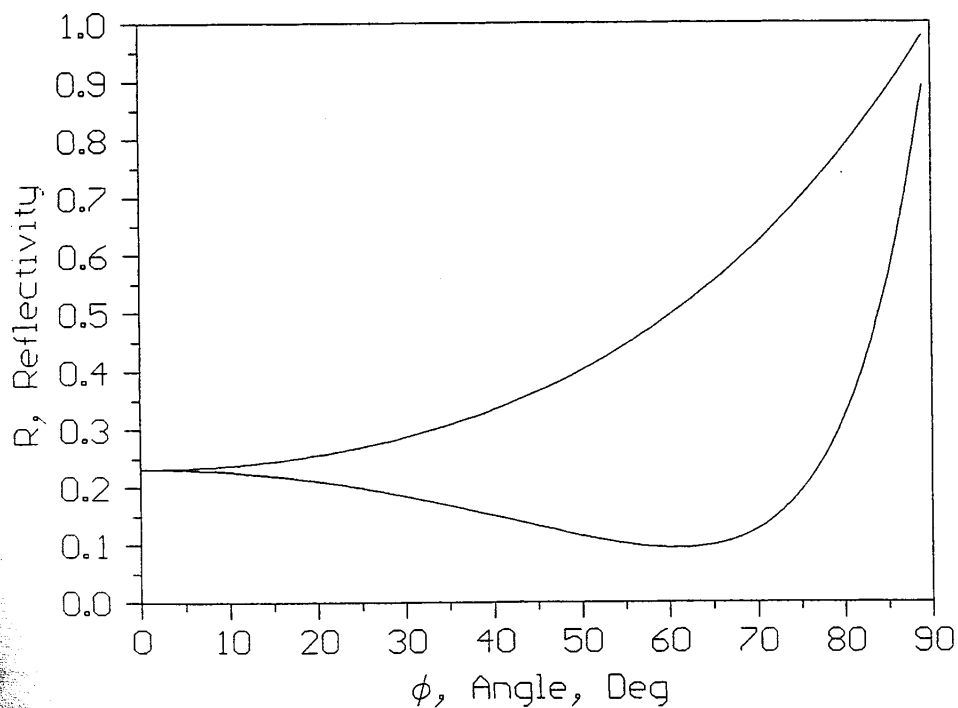


Figure 2 — *Fresnel Reflection from a Lossy Medium.*

Furthermore, the laws of thermodynamics require that the absorptivity equal the emissivity. Thus, emissivity is related to reflectivity by

$$\epsilon_x(\lambda, \theta) = A_x(\lambda, \theta) = 1 - R_x(\lambda, \theta),$$

where ϵ is the emissivity, A is the absorptivity, R is the reflectivity, and all are functions of the wavelength λ , the angle of observation θ , and the state of polarization, x . As an example, Figure 2 shows the Fresnel reflectivity of a surface having an index of refraction equal to $1.5 + 1.25i$. At the grazing angles typical of forward-looking mine detection, a significant difference exists between the two polarizations. This plot is for a smooth surface typical of a man-made object. For rough surfaces, the results become more complicated, involving details of the surface texture.

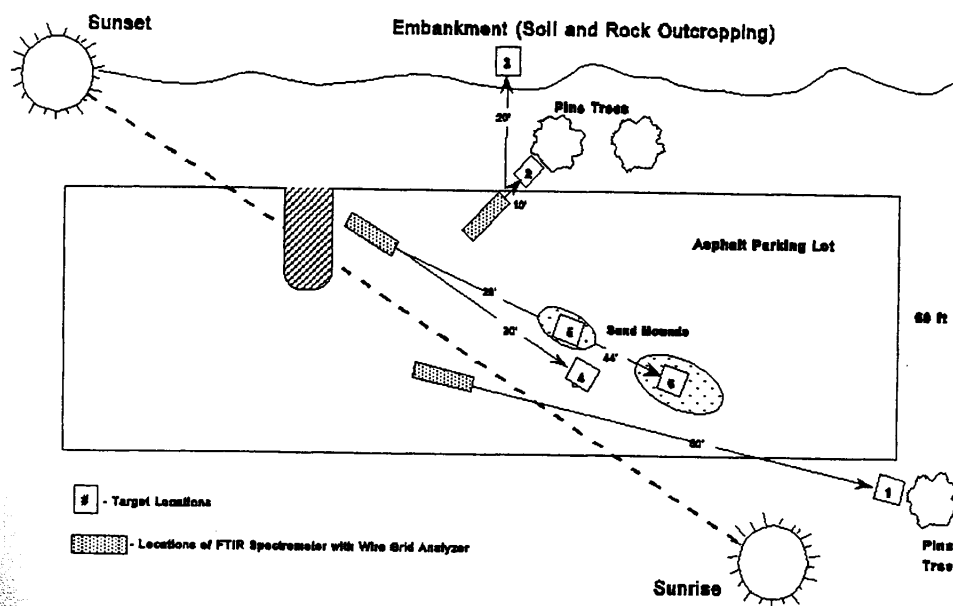
Spectral measurements were made using a FTIR spectrometer with a wire-grid polarizer before the lens against realistic target scenes on the test range shown in Figure 3.

Polarimetric difference images showed greatest contrast at polarization angles of zero and 60 degrees. Figure 4 shows results for a plastic frisbee on a sand background. The left panel shows the two spectra and the right panel shows their difference. The difference is particularly significant in view of the fact that the plastic only subtends 7 percent of the field of view. Interestingly, the polarimetric signature depends strongly on wavelength, and a polarimetric sensor could see a positive, negative, or zero signal, depending on the wavelength band.

Table 1 summarizes the polarimetric data. When the polarimetric signatures are converted to degree of linear polarization (DOLP), and corrected for the fraction of the field of view that is filled by the target, it is evident that strong polarimetric signatures exist for several targets.

Table 1
Key Polarimetric Results

	Sandy Asphalt	Tan CARC	Frisbee	Green CARC
DOLP (Tgt and Bkg)	0.02	0.02	0.01	0.02
Fractional FOV	?	0.24	0.07	0.15
Angle Normal to LOS	83	59	75	69
DOLP (Tgt Only)	?	0.08	0.14	0.13



ARS-68-03 14

Figure 3 — Outdoor Test Range.

Polarimetric differences for a tan CARC (chemical-agent resistant coating) target are shown in Figure 5. In this case, the CARC target is placed against sandy asphalt, which has a net polarimetric signature, integrated over the spectrum, as indicated by the degree of linear polarization shown in Table 1, similar to that of the CARC. However, the spectral signatures are noticeably different, as are the spectral polarimetric signatures, suggesting that both polarimetry and spectral data are important.

The possibility also exists for detecting soil disturbances associated with mine burial in a spectral band. Figure 6 shows spectra obtained in the laboratory for a "chunk" of soil removed from near the test range and for the same soil after being pulverized. The chunk was placed on an IRTAN window in the laboratory spectrometer, and is not considered representative of soil which will be encountered *in situ*. Nevertheless these results suggest that soil characteristics are dependent upon recent activity. These results

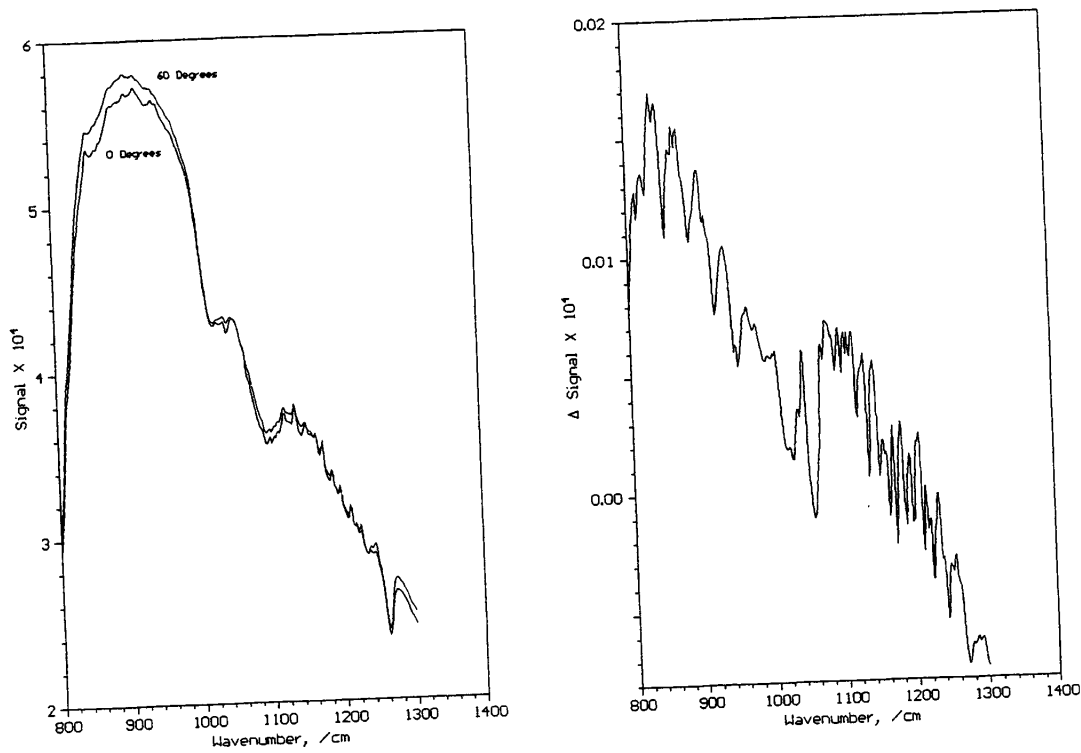


Figure 4 — Polarimetric Spectra for Plastic Frisbee.

are consistent with recent data in the literature [Johnson et. al.], which suggest that the increased reflectivity in the 8 to 10 micrometer region of the spectrum is caused by an increase in the number of small particles which are created when the soil is disturbed, but are later removed through erosion by wind and water. Polarimetric signatures of these samples were not investigated, but such a study is proposed for the future.

In conclusion, polarimetric signatures in the band from 8 to 14 micrometers are useful in detecting surface objects. Hyperspectral images are also useful and appear to be complementary. Combined polarimetric, hyperspectral FIR imaging may yield additional information about landmines scattered on the surface. Preliminary measurements suggest that this technique can be used to detect subsurface landmines as well, indirectly by measuring soil disturbances.

Future plans include, in the short term, additional measurements and polarimetric images using an uncooled microbolometer camera and a wire-grid polarizer. In the long term, filters will be added to produce hyperspectral, polarimetric images. More careful studies of different types of surfaces and soils will continue, using the existing FTIR instrument.

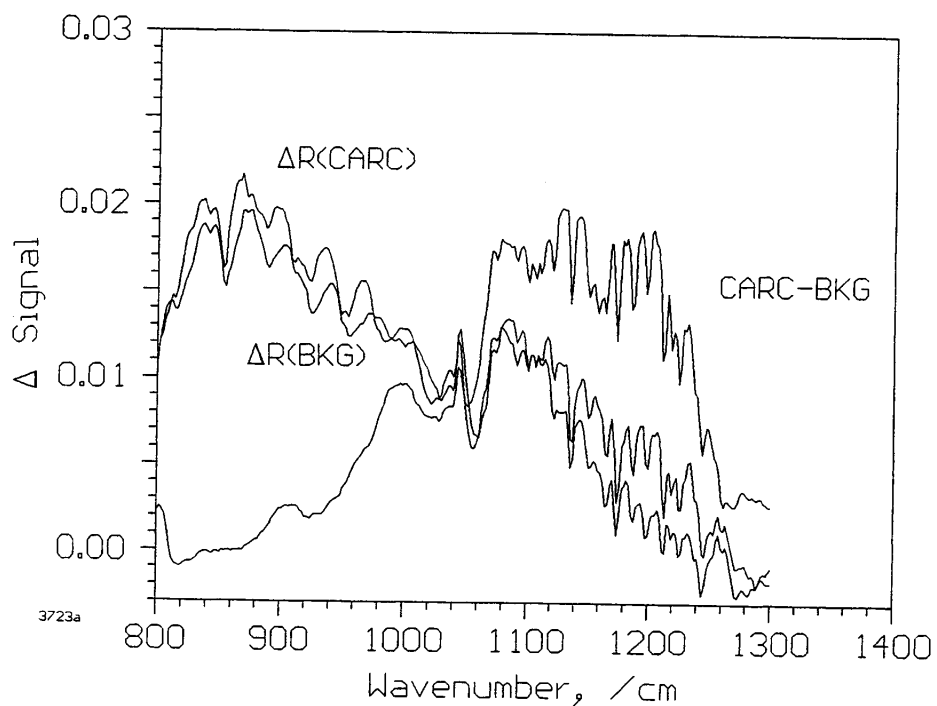


Figure 5 — Polarimetric Spectra for Tan CARC.

4. ACKNOWLEDGEMENT

This work was sponsored by Grant number DAAG55-97-1-0013 from the Army Research Office, Research Triangle Park, North Carolina to Northeastern University.

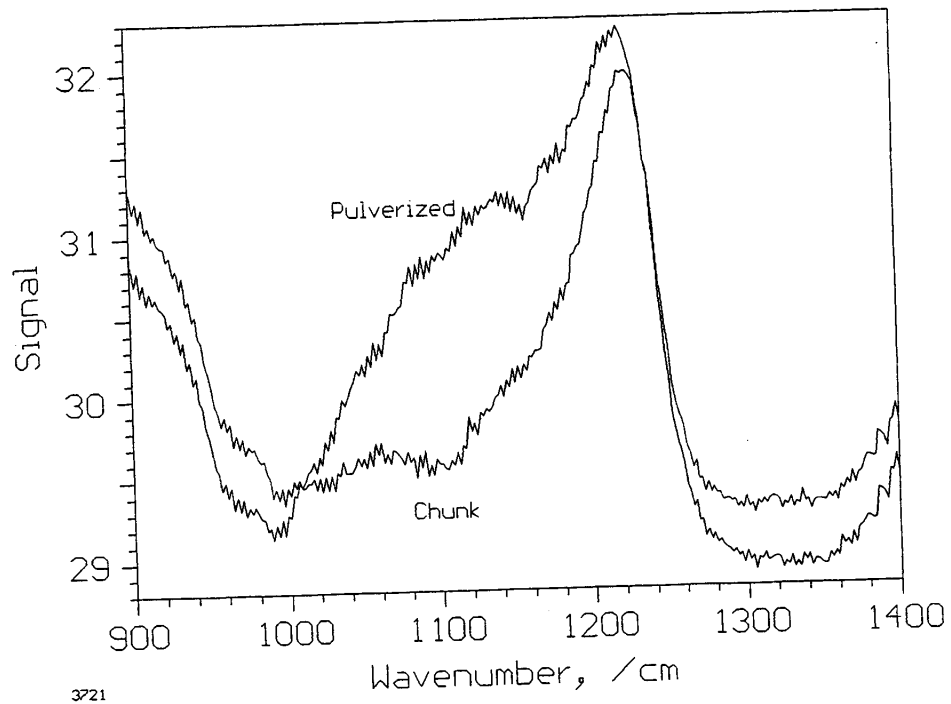


Figure 6 — Reflectance Spectra for Soil, Before and After Pulverizing.

5. REFERENCES

Johnson, et. al. "Infrared Measurements of Pristine and Disturbed Soils," *Remote Sensing of the Environment* 64, Pp. 34-46. 1998.

Laser-Induced Acoustic Waves for Measurement and Imaging

by

Wen Li

Ronald A. Roy*

Robin O. Cleveland*

Lawrence J. Berg**

Charles A. DiMarzio

Center for Electromagnetics Research

Northeastern University

235 Forsyth Building

Boston, Massachusetts 02115

*Aero & Mech Engineering

Boston University

110 Cummington Street

Boston, Massachusetts 02215

**Laser Science Inc.

8E Forge Parkway

Franklin MA 02038

ABSTRACT

A short pulse of laser light can act as a source of acoustic energy for acoustic imaging. Although there are a number of mechanisms by which the light pulse may generate sound, all require a pulse of high peak power density and short duration. In this work, we address examples where the material is highly absorbing at the laser wavelength, and the sound is generated near the surface. In these cases, there exist two different mechanisms which can convert the light to sound. The first is heating followed by expansion, and the second is generation of a plasma in the air above the surface. In the first case, sound generation occurs in the medium of interest and the energy efficiency can be very high, in the sense that no reflection losses occur. We present two applications from our own research.

1. INTRODUCTION

Acoustic imaging is useful in many different applications, ranging from the everyday example of a carpenter locating a stud by tapping the wall, to advanced ultrasound imaging of the body. A common requirement in all applications is the coupling of energy from a sound source into the material of interest. A pulsed laser offers several advantages as a

source of acoustic energy. First, it is remote. The laser can be a long way from the material to be tested [Li, et al.]. If the acoustic detection is accomplished with a laser vibrometer, then the complete acoustic imaging task can be performed from a long distance away [DiMarzio, et al.], for example, on hostile targets. Because of the good resolution possible with lasers, it is possible to make remote measurements with good transverse resolution, limited only by the size of the optics. Second, it is well calibrated. Many lasers are capable of delivering repeatable pulses. Third, it may be very efficiently coupled. Depending on the exact mechanism of interaction, sound may be generated in the medium itself, eliminating the losses associated with reflection at the surface. This will be discussed in more detail later. Fourth, optical properties (eg. spectroscopic) may be sampled acoustically, providing novel information.

Generation of sound by light is covered by several different names corresponding to different physical phenomena. We will begin with a cursory review some of the better known examples, before turning to the applications of interest in this paper. In an early application [Oraevsky], described as the photo-acoustic effect, light impinges on a sample of gas. If the wavelength of the light is tuned so that the gas is absorbing, a sound is produced, and detected by a sensitive microphone. The shape and duration of the acoustic pulse is determined by the interaction of the light and the medium. The strength of the acoustic signal, plotted as a function of the laser wavelength, yields the absorption spectrum of the gas.

In opto-acoustic imaging, light is incident on biological tissue. The tissue is weakly absorbing so that absorption occurs along the path of the light. In this case, an acoustic signal is generated which propagates to the detectors. The shape of the acoustic pulse is determined by the travel time from each point along the laser beam to the detectors and the absorption profile of the tissue at the laser wavelength. The primary absorber of light in tissue is hemoglobin, which has a highly structured optical absorption spectrum, which changes when oxygen is bound to it. As a result opto-acoustic imaging can detect blood vessels, discriminating between veins and arteries [Oraevsky, et al.].

In contrast to these examples, in the applications to be discussed in the present paper, the optical absorption is strong, and most of the light is absorbed within a few optical wavelengths of the surface of the medium. Possible mechanisms of interaction are optical radiation pressure, various forms of heating and expansion, and plasma generation.

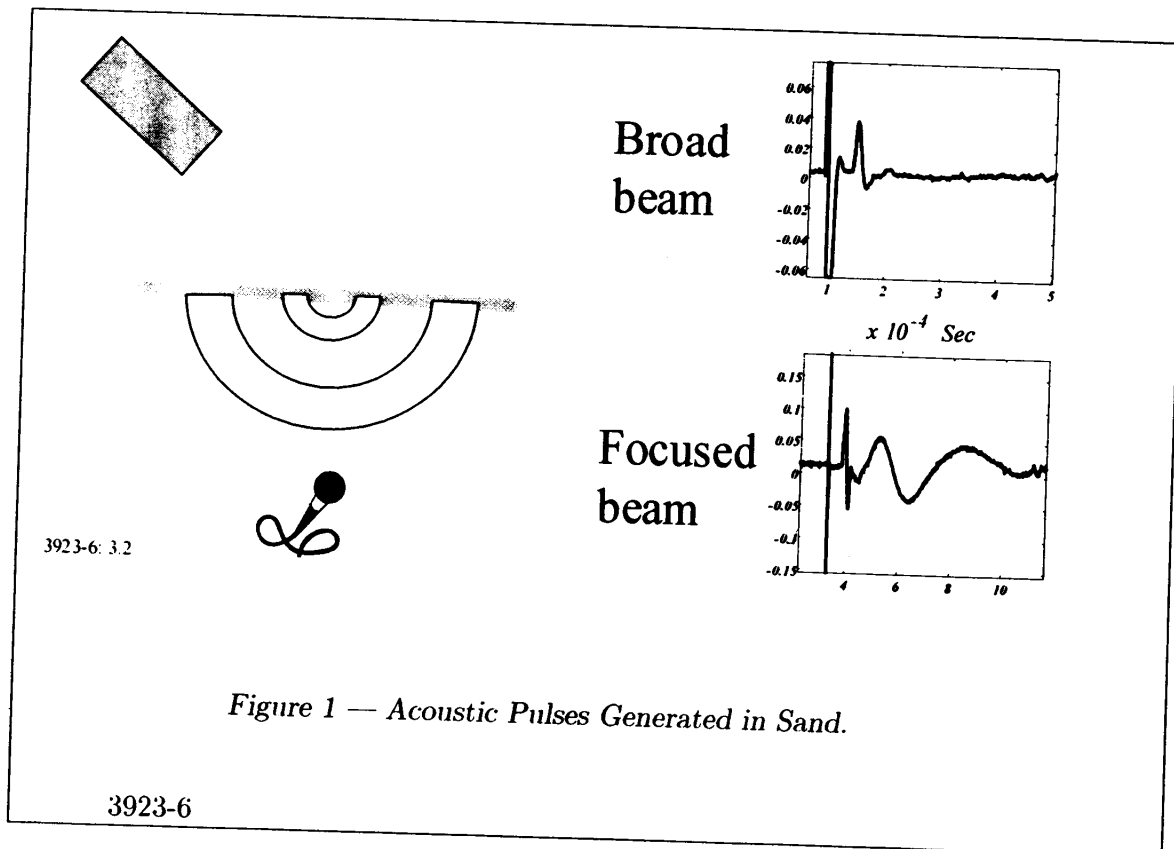
2. DETECTION OF SHALLOW-BURIED LANDMINES

Pulses of light have been used experimentally to detect buried objects such as landmines [Li et. al.]. Figure 1 shows an example of an acoustic pulse generated in sand by a 100 nanosecond, 150 millijoule pulse of light from a carbon-dioxide laser. The receiver is a B & K model 8103 hydrophone with a bandwidth of 100 kHz, buried a few centimeters below the surface. As the sand is strongly absorbing at the 10.6-micrometer wavelength of

the laser, most of the light is absorbed, rather than reflected. The absorption depth is of the order of the wavelength, so most of the light is absorbed in the first few micrometers of soil. This light heats the sand faster than thermal diffusion can cool it, with the result that the sand expands at the point where the laser is incident, and produces an acoustic pulse. In most acoustic imaging systems [Sabatier and Xiang; Scott et al.; Donskoy and Ekimov], the sound is generated in the air above the soil, and most of it is reflected because of the impedance mismatch at the interface. In the case of laser-induced sound, at least in this configuration, with 150 mJ of energy in a 1-cm diameter, the light is absorbed in the sand, the acoustic pulse is generated there, and a substantial fraction of the laser energy is converted to sound in the sand. Although the laser pulse is only 100 nanoseconds, the resulting sound pulse width is a few microseconds, because of the thermal properties of the sand.

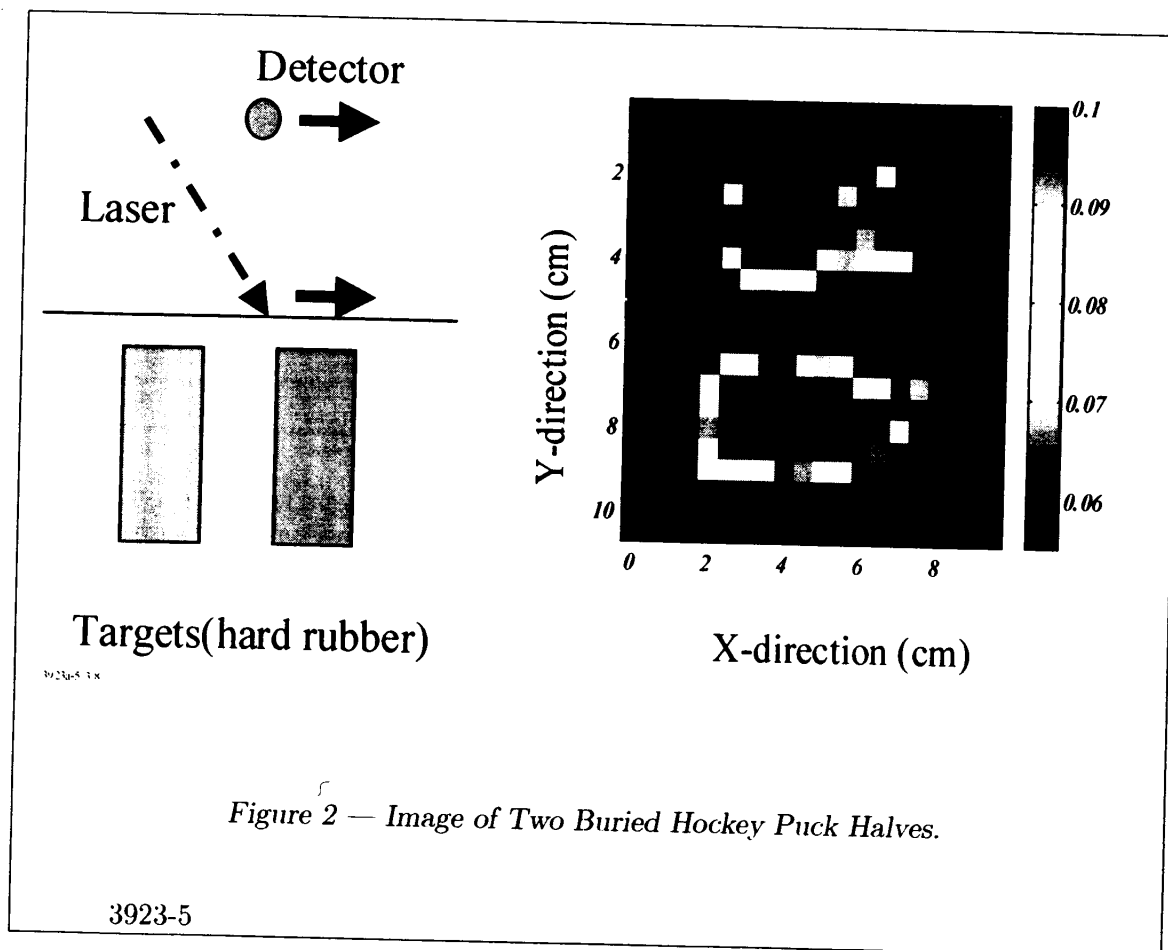
The times and energy densities discussed above are consistent with other qualitative observations we have made over the years [unpublished]. For example, one of us used this effect to locate a carbon-dioxide laser beam in order to align a laser radar, having 20 millijoules of energy in a pulse length of 2 microseconds and pulse-repetition frequency of 140 Hertz. The 140-Hz tone was readily apparent to the ear when the beam impinged on a variety of surfaces, even though the beam diameter was almost 30 centimeters. In contrast, a recent attempt at generating sound in sand with a laser pulse length in the tens of microseconds failed. In the latter case, although the pulse energy was sufficient, the longer pulse meant that power was delivered so slowly that thermal diffusion cooled the sand during the pulse, leading to a lower peak temperature and less expansion of the sand. Calculations have shown that radiation pressure of light produces a lower acoustic pulse than the heating effect under our conditions. As the pulse length becomes longer, of course, the heating effect becomes smaller, while the effect of radiation pressure remains substantially the same.

Figure 2 shows the result of scanning a laser beam over the surface of the sand in two dimensions, and processing the signals to determine the time of reflection from a pair of buried hockey puck halves. The color shows the strength of the strongest reflected signal. Because the ground location was known approximately, the ground signal was removed, and the strongest remaining signal was used. In addition to the signal strength, this technique also provides depth information, which is useful in imaging as well. The receiver was a barium titanate transducer with a center frequency of 29 kHz and a bandwidth of about 5 kHz, located above the surface. Other receivers such as the 100-kHz hydrophone mentioned earlier, and an inexpensive microphone were also used, but this transducer produced the best results. The narrow bandwidth removed low-frequency noise, provided high sensitivity and reduced noise. The frequency was sufficiently high to provide good depth information and transverse resolution, but not so high as to be absorbed before reaching the buried object. The sand box was moved while the laser and microphone remained in the same locations. The halves are oriented so that their flat surfaces are toward the surface of the sand, and are placed about 1 cm apart, and about 1 centimeter deep. This



depth is particularly important for the demining application, because ground-penetrating radars have difficulty discriminating shallow buried objects from surface returns. Note that the shape and space between the pucks are easily resolved. The pulse is composed of frequencies up to tens of kilohertz, so the wavelengths are on the order of a few centimeters and longer. Thus, this is an example of near-field imaging.

Increasing the power density raises the peak temperature and results in an increase in performance, but this increase has its limits. We have performed experiments with different laser spot sizes, and found that as the spot size is reduced, other effects begin to become important. One of these appears to be the generation of a plasma in the air above the sand. This is evident by the appearance of visible light being emitted from the contact point, in spite of the incident light having a wavelength well into the infrared. To an observer of the experiment, the resulting sound is impressive. A 18-mJ pulse produces a "pop" easily detected a few feet away and a 150-mJ pulse produces an explosive sound which attracts the attention of anyone in the laboratory. However, as impressive as this may be, the detection of buried objects has not been as successful. In this case, the sound is generated in the air above the sand. As seen in Figure 1, this produces a very different acoustic pulse with lower frequencies, and a slower sound speed. This wave is less absorbed



in the sand, but it must first pass through the air-sand interface where the loss to reflection is high. Furthermore, the plasma itself absorbs the light, and the plasma itself thus moves up away from the sand during the light pulse. Thus it appears that the best approach for detection of buried objects may be to use a pulse shorter than a few microseconds, with energy density just below that sufficient to initiate the plasma. This conclusion is offered somewhat tentatively, because the use of lower frequencies shows great promise for mine detection in view of their greater penetration. Because this application is based on near-field imaging, the resolution is not determined by the wavelength, and the normal benefits of smaller wavelengths are not relevant. Research is currently under way to determine the best compromise between the ability to transfer energy to the medium and the frequency distribution of that energy. To date, most of the detection has been done by detecting the acoustic signal directly, but in the future, a laser vibrometer will make the technique completely remote, and perhaps more sensitive.

3. NON-DESTRUCTIVE TESTING OF PAPER

The second application of Laser Induced Acoustic (LIA) sensing is in the field of Non-Destructive Testing (NDT). When employed for NDT, LIA used to generate an acoustic pulse, either through a process of localized thermal expansion, an ablation shock wave or both. The generated acoustic pulse then either propagates through the test medium or on the surface. Surface detection is performed through interferometric techniques or Doppler shift detection via a coherent laser sensor.

Detection of the acoustic wave reveals information about the test medium in several potential ways. Interruption or reflection of the acoustic wave can pinpoint cracks or localized faults - such methods have been employed to interrogate the quality of aircraft skins or other critical components. Detection of the magnitude and time delay at a point removed from the initiation point of the acoustic wave can reveal information related to the material properties, such as density, stiffness and other mechanical properties.

Recent work performed at the Lawrence Berkeley Laboratory with the support of the U.S. Department of Energy, in collaboration with one of us, at Laser Science, Inc., has used the latter technique to measure the quality of paper, non-destructively and *in-situ*. This differs from the mine-detection application in several ways. It is a single-point measurement, and imaging is not required. The resulting acoustic wave is inherently two-dimensional, being contained within the paper. The first parameter to be measured is the speed of the wave and depth of penetration is not an issue, so high frequencies are of primary importance.

The key elements of paper quality include overall strength (tensile and compressive) fiber orientation and fiber strength. Classical methods to check these quality control parameters include cutting samples from the large paper rolls after the manufacturing process is completed and the large "webs" are shut off. The paper is then inspected and the manufacturing process modified to correct any errors and improve the quality of the paper. This is a time-consuming operation that costs money while wasting significant amounts of material and energy. Contact methods employing ultrasonics have been developed and employed that used transducers to generate and sense acoustic waves to monitor paper quality: However, the pressures needed to interface the transducers to the paper roll are only adequate for the heavier grades of paper and damage the lighter grades.

The non-destructive, and in this case totally non-contact, technique employing LIA uses a YAG or TEA CO₂ laser to generate a sub-microsecond laser pulse that is focused onto the paper that is streaming between the paper rollers in the web at 30 meters per second. As the pulse duration is quite short relative to the speed of the moving paper, the acoustic pulse that is generated is essentially a point source. In conjunction with the firing of the pulsed laser, a laser-based interferometer "stares" at a probe point on the paper which is at a fixed location (several millimeters) offset from the point where the acoustic pulse is generated. Because the paper is moving, the probe point (where the

acoustic pulse is detected) must actually travel at a speed equal to that of the moving paper to maintain the fixed offset between acoustic point source and probe point. This is accomplished through the use of a rotating mirror and optical encoder, which essentially freezes the paper motion relative to the probe point and synchronizes the pulsed laser.

The key parameter measured in this technique is the time delay between pulsed laser firing (start of acoustic pulse) and detection of the pulse at the probe point. Knowing then the fixed offset distance provides the acoustic velocity of the wave in the paper, which is then related to elastic properties of the medium. Empirical relationships have already been established to relate elasticity to properties of strength.

4. CONCLUSIONS

Laser-induced acoustic pulses can be used for a variety of applications in many different types of media. The mechanisms of interaction are many, and vary with the characteristics of the media, the wavelength of the laser light, the pulse length, total energy, and spot size. Different mechanisms generate different acoustic pulses which can be tailored to the application. Two examples have been considered in which the medium is highly absorbing. In both cases, the propagation time of the acoustic wave is of importance. In the landmine case, near-field imaging is possible because the objects of interest are a few wavelengths from the surface. In the second example, the goal is to measure the speed of sound in a thin medium. In both cases, the laser offers significant advantages over other acoustic imaging techniques because of its remote measurement capability with good transverse resolution, its repeatability, and its efficiency.

5. REFERENCES

- Oraevsky, Alexander A., Tutorial on Opto-Acoustic Imaging presented at *OSA Annual Meeting*, Baltimore, MD. October 1998.
- Oraevsky, Alexander A.; Esenaliev, Rinat O.; Jacques, Steven L.; Tittel, Frank K.; "Laser optic-acoustic tomography for medical diagnostics: principles," *Proc. SPIE Vol. 2676, Biomedical Sensing, Imaging, and Tracking Technologies I*, Robert A. Lieberman; Halina Podbielska; Tuan Vo-Dinh; Eds. April 1996. p. 22-31.
- Li, Wen. Stephen W. McKnight, Charles A. DiMarzio, and Ron Roy, "Laser-Induced Acoustic Detection of Shallow-Buried Objects," *Subsurface Sensors and Applications, Proc. SPIE 3752*, October 1999. Pp. 285-291.
- Ridgeway, et al. *Journal of Ultrasonics*, Vol. 37, 1999 pp.395-403.
- DiMarzio, Charles A., Wen Li, Lawrence J. Berg, and James Sabatier, "Toward a Laser-Based, Non Contact Acoustic Landmine Imager," *Proc. SPIE, Vol. 4038, Detection and Remediation Technologies for Mines and Minelike Targets, V*, Publication Pending in 2000.
- Sabatier, James M. and Ning Xiang, "Laser-Doppler-based acoustic-to-seismic detection of buried mines," *Proc. SPIE Vol. 3710, Detection and Remediation Technologies for Mines and Minelike Targets IV*, Abinash C. Dubey; James F. Harvey; J. Thomas Broach; Regina E. Dugan; Eds., August 1999. p. 215-222.
- Donskoy, D. M., and A.E. Ekimov, "Seismo-Acoustic Landmine Detection and Discrimination: Physical Modeling and Lab/Field Testing," *Proc. SPIE, Vol. 4038, Detection and Remediation Technologies for Mines and Minelike Targets, V*, Publication Pending in 2000.
- Scott, W. R. Jr., G. D. Larson and J. S. Martin, "Simultaneous Use of Elastic and Electromagnetic Waves for the Detection of Buried Landmines," *Proc. SPIE, Vol. 4038, Detection and Remediation Technologies for Mines and Minelike Targets, V*, Publication Pending in 2000.

Imaging Dielectric and Conductivity from GPR Measurements

Alan J. Witten

School of Geology and Geophysics, University of Oklahoma,
810 Energy Center, Norman, OK 73019-0628

ABSTRACT

It has been shown that, for non-conductive backgrounds and non-conductive spatial heterogeneities, two- and three-dimensional images of spatial variations in wave speed can be reconstructed from broadband ground penetrating radar (GPR) measurements. In diffraction tomography, the reconstructed image is the so-called object function which, for non-conducting heterogeneities, is one minus the square of the real refractive index. In cases where spatial variations in electrical conductivity exist, the object function is complex with real part related to relative changes in the real refractive index and the imaginary part representing the spatial variations in conductivity. Thus, by considering the complex object function, it is possible to reconstruct images of both wave speed and conductivity. The procedure presented here is used to reconstruct images of wave speed and conductivity for several buried targets.

Keywords: diffraction tomography, ground penetrating radar, wave speed, electrical conductivity

1. INTRODUCTION

Tomographic imaging has been applied in geophysics to wave-based data to reconstruct images of geologic structure and isolated buried objects. There are various imaging procedures ranging from straight ray algorithms to non-linear inversions. The particular type of imaging algorithm considered here is diffraction tomography¹. Geophysical diffraction tomography (GDT) was first considered for a multi-bistatic transmission mode measurement geometry using a single frequency and the synthesis of multiple direction of insonifying plane waves². This work has been extended to a multi-monostatic measurement geometry where wave sources and receivers are assumed to be co-located and moved in unison along a line or over a plane^{3,4}. Such a measurement geometry is typical of GPR measurements and, for this geometry, broadband information is used.

Traditionally, GDT has been used to reconstruct images of the spatial variations in wave speed. In many applications, the problem is one of target discrimination and, as such, there may exist many buried targets of similar size, shape, and wave speed while only a relatively small number of these targets are of interest. For example, land mines are relatively small and may appear as many other near-surface naturally-occurring or man-made objects. If the object of a geophysical investigation is to locate all possible land mines for subsequent remediation, the cost associated with the removal of all false positives could be prohibitive. An even worse situation is one where land mines exist in a host soil where the wave speed of the land mine is nearly that of the soil. In such cases, land mines may not be detected. It is clear that there is a need to develop better techniques for the discrimination of buried targets and one possible approach to this problem is to reconstruct images of wave speed and electrical conductivity of buried objects. By considering two physical properties, there is more information for target discrimination and less likelihood of a target being "invisible".

For transmission mode radio wave measurements, it is possible to reconstruct wave speed and electrical conductivity using straight ray algorithms. In these methods, perturbations in travel time are associated with spatial variations in wave speed while amplitude variations are associated with attenuation induced by spatial variations in electrical conductivity. Apart from the known limits of straight ray algorithms, there is an addition problem

associated with the separation of wave speed and attenuation from travel time and amplitude. The Green's function for wave propagation is complex so that there is a blending of real (propagation) and imaginary (attenuation) resulting in the travel time being corrupted by attenuation and vice versa. While diffraction tomography has its own limitations, it does not assume any explicit separation of travel time and amplitude and, subject to a weak scattering approximation, rigorously treats the blending of real and imaginary parts^{5,6}. By considering the imaging reconstruction to be complex, a method is established here whereby separate images of wave speed and electrical conductivity can be obtained.

2. MULTI-MONOSTATIC BROADBAND IMAGING

The GDT algorithm has previously been published where only spatial variations in wave speed were considered^{3,4}. Since the separate imaging of wave speed and conductivity is based on this algorithm, it is summarized here. Since the examples given later are based on one-dimensional measurements over a line on the ground surface, only the two-dimensional imaging algorithm is presented. This algorithm can easily be generalized to three dimensions for data acquired over a plane.

The reconstruction of wave speed from broadband multi-monostatic measurements is based on the inversion of the frequency domain wave equation subject to the Born approximation⁷ given by

$$u_s(\mathbf{r}) = k_0^2 \int d\mathbf{r}' G(|\mathbf{r} - \mathbf{r}'|) u_0(\mathbf{r}') O(\mathbf{r}'), \quad (1)$$

where u_s is the temporally Fourier transformed time-domain scattered field, G is the Green's function for the scalar Helmholtz equation, u_0 is the incident wave, for a non-conducting background and spatial heterogeneities, $k_0 = \omega \sqrt{\mu_0 \epsilon_0}$ is the background wavenumber for frequency ω , magnetic permeability μ_0 , and dielectric constant ϵ_0 , $O(\mathbf{x}) = 1 - (c_0^2/c^2(\mathbf{x}))$ is the object function that defines the spatial variations in wave speed $c(\mathbf{x}) = 1/\sqrt{\mu_0 \epsilon(\mathbf{x})}$ associated with spatial variations in dielectric $\epsilon(\mathbf{x})$ in terms of the background wave speed $c_0 = 1/\sqrt{\mu_0 \epsilon_0}$. It has been assumed in Eq. (1) that the incident field u_0 arriving at the receiver has been removed from the data either by appropriate time gating or filtering so that this component of the data is absent in u_s .

For a co-located point source and receiver, the incident field u_0 is given by $G(|\mathbf{r} - \mathbf{r}'|)$ so that Eq. (1) becomes

$$u_s(\mathbf{r}) = k_0^2 \int d\mathbf{r}' G^2(|\mathbf{r} - \mathbf{r}'|) O(\mathbf{r}'). \quad (2)$$

Assuming that the measurements are made at positions ℓ on a line in an x, z coordinate system $\mathbf{r} = (\ell, 0)$ where z is measured vertically upward, the plane wave expansion for the Green's function G can be used⁸, the data spatially Fourier transformed with respect to measurement location ℓ

$$\tilde{u}_s(\kappa) = \int d\ell e^{-i\kappa\ell} u_s(\ell, 0),$$

and a resulting integral approximated by stationary phase to give

$$\tilde{u}_s(\kappa) = \frac{ik_0 e^{i\pi/4}}{4\sqrt{2\pi} \sqrt{4k_0^2 - \kappa^2}} \tilde{O}(\kappa, \sqrt{4k_0^2 - \kappa^2}), \quad (3)$$

where

$$\tilde{O}(\kappa, \sqrt{4k_0^2 - \kappa^2}) = \int d\mathbf{x} e^{-i(\kappa x + \sqrt{4k_0^2 - \kappa^2} z)} \frac{O(\mathbf{x})}{\sqrt{|z|}}$$

and it is clear that Eq. (3) relates the one-dimensional spatial Fourier transform of the data to the two-dimensional spatial Fourier transform of the object function divided by the square root of the depth. This equation can be used to represent the Fourier transform of $O_1 = O/\sqrt{|z|}$ in terms of the data as

$$\tilde{O}_1(\mathbf{K}) = -\frac{i}{k} 4\sqrt{2\pi} \sqrt{4k_0^2 - \kappa^2} e^{-i\pi/4} \tilde{u}_s(\kappa), \quad (4)$$

where $\mathbf{K} = (\kappa, \sqrt{4k_0^2 - \kappa^2})$.

Images of object function and hence wave speed can be reconstructed by inverting the two-dimensional spatial Fourier transform given by Eq. (4). Typically, data is acquired over uniform spacings in ℓ and the spatial and temporal Fourier transforms of u_s result in \tilde{u}_s being known at uniform intervals in ω (or alternatively k_0) and κ . To avoid \mathbf{K} -space interpolations necessary to implement the inverse Fourier transform, this inverse transform is expressed as integrals over κ and k_0 to give

$$O(\mathbf{x}) = \mathcal{L}[\tilde{u}_s], \quad (5)$$

where

$$\mathcal{L}[\] = -\frac{4i\sqrt{2}}{\pi\sqrt{\pi}}\sqrt{|z|}e^{-i\pi/4} \int dk_0 \int \frac{d\kappa}{\sqrt{4k_0^2 - \kappa^2}} e^{i(\kappa x + \sqrt{4k_0^2 - \kappa^2} z)} [\], \quad (6)$$

and O is taken to be real. Figure 1 shows the \mathbf{K} -space coverage over which \tilde{O}_1 is known. From this coverage, expectations of the images resulting from the application of Eqs. (5) and (6) can be deduced. First note that there

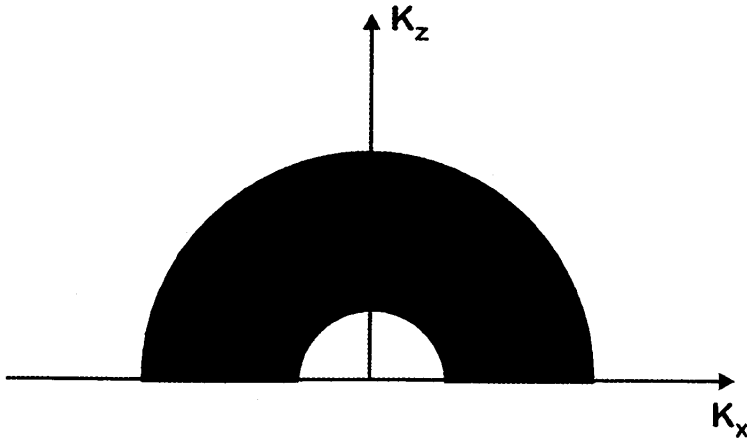


Figure 1: Illustration of the \mathbf{K} -space coverage of the multi-monostatic imaging algorithm where $\mathbf{K} = (K_x, K_z)$.

is no coverage at the origin. This implies that there is no spatial DC information present and the reconstructions will represent only the boundaries associated with changes in wave speed. This is common to linearized reflection mode imaging. Another artifact associated with the lack of \mathbf{K} -space coverage at the origin is the reconstruction of O will necessarily have a zero mean. Experience with this algorithm has established that, for features that are large compared to a wavelength, a zero mean is produced by filling the otherwise hollow interior of a feature with an opposite contrast. For targets that are small compared to a wavelength, a zero mean is produced by a ring around the feature having an opposite contrast. Concentric rings around the image of objects will occur as a result of bandwidth limitations. Although the coverage shown in Fig. 1 indicated information at $K_z = 0$ this can only be achieved when measurements extend out to $\ell = \pm\infty$. Since can never be the case, the lateral boundaries of features can never be reconstructed.

3. IMAGING OF WAVE SPEED AND ELECTRICAL CONDUCTIVITY

It can be shown that the procedure outlined in Sect. 2 can be directly used to separately image wave speed and electrical conductivity. In the more general case of a non-conducting background and a heterogeneity that has both

a difference in wave speed and electrical conductivity with respect to the background, the object function can be expressed as

$$O(\mathbf{x}) = 1 - \frac{k^2(\mathbf{x})}{k_0^2} = 1 - \frac{\mu_0 \epsilon(\mathbf{x}) \omega^2 + i \mu_0 \sigma(\mathbf{x}) \omega}{\mu_0 \epsilon_0 \omega^2}, \quad (7)$$

where $\epsilon(\mathbf{x})$ and $\sigma(\mathbf{x})$ are spatial variations in dielectric and electrical conductivity, respectively. It is clear from this representation that the object function is complex where $O = O_R + iO_I$ and

$$O_R(\mathbf{x}) = 1 - \frac{\epsilon(\mathbf{x})}{\epsilon_0} = 1 - \frac{c_0^2}{c^2(\mathbf{x})}, \quad (8)$$

and

$$O_I(\mathbf{x}) = \frac{\sigma(\mathbf{x})}{\epsilon_0 \omega}. \quad (9)$$

Similarly, the two-dimensional spatial Fourier transform of the object function is complex where, from Eqs. (8) and (9),

$$\tilde{O}_R(\mathbf{K}) = 1 - \frac{\tilde{\epsilon}(\mathbf{K})}{\epsilon_0}, \quad (10)$$

$$\tilde{O}_I(\mathbf{K}) = \frac{\tilde{\sigma}(\mathbf{K})}{\epsilon_0 \omega}, \quad (11)$$

and $\tilde{O} = \tilde{O}_R + i\tilde{O}_I$.

From Eq. (4), the real part of \tilde{O}_1 is

$$\tilde{O}_{1R}(\mathbf{K}) = \Re \left\{ -\frac{i}{k} 4\sqrt{2\pi} \sqrt{4k_0^2 - \kappa^2} e^{-i\pi/4} \tilde{u}_s(\kappa) \right\}$$

so that the image of wave speed variations is simply the real part of $O(\mathbf{x})$ given by Eqs. (5) and (6). Since the two-dimensional Fourier transform of the spatial variations in electrical conductivity relative to the background dielectric ϵ_0 is $\sigma(\mathbf{K})/\epsilon_0 = \omega \tilde{O}_I(\mathbf{K})$, the two-dimensional spatial Fourier transform of this relative conductivity divided by the depth is given by

$$\omega \tilde{O}_{1I}(\mathbf{K}) = -\omega \Im \left\{ \frac{i}{k} 4\sqrt{2\pi} \sqrt{4k_0^2 - \kappa^2} e^{-i\pi/4} \tilde{u}_s(\kappa) \right\}.$$

It is now clear that spatial variations in wave speed can be reconstructed using

$$1 - \frac{c_0^2}{c^2(\mathbf{x})} = \Re \{ \mathcal{L}[\tilde{u}_s] \}, \quad (12)$$

and spatial variations in relative electrical conductivity can be reconstructed using

$$\frac{\sigma(\mathbf{x})}{\epsilon_0} = \Im \{ \mathcal{L}[\omega \tilde{u}_s] \}, \quad (13)$$

where \mathcal{L} is defined in Eq. (6).

This procedure is only valid when the background is non-conducting. In cases where there is a background conductivity (σ_0), the background wavenumber $k_0 = \sqrt{\mu_0 \epsilon_0 \omega^2 + i \mu_0 \sigma_0 \omega}$ is complex and the integral transform given in Eq. (3) is not a Fourier transform. Although more complicated and less computationally efficient, separate imaging of spatial variations in wave speed and conductivity can still be accomplished.

4. RESULTS

The reconstruction algorithms for wave speed and electrical conductivity given by Eqs. (12) and (13) have been applied to broadband GPR data acquired at two sites. The first is a test site where 400 MHz center-frequency antennas were used over a variety of known shallowly buried objects of similar size and shape but composed of different

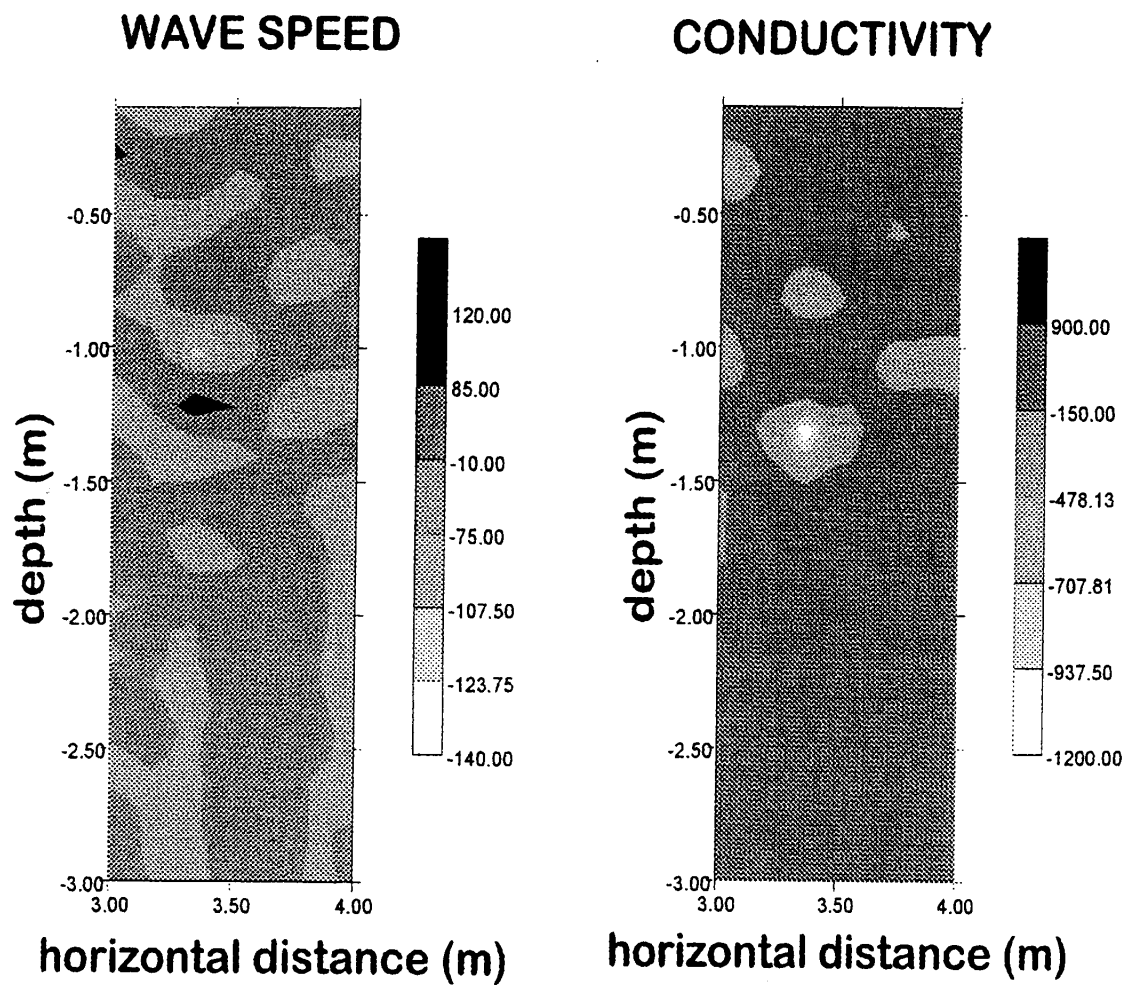


Figure 2: Gray scale image of wave speed and electrical conductivity for a polystyrene disk. Darker shades of gray are associated with larger values of wave speed and conductivity.

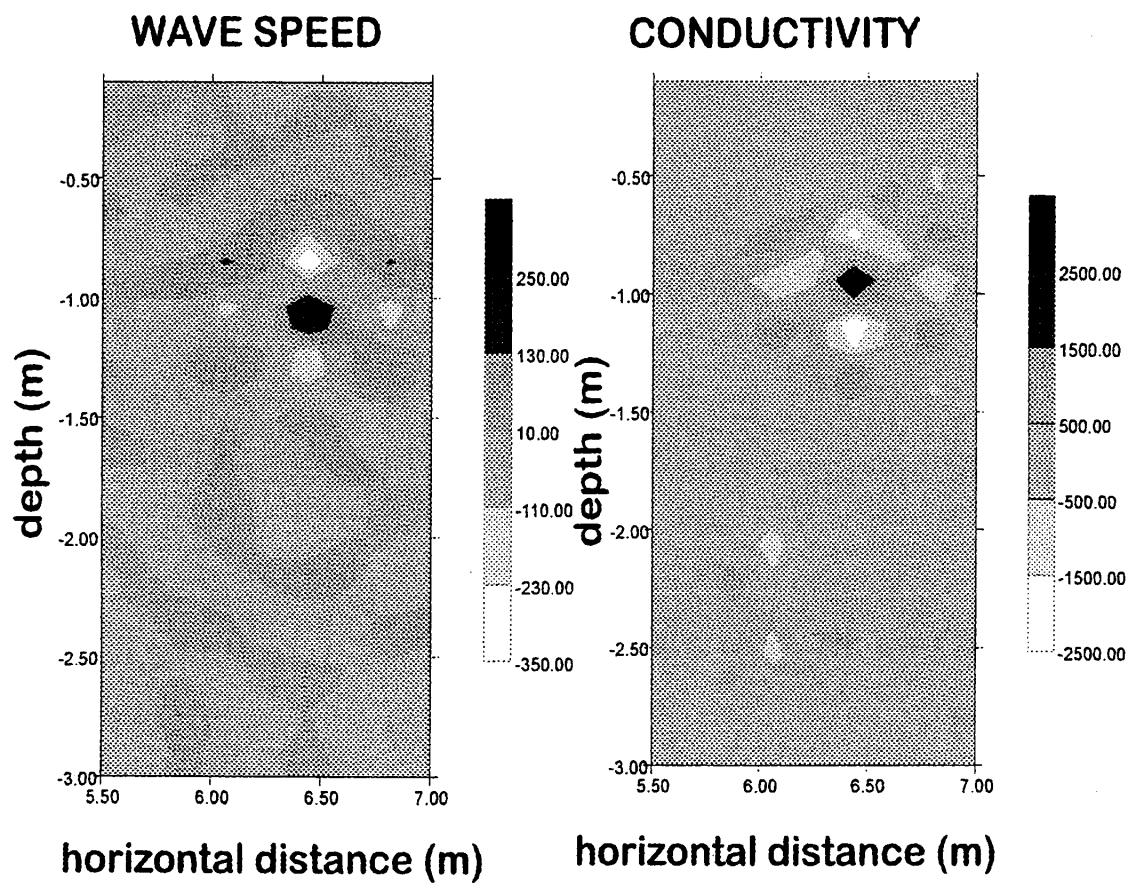


Figure 3: Gray scale image of wave speed and electrical conductivity for a concrete tube. Darker shades of gray are associated with larger values of wave speed and conductivity.

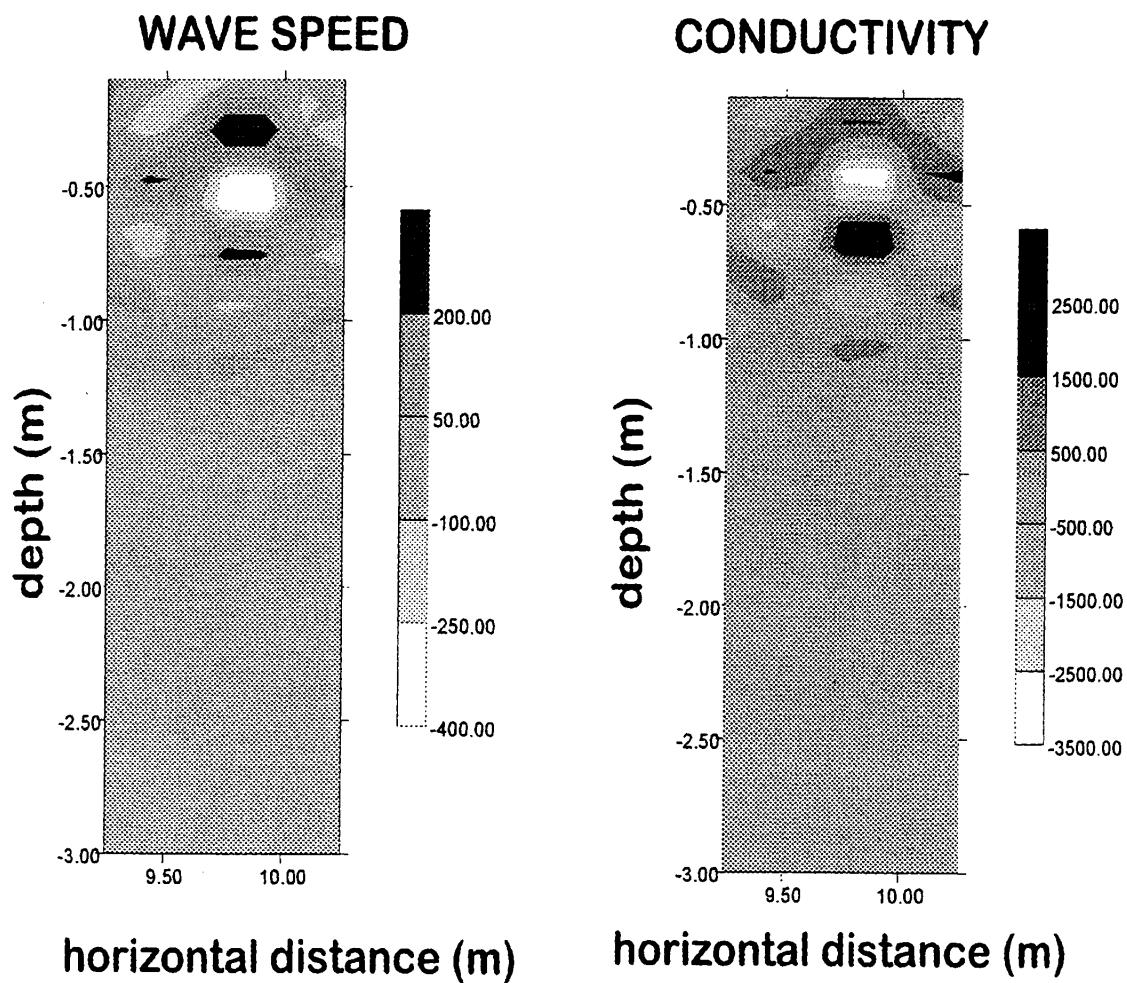


Figure 4: Gray scale image of wave speed and electrical conductivity for a iron tube. Darker shades of gray are associated with larger values of wave speed and conductivity.

materials. The second site contains a shallowly buried land mine of unknown type, size, and depth. At the land mine site, 1000 MHz center-frequency antennas were used.

Figures 2, 3, and 4 display reconstructed values of wave speed and conductivity as gray scales for three different targets at the test site. The images of the polystyrene disk (Fig. 2) and the iron tube (Fig. 4) yield the anticipated results that the polystyrene disk exhibits a high relative wave speed and low conductivity and the iron tube exhibits low relative wave speed and high conductivity. For the concrete tube, both the relative wave speed and conductivity are high. The sizes of three targets are small relative to wavelength and, as a result, the boundaries of the features blur together making them appear solid rather than hollow as would be expected since there is no spatial DC information. The ringing evident in some images is a result of both the zero-mran requirement and bandwidth limitations.

Figure 4 presents gray scale images of wave speed and electrical conductivity, similar to those given in Figs. 2, 3, and 4; for the data acquired at the land mine site. In this figure, the presence and location of the mine is obvious.

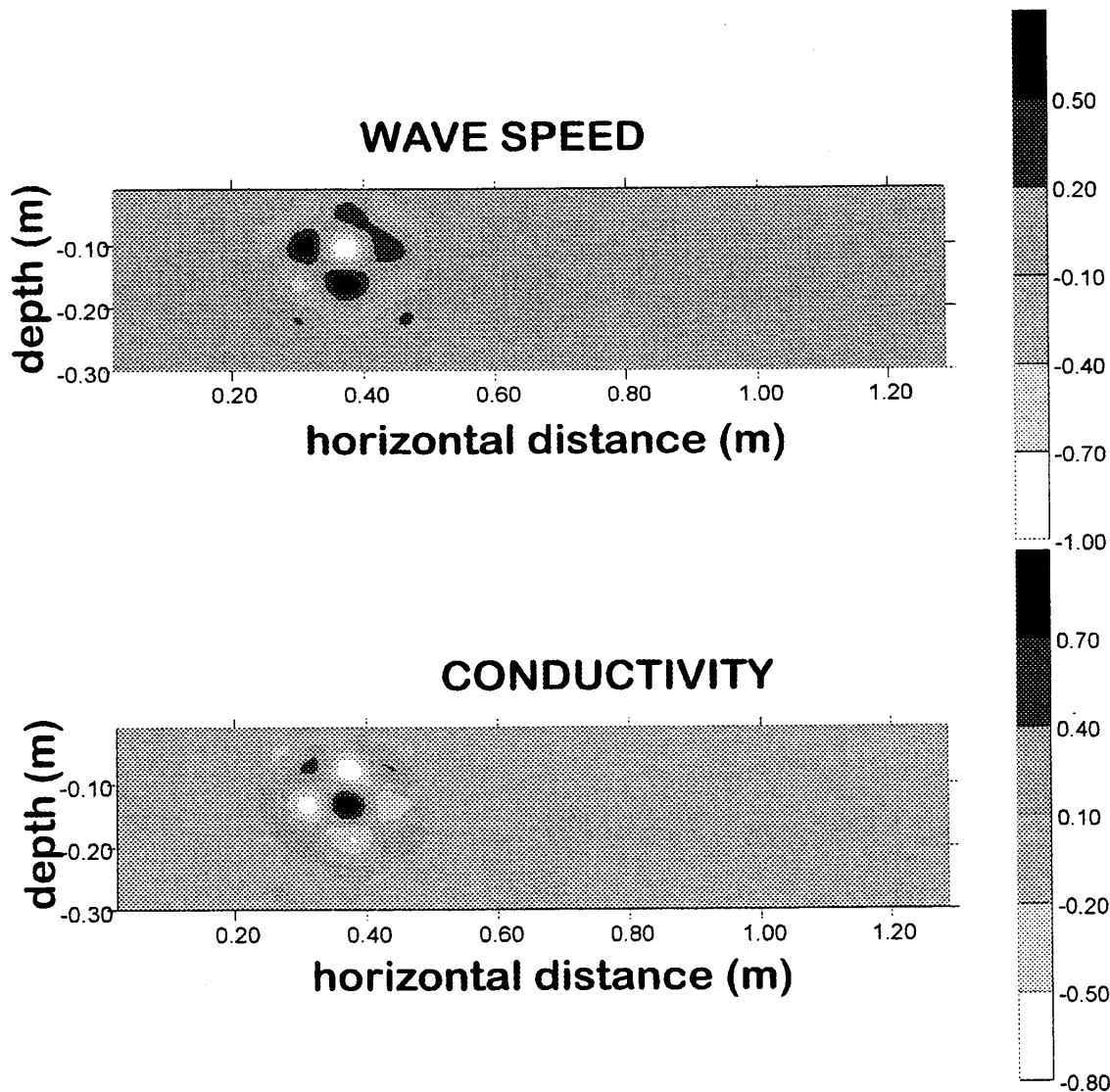


Figure 5: Gray scale image of wave speed and electrical conductivity for a buried land mine. Darker shades of gray are associated with larger values of wave speed and conductivity.

Because these images reveal that this feature has a wave speed greater than the background and an electrical conductivity less than the background, it is suspected that this is plastic mine. An interesting feature in these images is the low wave speed and high electrical conductivity in the center. This could be an image artifact associated with zero-mean object function requirement (Sect. 2) or it could be real. If the center of the image actually has a reverse contrast, then a zero-mean may approximately be realized without any image artifacts. If this central feature is real, it could be the metallic firing pin components of the mine.

5. CONCLUSIONS

A method has been developed, within the formalism of diffraction tomography and for a multi-monostatic measurement geometry, whereby spatial variations in wave speed (dielectric) and electrical conductivity can be reconstructed from broadband GPR measurements. When applied to known buried objects, the results proved consistent with the material composition of these objects.

Although the procedure described here is for multi-monostatic measurements and for a non-conducting background in two dimensions, it can readily be generalized to three dimensions and for a conducting background. Furthermore, it can also be extended to multi-bistatic transmission and reflection geometries.

6. REFERENCES

1. A.J. Devaney, "A filtered backpropagation algorithm for diffraction tomography," *Ultrason. Imaging*, **4**, pp.336-350, 1982.
2. A.J. Devaney, "Geophysical diffraction tomography," *IEEE Trans. Geosci. and Remote Sensing*, **GE-22**, pp.3-13,1984.
3. J.E. Molyneux and A.J. Witten, "Diffraction tomographic imaging in a monostatic measurement geometry," *IEEE Trans. Geosci. and Remote Sensing*, **31**, pp. 507-511, 1993.
4. A.J. Witten, J.E. Molyneux, and J.E. Nyquist, "Ground penetrating radar tomography: algorithms and case studies," *IEEE Trans. Geosci. and Remote Sensing*, **32**, pp. 461-467, 1994.
5. A.J. Devaney, "Structure determination from intensity measurements in scattering experiments," *Phys. Rev. Letters*, **62**, pp. 2385-2388, 1989.
6. A.J. Witten and J.E. Molyneux, "Geophysical diffraction tomography: validity and implementation," *Geophysical Inversion*, J.B. Bednar, L.R. Lines, R.H. Stolt, and A.B. Weglein, eds., pp. 354-369, SIAM, Philadelphia, 1992.
7. M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon, New York, 1983.
8. P.M. Morse and H. Feshbach, *Methods of Theoretical Physics, Part 1*, McGraw-Hill, New York, 1953.

Statistical modeling of rough surface scattering for ground-penetrating radar applications

George A. Tsihrintzis, Carey M. Rappaport, Scott C. Winton, and Peter M. Johansen

Center for Electromagnetics Research

Department of Electrical and Computer Engineering

Northeastern University

Boston, MA 02115, USA.

ABSTRACT

Rough surface clutter is a significant source of interference in non-specular ground penetrating radar (GPR) applications that needs to be suppressed to maintain high performance in the signal processing. Our research is in the directions of (i) development and testing of flexible parametric models for the statistical distribution of clutter that rely on the theories of alpha-stable random processes, (ii) establishment of bounds on the performance of signal processing algorithms, and (iii) design and analysis of robust, non-Gaussian signal processing algorithms based on the statistical clutter models. Synthetic data simulated with FDTD techniques are extensively used.

Keywords: Rough surface scattering, ground penetrating radar, FDTD modeling.

1. INTRODUCTION

A buried mine sensing modality with great opportunity for detailed physics-based modeling is Ground Penetrating Radar (GPR). To maximize the potential of mine detection with GPR, the radar must be wideband, to use as much of the frequency-dependent scattering information as possible. Within this area, developing accurate computational and statistical models of the mine/soil/air environment is of paramount importance. These models can be used to determine if and under what conditions a particular mine-like target can be observed and what are optimal detection procedures and help design new sensors to best collect the data needed for the detection algorithms. Particularly important are stochastic rough surface studies with FDTD simulation, since surface clutter is the major source of interference obscuring the signal from the mine (signal of interest).

The exact wave equations result in a nonlinear model for the wave-surface interaction, one that is analytically difficult to invert. Therefore, approximations are usually made to the exact nonlinear model, which are valid and

sufficient only under restricting conditions. Most common approximations arise (i) either from perturbations to corresponding solvable "smooth" surface problems [5, 3, 2] or (ii) from the Kirchhoff approximation [1]. Perturbation solutions apply in the limiting case when the deviations of the rough surface from a smooth one are small with possibly large slopes relative to the wavelength. On the other hand, the Kirchhoff approximation is applicable if the irregularities of the surface have large radii of curvature relative to the wavelength. Both methods break down when incidence and/or observation angles approach grazing and none applies when the surface is rough on length scales comparable to the wavelength. Recently, field-phase perturbation techniques were reported [6] which reduced to the usual (field) perturbation techniques or the Kirchhoff approximation in the low and high frequency limits, respectively, but were also valid in intermediate frequencies.

In this paper, we present a framework for designing algorithms for detection of radar returns from buried mines in the presence of rough surface clutter. The rough surface height function is modeled as a realization from a Gaussian process and the electric field clutter statistics are related to the surface statistics via first order perturbation analysis. It is shown that the effect of clutter is minimized if the receiver is placed in the backscatter region of the mine. More specifically, the paper is organized as follows: Section 2 reviews first-order field perturbation analysis of electromagnetic scattering from a rough surface and relates field statistics to rough surface statistics. Section 3 is concerned with mine target backscatter simulation in both the frequency and time domains and preliminary conclusions on the "optimum" receiver location. Finally, Section 4 summarizes the paper and points to related future research.

2. ELECTROMAGNETIC ROUGH SURFACE SCATTERING

A. Electric Field Equations

We are concerned with the reflection of plane electromagnetic waves from a surface $z = h(x, y)$, which is almost, but not quite flat as illustrated in Fig. 1. As in [3], we are going to assume that the surface is a realization from a periodic random process with period L , thus

$$z = h(x, y) = \sum_{m,n=-\infty}^{\infty} S_{m,n} \exp[-i \frac{2\pi}{L}(mx + ny)]. \quad (2-1)$$

The upper medium ($z > h(x, y)$) is non-dispersive with dielectric constant $\epsilon_0 = 8.854 \times 10^{-12}$ farad/meter, permeability $\mu_0 = 1.257 \times 10^{-6}$ henry/meter, and conductivity $\sigma_0 = 0$, while the lower medium ($z < h(x, y)$) is dispersive with frequency-dependent relative dielectric constant ϵ_r , permeability μ_0 , and frequency-dependent conductivity σ .

A horizontally polarized monochromatic electromagnetic plane wave with electric field

$$E_x^i = 0, \quad E_y^i(x, y, z) = \exp[-ik_0(\sin \theta x - \cos \theta z)], \quad E_z^i = 0 \quad (2-2)$$

is incident on the two media interface, where $k_0 = \omega \sqrt{\mu_0 \epsilon_0}$ is the wavenumber at the frequency ω of the wave and θ is the angle between the direction of propagation of the incident wave and the positive z axis assumed such that $\sin \theta = \frac{2\pi \nu}{k_0 L}$ for some integer ν .¹ In the absence of interface roughness, the interaction of the incident wave and the

¹It is assumed that the time-dependent wave is $E^i(x, y, z) \exp(i\omega t)$.

interface would result in a *specularly* reflected wave with electric field

$$E_x^r = 0, \quad E_y^r(x, y, z) = R \exp[-ik_0(\sin \theta x + \cos \theta z)], \quad E_z^r = 0 \quad (2-3)$$

for $z > h(x, y)$, where the specular reflection coefficient R is given by

$$R = \frac{1 - \rho}{1 + \rho}, \quad \rho = \frac{\sqrt{\epsilon_r + \sigma/(i\omega\epsilon_0) - \sin^2 \theta}}{\cos \theta} \quad (2-4)$$

Similarly, the specular wave propagating in the lower medium has electric field

$$E_x^t = 0, \quad E_y^t(x, y, z) = T \exp[-ik_0(\sin \theta x + a\sqrt{\epsilon_r + \sigma/(i\omega\epsilon_0) - \sin^2 \theta} z)], \quad E_z^t = 0, \quad (2-5)$$

for $z < h(x, y)$, where the transmission coefficient T is given by

$$T = \frac{2}{1 + \rho} \quad (2-6)$$

with ρ as in Eq.(2-4). Thus, no depolarization occurs in the absence of interface roughness.

When the two media interface is rough, however, Eqs.(2-3) through (2-6) are no longer valid and both non-specular reflection and transmission and depolarization of the incident wave occur. The resulting exact scattered wave is non-trivial to compute and, in general, one needs to resort to numerical techniques. However, under the assumption of "low" frequency ω , "small" relative dielectric constant ϵ_r and conductivity σ , and "small" maximum interface height function (maximum of $|h(x, y)|$), perturbation corrections to Eqs.(2-3) and (2-5) become valid first-order approximations. The resulting expression for the y -component of the electric field is [3]

$$E_y = \begin{cases} E_y^r + \sum_{m,n=-\infty}^{\infty} A_{m,n} \exp[-i\frac{2\pi}{L}(mx + ny) - ia(m, n)z], & \text{for } z > h(x, y) \\ E_y^t + \sum_{m,n=-\infty}^{\infty} B_{m,n} \exp[-i\frac{2\pi}{L}(mx + ny) + ib(m, n)z], & \text{for } z < h(x, y) \end{cases} \quad (2-7)$$

with similar expressions for the x - and z -components E_x and E_z of the electric field. In Eq.(2-7), we have defined

$$ia(m, n) = \sqrt{-k_0^2 + \frac{4\pi^2}{L^2}(m^2 + n^2)} \quad (2-8)$$

$$ib(m, n) = \sqrt{-k_0^2(\epsilon_r + \frac{\sigma}{i\omega\epsilon_0}) + \frac{4\pi^2}{L^2}(m^2 + n^2)}, \quad (2-9)$$

where $a(m, n)$ is either positive real or negative imaginary and the same would be true for $b(m, n)$ if the lower medium were non-conducting. To first order, the coefficients $A_{m,n}$ and $B_{m,n}$ are given by [3]

$$A_{m,n} = B_{m,n} = \frac{2iUS_{m-n,n}}{d(m, n)} \left[\frac{4\pi^2 n^2}{L^2 D_{m,n}} - 1 \right], \quad (2-10)$$

where

$$U = \frac{k_0^2 T}{2} (\epsilon_r - 1 + \frac{\sigma}{i\omega\epsilon_0}) \quad (2-11)$$

and

$$D_{m,n} = \frac{4\pi^2}{L^2}(m^2 + n^2) - b(m, n)c(m, n) \quad (2-12)$$

with

$$c(m, n) = \sqrt{k_0^2 \left(\epsilon_r + \frac{\sigma}{i\omega\epsilon_0} \right) - \frac{4\pi^2}{L^2} (m^2 + n^2)} \quad (2-13)$$

and

$$d(m, n) = b(m, n) + c(m, n). \quad (2-14)$$

We see that besides the specularly reflected and transmitted components, the electric field also contains a non-specular reflection and transmission due to the interface roughness. Additionally, depolarization occurs as the electric field vector also has x - and z -components.

B. Relation of Surface Statistics to Clutter Statistics

Consider now the surface height function as a realization from a periodic Gaussian random process of a certain covariance function. Since the Fourier series coefficients $S_{m,n}$ are given via a linear transformation of the surface height function, they constitute a Gaussian vector of a certain covariance matrix \underline{R} . Similarly, since the electric field clutter in Eq.(2-7) is a linear functional of the Fourier series coefficients $S_{m,n}$, the clutter will constitute a Gaussian process of some covariance function. Relating the covariance function of the electric field clutter to that of the Fourier series coefficients, we get:

$$\begin{aligned} \langle E_y(x, y, z) E_y^*(x', y', z') \rangle &= 4|U|^2 \sum_{m,n=-\infty}^{\infty} \frac{\sigma_{m,n}^2}{|d(m, n)|^2} \left[\frac{4\pi^2 n^2}{L^2 D_{m,n}} - 1 \right] \left[\frac{4\pi^2 n^2}{L^2 D_{m,n}^*} - 1 \right] \\ &\quad \exp \left[-i \frac{2\pi}{L} (m(x - x') + n(y - y')) - i(b(m, n)z - b^*(m, n)z') \right], \end{aligned} \quad (2-15)$$

where

$$\sigma_{m,n}^2 = \langle |S(m - n, n)|^2 \rangle. \quad (2-16)$$

3. MINE TARGET BACKSCATTER SIMULATION FROM ROUGH DISPERSIVE GROUND

A. Frequency Domain Simulation

Consider again the configuration in Fig. 1, in which a cylindrical object (mine) is buried in soil with its center located at $(0, 0, z_m)$ ($z_m < 0$) and interrogated by a monochromatic plane wave. We define the (frequency-dependent) object function

$$O(x, y, z; \omega) = \begin{cases} \Delta k^2 \equiv \omega^2 \mu \epsilon_0 (\Delta \epsilon_r - \frac{\Delta \sigma}{i\omega\epsilon_0}), & \text{if } |z - z_m| \leq \delta z \text{ and } \sqrt{x^2 + y^2} \leq \Delta \rho \\ 0, & \text{else,} \end{cases} \quad (3-1)$$

where $\Delta \epsilon_r$ and $\Delta \sigma$ are the differences in dielectric constant and conductivity, respectively, between the object and the soil.

Under the assumption that the interface between air and soil is perfectly flat, the signal due to the mine is given to first-order by

$$E_y^m(x, y, z) = \iint dK_x dK_y e^{i(K_x x + K_y y)} B(K_x, K_y, K_{0x}, K_{0y}, \omega) e^{im_1(K_x, K_y)z} I(K_x, K_y), \quad (3-2)$$

where

$$I(K_x, K_y) = \pi \Delta k^2 \Delta z \Delta \rho \operatorname{sinc} \left[\frac{\Delta z}{2} (m_2(K_x, K_y) + m_2(K_{0x}, K_{0y})) \right] \frac{J_1 \left[\frac{\Delta \rho}{2} \sqrt{(K_x - K_{0x})^2 + (K_y - K_{0y})^2} \right]}{(K_x - K_{0x})^2 + (K_y - K_{0y})^2} \cdot e^{-i[m_2(K_x, K_y) + m_2(K_{0x}, K_{0y})z_m]} \quad (3-3)$$

and

$$B(K_x, K_y, K_{0x}, K_{0y}) = -i \frac{\mu \omega [(K_y^2 + m_1 m_2)(K_{0y}^2 + m_1^0 m_2^0) + K_x i K_{0x} (K_y K_{0y} - m_1 m_1^0)]}{16 \pi^4 (m_1 + m_2)(K_x^2 + K_y^2 + m_1 m_2)(m_1^0 + m_2^0)(K_{0x}^2 + K_{0y}^2 + m_1^0 m_2^0)} \quad (3-4)$$

In Eqs.(3-2), (3-3), (3-4), we have defined

$$K_{0x} = k_0 \sin \theta \quad (3-5)$$

$$K_{0y} = 0 \quad (3-6)$$

$$m_1(K_x, K_y) = -\sqrt{k_0^2 - K_x^2 - K_y^2} \quad (3-7)$$

$$m_2(K_x, K_y) = \sqrt{k_0^2 \left(\epsilon_r - i \frac{\sigma}{\omega \epsilon_r} \right) - K_x^2 - K_y^2} \quad (3-8)$$

$$m_1^0 = m_1(K_{0x}, K_{0y}) \quad (3-9)$$

$$m_2^0 = m_2(K_{0x}, K_{0y}). \quad (3-10)$$

We simulated the field in Eq.(3-2) with frequency $\omega = 1$ GHz, angle of incidence $\theta = \frac{\pi}{3}$, relative dielectric constant of lower medium $\epsilon_r = 4$, relative dielectric constant of mine $\epsilon_{mine} = 3.8$ (or $\Delta \epsilon_r = -0.2$), conductivity of lower medium $\sigma = 0.001$, and conductivity of mine $\sigma_{mine} = 0$ (or $\Delta \sigma = -0.001$). The mine was buried 5 cm in the soil ($z_m = -0.05$) had a radius of 2 cm ($\Delta \rho = 0.02$) and a height of 2 cm ($\Delta z = 0.02$). The top plot in Fig. 2 is a contour plot of the magnitude (intensity) $|E_y|$ of the electric field in Eq.(3-2) on the plane $z = 0.2$, i.e., 20 cm above the air/soil interface. The bottom plot in Fig. 2 is a plot of the values of the electric field intensity $|E_y|$ along the x -axis at $z = 0.02$. From Fig. 2, it is clear that, for the given simulation parameters, the mine's scattered field propagates mostly in the backward direction. This is in contrast with the observation that reflections from the air/soil interface are mostly in the forward direction. Thus, the ratio of the mine signal over the clutter is maximized in the backscatter direction and that is the appropriate receiver location for maximum detection performance.

B. Time Domain Simulation

The frequency domain analysis presented above is based on first order approximations to the interaction of electromagnetic waves with the mine/air/soil environment. A more reliable analysis needs to be based on the exact interaction between the electromagnetic waves and the mine/air/soil environment, as such can be provided by numerical techniques. Additionally, the analysis needs to be done in the time domain, to reveal the time separation between signals. In our work, we have been combining dispersive soil FDTD modeling with new research in statistical rough surface modeling to give a realistic prediction of the scattering from ground with and without a mine target. Our results show clearly that even with the very low dielectric contrast of a non-metallic mine buried in sand, the backscattered signal from the mine can be readily observed and interpreted. The important aspects of the sensing is that the excitation pulse be short and inclined to the ground surface.

Our test case consists of a small circular antipersonnel mine; 4 cm in diameter, with no metal components—only TNT—embedded in either 10% moist clay loam (which is quite dispersive) or dry sand (which has a lower dielectric constant than TNT). The ground surface is chosen to be quite rough, with height statistically varying with position as much as ± 15 cm over a 1 m path. The mine target is buried close to the surface (making it notoriously difficult to sense with radar), within 5 cm. Two conditions enable a radar to detect a mine in this challenging case: 1) the radar pulse must be short enough to allow scattering from the ground surface to separate in time from the scattering from the buried mine, and 2) the incident pulse must be inclined relative to the surface normal, so that the specular reflection propagates away from the source. By positioning a receiver near the surface, slightly behind the suspected mine position, and observing the backscattered signal at the time when a signal leaving the suspected mine would arrive at the receiver, a received signal above a nominal clutter threshold would unambiguously indicate the presence of a mine. For our simulations, a 2 GHz modulated gaussian pulse, incident at 30 deg. from normal, with a single receiver 20 cm above the ground surface is sufficient to highlight the presence of an entirely plastic mine.

Observing the total field as a function of position with the mine present (Figs. 3 and 4) and the field due to just the ground without the mine (Figs. 5 and 6) shows the difference (the mine's backscattered field) (Figs. 7 and 8) quite clearly in the backscattered direction. While most of the wave reflected from the ground surface propagates mostly forward, the field scattered from the mine propagates in all directions, and this backward propagation can be distinguished even in the presence of the incident and specularly reflected fields. That is, by positioning detectors behind the target, illuminating the ground with an inclined wave, and time gating the received signal to correspond to the arrival time from a target buried at a given distance, the measured signal (exceeding a tunable threshold) will be due solely to the buried object. A receiver positioned as in Fig. 3 (or 4), would measure the signals given in Figs. 4 and 6 with and without the mine present, respectively.

Our work continues with the study of polarization effects (specifically TM polarization with Brewster angle incidence), the clutter effects of buried rocks and moisture variations, incident pulse tailoring for specific targets, and sensor placement for optimal detection.

4. SUMMARY AND FUTURE WORK

In this paper, we attempt to combine dispersive soil FDTD modeling with new research in statistical rough surface modeling to give a realistic prediction of the scattering from ground with and without a mine target and to allow the application of optimum hypothesis testing theory to solve the problem of detection of radar returns from buried mines in the presence of rough surface clutter. For the rough surface, we proposed a statistical model that relies on the theories of sub-Gaussian random processes [4] and is a generalization of the Gaussian model that allows for impulsiveness.

We reviewed first order field perturbation electromagnetic scattering from rough surfaces and related the electric field clutter statistics to the statistics of the rough surface. We showed that within first order theory, a cylindrical mine scatters a plane wave mostly in the backward direction, as opposed to ground reflections which are mainly in the

forward direction. We verified this result with numerical (FDTD) electromagnetic wave propagation and scattering analysis.

Our work continues with the study of polarization effects (specifically TM polarization with Brewster angle incidence), the clutter effects of buried rocks and moisture variations, incident pulse tailoring for specific targets, and sensor placement for optimal detection. Future work will also be in the directions of validation of the sub-Gaussian model on real data and design and performance documentation of algorithms for detection of buried mines for a large variety of realistic models for the signals returned by the mine.

ACKNOWLEDGEMENTS

This work was supported by ARO Grant DAAG55-97-1-0013.

References

- [1] P Beckmann and A Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces*. Pergamon, New York, NY, 1963.
- [2] K M Mitzner. Effect of small irregularities on electromagnetic scattering from an interface of arbitrary shape. *J. Math. Phys.*, 5:1776-1786, 1964.
- [3] S O Rice. Reflection of electromagnetic waves from slightly rough surfaces. *Comm. Pure Appl. Math.*, 4:351-378, 1951.
- [4] G A Tsihrintzis and C L Nikias. Modeling, parameter estimation, and signal detection in radar clutter with alpha-stable distributions. In *1995 IEEE Workshop on Nonlinear Signal and Image Processing*, Neos Marmaras, Halkidiki, Greece, June 1995.
- [5] P C Waterman. Scattering by periodic surfaces. *J. Acoust. Soc. America*, 57:791-802, 1975.
- [6] D Winebrenner and A Ishimaru. Investigation of a surface field perturbation technique for scattering from rough surfaces. *Radio Sci.*, 20:161-170, 1985.

Imaging and Detection of Mines from Acoustic Measurements

Alan Witten^a, C. DiMarzio^b, W. Li^b, and S.W. McKnight^b

^aSchool of Geology and Geophysics, University of Oklahoma, 810 Energy Center, Norman, OK 73019-0628

^bCenter for Electromagnetic Research, Northeastern University, Boston, MA 02115

ABSTRACT

A laboratory-scale acoustic experiment is described where a buried target, a hockey puck cut in half, is shallowly buried in a sand box. To avoid the need for source and receiver coupling to the host sand, an acoustic wave is generated in the subsurface by a pulsed laser suspended above the air-sand interface. Similarly, an airborne microphone is suspended above this interface and moved in unison with the laser. After some pre-processing of the data, reflections from the target, although weak, could clearly be identified. While the existence and location of the target can be determined by inspection of the data, its unique shape can not. Since target discrimination is important in mine detection, a three-dimensional imaging algorithm was applied to the acquired acoustic data. This algorithm yielded a reconstructed image where the shape of the target was resolved.

Keywords: laser-acoustics, diffraction tomography, maximum likelihood estimation, mine-like targets

1. INTRODUCTION

A fundamental limitation in the use of acoustic methods for problems such as mine detection is associated with acoustic coupling to the ground. The need for mechanical contact between both sources and receivers makes the deployment of sensors slow and, hence, impractical for applications such as mine detection that require rapid data acquisition. The use of laser-induced compressional waves can eliminate the need for source-soil coupling and, with the laser source, acoustic energy from shallowly buried objects can potentially be recovered with a near-surface airborne microphone. Such measurements eliminate the need for any sensor coupling to the ground surface.

In this paper, laser acoustic measurements are described. In a laboratory setting over a simple, shallowly-buried mine-like target in a multi-monostatic measurement geometry a laser source and an adjacent microphone are moved in unison over a plane above the ground surface.

Subsequently, a three-dimensional imaging algorithm is applied to these measurements, with the hope of both recovering buried targets from within signal noise and, furthermore, distinguish mine-like targets from other targets that might be encountered in the shallow subsurface.

2. THE THREE-DIMENSIONAL IMAGING ALGORITHM

The three-dimensional imaging algorithm used here is an extension of a two-dimensional multi-monostatic filtered backpropagation algorithm developed by Molyneux and Witten¹. Here it is assumed that a co-located acoustic source and receiver are moved in unison along the horizontal plane $z = 0$. The measured acoustic wave field $\bar{u}(\mathbf{r}, t)$ is the total field measured at point \mathbf{r} on the measurement plane at time t . This total field can be represented by the superposition of the incident field \bar{u}_0 , the field that would be measured in the a homogeneous medium, and a perturbed field \bar{u}' associated with any scatterers that are present.

The imaging algorithm is implemented in the frequency domain so that temporal Fourier transformed wave fields

$u(\mathbf{r}, k_0)$, $u_0(\mathbf{r}, k_0)$, and $u'(\mathbf{r}, k_0)$ are introduced and defined as

$$f(\mathbf{r}, k_0) = \int dt \bar{f}(\mathbf{r}, t) e^{i c_0 k_0 t},$$

where \bar{f} and f represent the total, incident, or perturbed wave fields and their temporal Fourier transform. The background wavenumber k_0 is defined to be ω/c_0 where ω is the frequency and c_0 is the assumed real wave speed in the propagating medium in the absence of any scatterers. For any scatterers located in the half-space $z < 0$ and invoking the Born approximation, the perturbed wave field can be represented as

$$u'(\mathbf{r}, k_0) = -i \frac{k_0^2}{4\pi} \int d\mathbf{x}' O(\mathbf{x}') u_0(\mathbf{x}') \int \frac{d\mathbb{Q}}{\sqrt{k_0^2 - \alpha^2}} e^{-i(\mathbb{Q} + \sqrt{k_0^2 - \alpha^2} \hat{\mathbf{z}}) \cdot (\mathbf{x}' - \mathbf{r})}, \quad (1)$$

where $\mathbf{x}' = (x', y', z')$, the integral over $\mathbb{Q} = (\alpha_x, \alpha_y, 0)$ is the Weyl expansion for scalar Green's function², $\hat{\mathbf{z}}$ is the unit vector in the vertical (upward) direction, and

$$O(\mathbf{x}) = 1 - \frac{k^2(\mathbf{x})}{k_0^2} \quad (2)$$

is the object function and $k(\mathbf{x})$ is the spatial variations in wavenumber that can be complex. For a point acoustic source located at \mathbf{r} , the incident field at \mathbf{x}' is

$$u_0(\mathbf{x}') = i \int \frac{d\mathbf{r}}{\sqrt{k_0^2 - \beta^2}} e^{i(-\sqrt{k_0^2 - \beta^2} \hat{\mathbf{z}}) \cdot (\mathbf{x}' - \mathbf{r})}. \quad (3)$$

Taking k_0 to be real, the real part of the object function

$$O_R(\mathbf{x}) = 1 - \frac{c_0^2}{c^2(\mathbf{x})} \quad (4)$$

where $c(\mathbf{x})$ is the spatial variations in wave speed, is one minus the square of the real refractive index. The imaginary part of the object function, O_I is the spatial variations in attenuation per unit length. Given a known background wave speed c_0 , the spatial variations in wave speed $c(\mathbf{x})$ and the attenuation per unit length that characterize scatterers can be recovered from the object function via Eqs. (2) and (4). Thus, an inversion of Eq. (1) subject to Eq. (3) whereby the object function is expressed as a function of the temporally Fourier transformed data u' can be an imaging algorithm.

A first step in the inversion of Eq. (1) is to seek a deconvolution by Fourier transforming the acquired data, u' with respect to measurement location \mathbf{r} . Defining

$$\tilde{u}'(\mathbf{l}, k_0) = \int d\mathbf{r} u'(\mathbf{r}, k_0) e^{-i\mathbf{l} \cdot \mathbf{r}},$$

substituting Eq. (3) into Eq. (1), the spatially Fourier transformed data becomes

$$\tilde{u}'(\mu, k_0) = \frac{k_0^2}{2} \int d\mathbf{x}' O(\mathbf{x}') e^{-i\mathbf{l} \cdot \mathbf{r}'} I \quad (5)$$

where

$$I = \int \frac{d\mathbb{Q}}{\sqrt{k_0^2 - \alpha^2} \sqrt{k_0^2 - (\alpha - \mu)^2}} e^{-i(\sqrt{k_0^2 - \alpha^2} + \sqrt{k_0^2 - (\alpha - \mu)^2}) z'}, \quad (6)$$

and $\mathbf{r}' = (x', y', 0)$.

The fact that the integral I remains in Eq. (5) is an indication that the spatial Fourier transform was not completely successful in the required deconvolution. The reason for this is that the measurement geometry is multi-monostatic where, by taking source and receiver to be coincident, eliminated one independent vector, either the source location

or the receiver location, that can be used for deconvolution. Had the measurement geometry been multi-bistatic where the source and receiver are positioned independently, two spatial Fourier transforms (one for the source position and one for the receiver position) could have been employed to completely deconvolve Eq. (1). For this reason, it is now necessary to deal with the integral given by Eq. (6). No exact, closed-form solution to this integral has been found; however, it has been shown¹ that for the support of O in the far-field of the measurement point, $k_0 z \gg 1$, the integral can be accurately approximated by stationary phase to give

$$\tilde{u}'(\mathbf{r}', k_0) = i \frac{k_0}{8\pi} \tilde{O}_1(\mathbf{K}), \quad (7)$$

where

$$\tilde{O}_1(\mathbf{K}) = \int d\mathbf{x}' \frac{O(\mathbf{x}')}{|z|} e^{-i\mathbf{K} \cdot \mathbf{x}'} \quad (8)$$

and

$$\mathbf{K} = (\mathbf{r}', \sqrt{4k_0^2 - \mu^2}) = (\mu_x, \mu_y, \sqrt{4k_0^2 - \mu^2}). \quad (9)$$

Equation (7) provides the needed relationship between the object function and the acquired data that can be inverted analytically. Before implementing this inversion, it is useful to recall that a far-field approximation was necessary to achieve the form given by Eq. (7). Consequently, this approximation has the potential to limit the applicability of any imaging algorithm based on Eq. (7) to scatterers that are much deeper than one wavelength. Numerical analyses performed by Molyneux and Witten¹ have established that, within the Born approximation, the invoked far-field approximation is more restrictive in the forward sense, Eq. (7), than in the inverse sense. In fact, there is no loss of image fidelity for scatterers in the near-field when compared with scatterers in the far-field.

The final form of the imaging algorithm can easily be established by expressing \tilde{O}_1 as a function of \tilde{u}' and inverting the indicated three-dimensional Fourier transform to give

$$O(\mathbf{x}) = -i \frac{4|z|}{\pi^2 c_0} \int d\omega \int d\mathbf{r}' \frac{\tilde{u}'(\mathbf{r}', k_0)}{\sqrt{4k_0^2 - |\mathbf{r}'|^2}} e^{i\mathbf{K} \cdot \mathbf{x}}. \quad (10)$$

In this equation, the integral over the vertical component of the wavenumber, K_z has been cast as an integration over the signal bandwidth. In order to obtain some expectation of the results that can be achieved from image reconstruction based on Eq. (10), it is necessary to characterize the \mathbf{K} -space coverage. It is clear from Eq. (9) that, for a particular frequency ω or its associated background wavenumber k_0 , this coverage will be an upper hemispherical surface of radius $2k_0$ centered at $K_z = k_0$. Exploiting bandwidth adds concentric hemispherical surfaces; however, there is no \mathbf{K} -space coverage at the origin. With no knowledge of \tilde{O}_1 at $\mathbf{K} = 0$, there can be no spatial DC information about the scatterers. Thus, for heterogeneities that are relatively large, images will be hollow. For smaller objects, their boundaries will blur together providing an image that is only apparently solid. This absence of spatial DC information is typical of reflection-mode signal processing subject to the Born approximation where only transitions between objects of differing properties can be recovered.

3. MAXIMUM LIKELIHOOD ESTIMATION

The use of an imaging algorithm, such as the one described above, requires some form of visual inspection of the reconstructed image to establish that any feature appearing in the image has the size and shape characteristic of a target of interest. This inspection step can be eliminated by considering a related signal processing algorithm that essentially produces an image related to likelihood of occurrence of a specified target. The approach considered here is referred to as maximum likelihood estimation³.

By defining the acquired data as $u'(\mathbf{r}, \omega)$ and a data template $U(\mathbf{r}, \omega; \mathbf{x}_0)$ to be either simulated or measured data for a target of interest located at \mathbf{x}_0 , it is clear that a good indication that this target of interest is present and located at \mathbf{x}_0 is that the error

$$E(\mathbf{x}_0) = \sum_{\omega} \int d\mathbf{r} [u'(\mathbf{r}, \omega) - U(\mathbf{r}, \omega; \mathbf{x}_0)]^2 \quad (11)$$

is below some small value characterized by the signal noise. It is recognized that Eq. (11) is simply the mean square difference, over all data, between the data and the data template. While this equation can be applied directly to identify targets of interest from measurements, its use will require either measured or simulated data templates for all targets of interest at many possible target locations.

The number of data templates can be reduced to a single data template for a given target type by first using Parseval's theorem to express the sum over all data given by Eq. (11) as the similar sum of the spatial Fourier transforms of the data and the data template

$$E(\mathbf{x}_0) = \sum_{\omega} \int d^1 \left[\tilde{u}'(\omega, k_0) - \tilde{U}(\omega, k_0; \mathbf{x}_0) \right]^2. \quad (12)$$

The relationship between the spatial Fourier transforms of the data and the object function derived for imaging can be used to analytically move the target. From Eq. (7), the data template can be expressed as

$$\tilde{U}(\omega, k_0; \mathbf{x}_0) = i \frac{k_0}{8\pi} \tilde{O}_1(\mathbf{K}; \mathbf{x}_0).$$

If it is now assumed that a data template is available for a target located at \mathbf{x}_{ref} , this data templated can be represented by

$$\tilde{U}_{ref}(\omega, k_0) = \tilde{U}(\omega, k_0; \mathbf{x}_{ref}) = i \frac{k_0}{8\pi} \tilde{O}_1(\mathbf{K}; \mathbf{x}_{ref}).$$

From the shift rule for integral transforms,

$$\tilde{O}_1(\mathbf{K}; \mathbf{x}_0) = \tilde{O}_1(\mathbf{K}; \mathbf{x}_{ref}) e^{-i\mathbf{K} \cdot (\mathbf{x}_0 - \mathbf{x}_{ref})},$$

or

$$\tilde{U}(\omega, k_0; \mathbf{x}_0) = \tilde{U}_{ref}(\omega, k_0) e^{-i\mathbf{K} \cdot (\mathbf{x}_0 - \mathbf{x}_{ref})}. \quad (13)$$

Equation (13) is the desired result in that it allows the synthesis of a data template at an arbitrary location \mathbf{x}_0 from a measured or simulated data template at any known location \mathbf{x}_{ref} . When Eq. (13) is used in Eq. (12), the error E can be computed by a combination of numerical integrations that are independent of target location and a spatial Fourier transform with respect to potential target locations.

4. THE LASER-ACOUSTIC EXPERIMENT

An experiment was conducted to acquire acoustic wave reflection data from a shallowly buried object in a sand box. The target was a 7.5 cm diameter hockey puck that was sliced in half radially and buried 1 cm deep in the sand box. To create acoustic waves in the sand, a pulsed laser was suspended above the surface of the sand. An impulsive acoustic wave was created by the interaction of the laser pulse with the air-sand interface⁴. The acoustic receiver was a transducer suspended 5 cm above the laser spot. The laser-transducer pair were moved in unison above the surface of the sand. Figure 1 illustrates this measurement geometry.

Data were acquired by moving the source-receiver pair along a series of parallel "scan" lines. Each scan line was 7 cm long and the transducer output voltage as function of time was measured at 15 points along each scan line. Twenty five scan lines were acquired over a length of 12 cm. (Fig. 2). This provided a uniform, horizontal (x, y) coverage at 0.5 cm intervals. At each measurement location, 1000 voltage measurements were acquired at 0.5 microsecond intervals. Data acquisition began (time zero) at a constant time after the time of the laser pulse.

Figure 3 is a plot of the output voltage versus time for one of the spatial measurement locations. The acoustic data are characterized by strong "ringing" presumably associated with a resonant mode of the transducer. This ringing made the identification of features in the time series associated with reflections quite difficult. For this reason, each time series was pre-processed with a filter to suppress the artifacts of resonance. This filter has the form

$$\tilde{u}'(t) = \int d\omega h(\omega) s(\omega) e^{-i\omega t}, \quad (14)$$

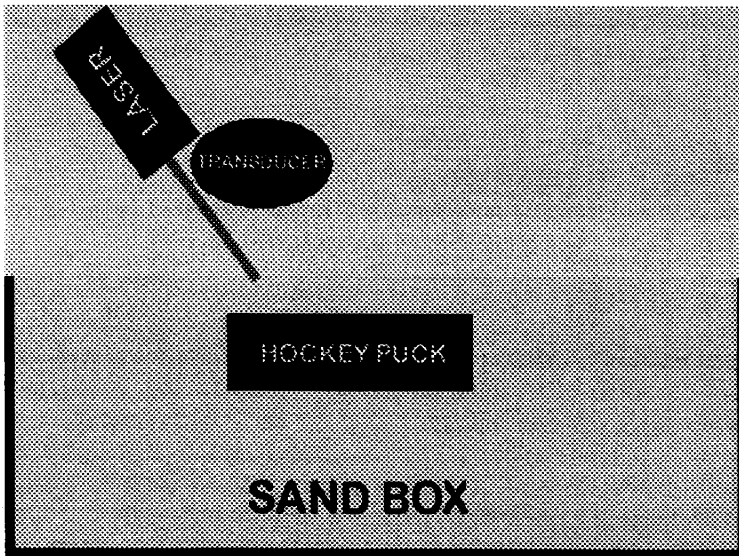


Figure 1: Illustration of the measurement geometry in a vertical cross-section.

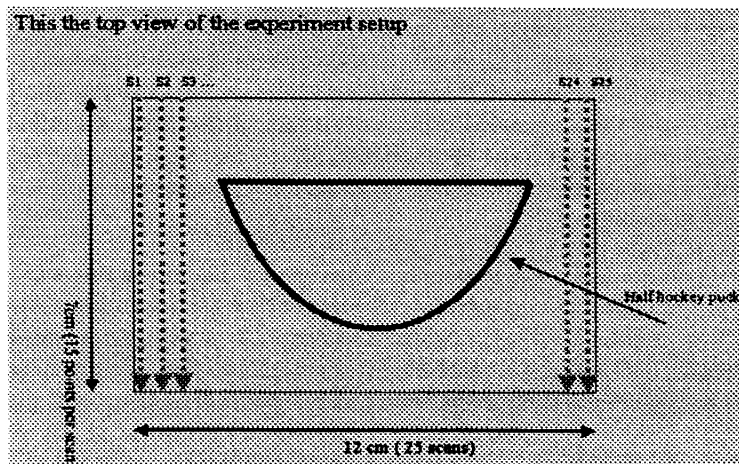


Figure 2: Illustration of the spatial measurement configuration.

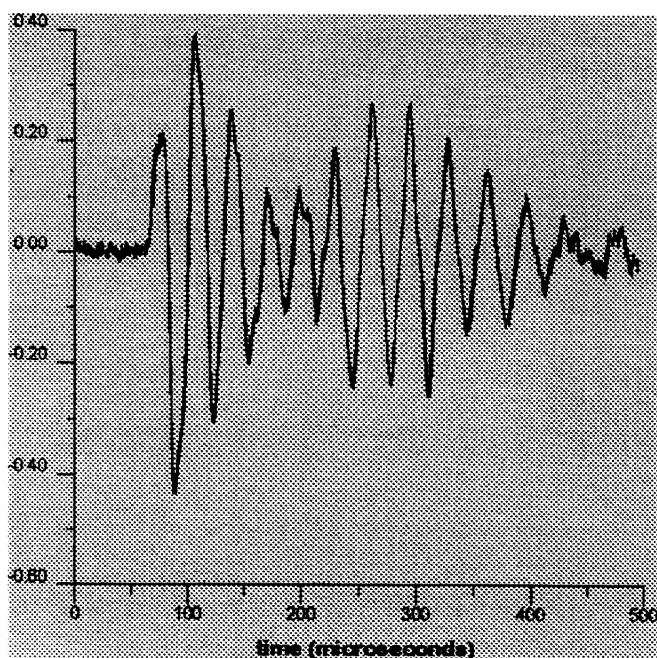


Figure 3: An example of the acoustic transducer output voltage as a function of time for one measurement location.

where

$$h(\omega) = \frac{s_{ref}^*(\omega)}{s_{ref}s_{ref}^* + \epsilon}, \quad (15)$$

s is the temporally Fourier transformed time series for any of the spatial measurement points in a particular scan line, s_{ref} is the temporally Fourier transformed time series for the first spatial measurement point in that scan line, the $*$ denotes complex conjugation, and ϵ is a constant included to eliminate any singularities in the filter h .

The filter described by Eqs. (14) and (15) has the added benefit of effectively moving the receiver location to surface of the sand. Figures 4(a) and (b) display the time series for every spatial measurement location along scan lines 1 and 12, respectively. In these figures, output voltage is displayed as gray levels with darker shades of gray associated with higher output voltages. Referring to Fig. 2, scan line 1 does not pass over the puck and, as such, there should be no evidence of a reflection from this feature. However, scan line 12 passes over the hockey puck at approximately its widest part so that a reflection from the sand-hockey puck interface should appear in the data. These data, as presented in Fig. 4, are consistent with these expectations. The pre-processed time series can also be presented as a time slice where output voltage at a fixed time is presented as a function of measurement position (x) along each scan line for all scan lines (y). Figure 5 is a gray scale presentation of the temporal data time-sliced at 53.5 microseconds. The horizontal surface of the hockey puck is clearly identified in this figure and its shape is reasonably well resolved.

5. IMAGING RESULTS

The three-dimensional imaging algorithm, Eq. (10), was applied to the pre-processed laser-acoustic data. Apart from the specification of the measurement geometry, the only user inputs to this algorithm are associated with the specification of the background wavenumber k_0 so that a background acoustic wave speed c_0 and a bandwidth must be specified. Most of the power in the acoustic signal was confined to a frequency band between about 10 and 60 kHz; however, there was limited spectral power out to about 250 kHz. Since a broader bandwidth produces better images, a bandwidth between 10 and 250 kHz was used for the reconstructions presented here. The background

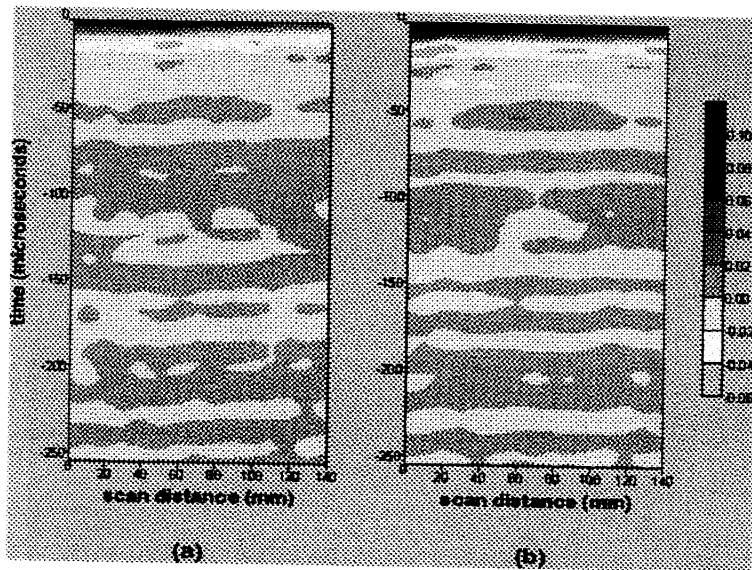


Figure 4: Transducer output voltage as function of time for (a) scan line 1 and (b) scan line 12. The horizontal axis is measurement position (x) along the scan line and the vertical axis is time in microseconds. The output voltage is displayed as gray scales having values defined by the palette on the right.

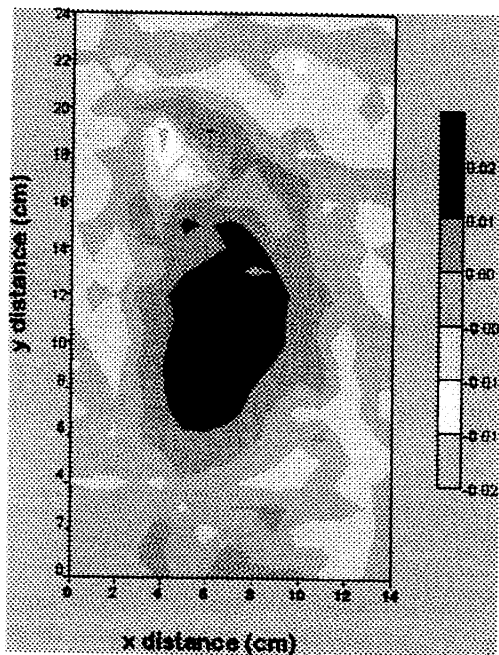


Figure 5: Gray scale presentation of output voltage time-sliced at 53.5 microseconds.

wavespeed used was 700 m/s, a value typical of poorly cemented sandstone.

Figures 6, 7, and 8 show vertical cross-sectional slices through the three-dimensional image below scan lines 1, 8, and 12, respectively. Each figure shows both real and imaginary parts of the reconstruction of the object function $O = O_R + iO_I$ where, as noted earlier, O_R can be considered to be spatial variations in relative reflectivity and O_I can be taken to be spatial variations in attenuation per unit length. These images reproduce some expected features. First, there is no evidence of the hockey puck in the vertical cross-section below scan line 1 (Fig. 6). The hockey puck does appear in the images of both O_R and O_I below scan lines 8 (Fig. 7) and 12 (Fig. 8) with the puck appearing wider below scan line 12 than scan line 8. Additionally, the puck appears with a positive contrast in O_R indicating that the wave speed of the puck is greater than the host sand. The O_I contrast of the puck is negative suggesting that the puck is less attenuating than the sand.

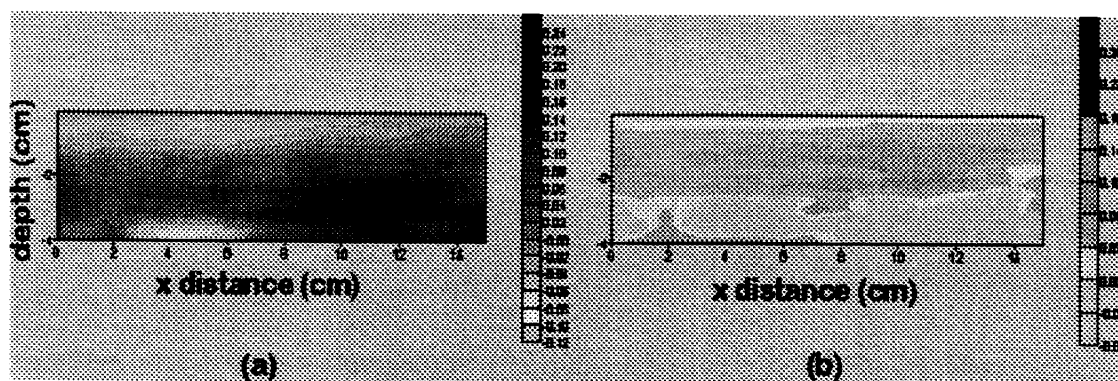


Figure 6: Reconstructed image of (a) O_R and (b) O_I through a vertical cross-section below scan line 1 displayed as gray scales.

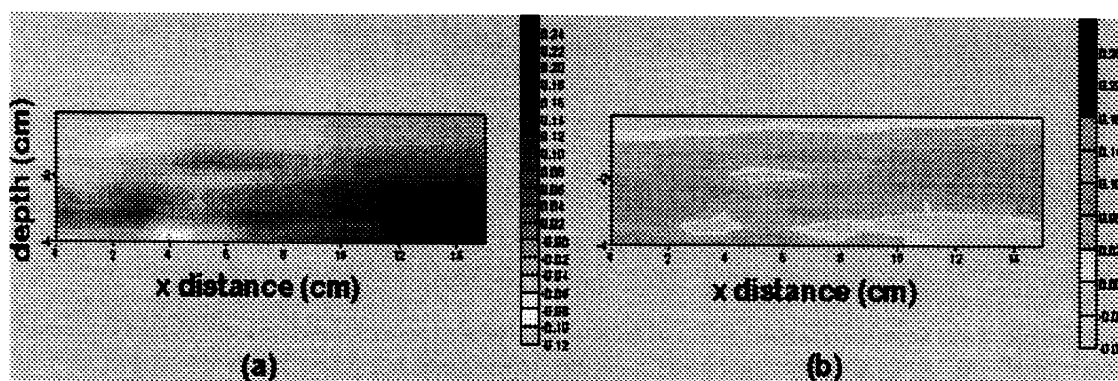


Figure 7: Reconstructed image of (a) O_R and (b) O_I through a vertical cross-section below scan line 8 displayed as gray scales.

One surprising result is that the O_I reconstruction is more robust than the O_R reconstruction. This fact can be seen by comparison of Figs. 7 and 8. The same plotting contrast was used for O_I ; however, different plotting contrasts were required to make the puck apparent in the O_R images in both vertical cross-sections. This fact becomes more evident when viewing a horizontal slice through the three-dimensional image at a depth of 1.8 cm (Fig. 9). Here the hockey puck can easily be distinguished in the O_I reconstruction but not in the O_R reconstruction.

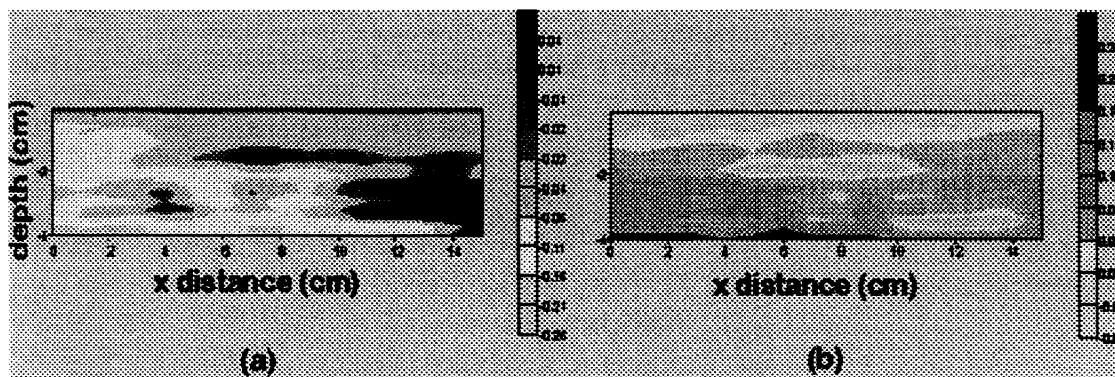


Figure 8: Reconstructed image of (a) O_R and (b) O_I through a vertical cross-section below scan line 12 displayed as gray scales.

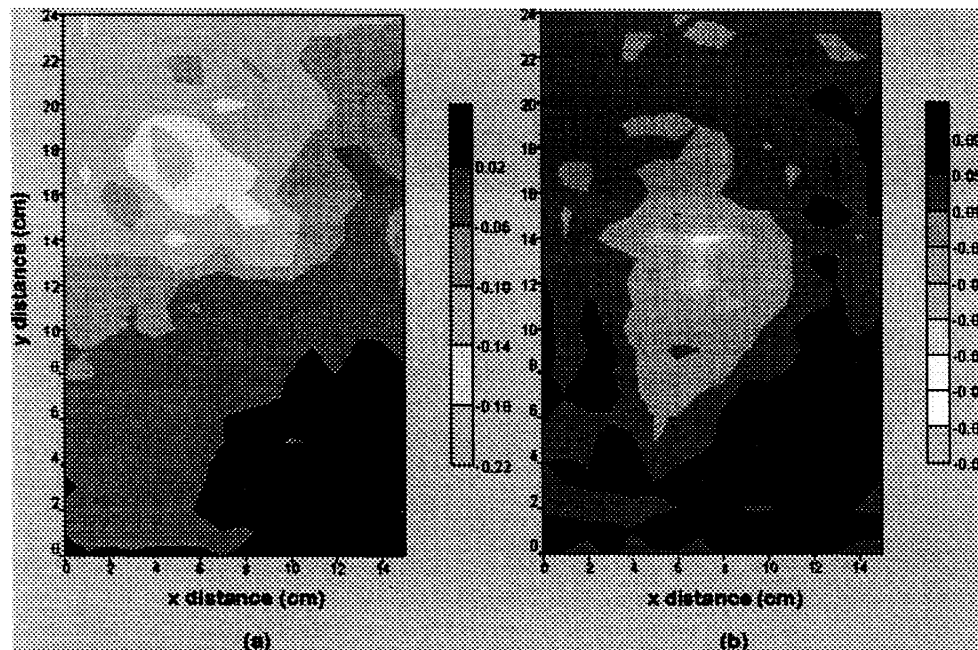


Figure 9: Gray scale image of the three-dimensional image of (a) O_R and (b) O_I horizontally sliced at a depth of 1.8 cm.

6. CONCLUSIONS

One of the objectives of this study was the evaluation of the efficacy of acoustic measurements of scattering from subsurface objects when the sensors are not in intimate contact with ground surface. In geophysical acoustic measurements, the need for mechanical coupling between both source and receiver with the ground surface is the rate limiting factor in data acquisition. For the demining problem, not only does this coupling issue slow the speed of data acquisition to point where it becomes impractical, it also presents a risk of inadvertent mine detonation.

It has been established here, at least for the ideal conditions of a shallow object in near-homogeneous background and

a planar interface, that reflections from this target can clearly be identified. As evident in Figs. 4 and 5, after some slight pre-processing the presence and location of the target can clearly be identified. Even under ideal conditions; however, the ground is quite heterogeneous where numerous man-made and naturally occurring objects may be found. For this reason, there is an ancillary discrimination problem where a detected buried object must be identified as being mine-like with some reasonable level of confidence. There are a number of attributes that can be considered for target discrimination. Here, the rather unique shape of the half hockey puck target was used and the objective of signal processing was to reconstruct this shape using a three-dimensional multi-monostatic reflection-mode imaging algorithm.

The quantity imaged is the object function which is complex having a real part that is associated with relative spatial variations in wave speed and an imaginary part that is associated with relative spatial variations in attenuation per unit length. The reconstructed three-dimensional image was displayed as two-dimensional cross-sections representing slices of this three-dimensional image that are horizontal and vertical. While both the real and imaginary parts of the object function produced images of vertical cross-sections where the target was clearly identifiable (Figs. 6a, 7a, and 8a), the wave speed contrast (the real part of the object function) varied with horizontal y -direction (Fig. 2). Consequently, the target could not be distinguished in the horizontally-sliced image based on the wave speed reconstruction (Fig. 9a). This was not the case for the reconstruction of attenuation per unit unit length. For this variable, the contrast was consistent from vertical cross-section to vertical cross-section so that the target location and shape could be clearly identified in the horizontally-sliced image (Fig. 9b).

While the imaging results obtained here suggest that attenuation per unit length is the more robust image, it is important to recall that the target appears as a relative low in this variable. This fact, in turn, suggests that there is some background attenuation. If this is indeed the case, the background wavenumber k_0 taken to be real in the imaging algorithm must, more properly, be taken to be complex. For a complex background wavenumber, the integral transform defined by Eq. (8) is no longer a spatial Fourier transform so that a different inversion procedure must be implemented such as the algebraic one given by Witten et. al.⁵. In principle, it is possible to apply such an algorithm here; however, in practice it is necessary to have some estimate of the imaginary part of k_0 . In light of this information, it useful to reconsider what was actually imaged. There are two important implications associated with a complex k_0 . First, it must be recognized that there is a depth-dependent attenuation and, as a result, a depth-dependent gain must be introduced into the imaging algorithm. This is not a significant factor here since the target was shallow and the absence of this gain factor only serves to slightly reduce the contrast of the reconstructed image. More important is the fact that a complex background wavenumber produced a blending of the real and imaginary parts of the object function that is not accounted for in the imaging algorithm. Thus, the quantities reconstructed are truly not the real and imaginary of the object function but some blending of the two. The interpretation of the real part and imaginary parts of the object function as being an indication of relative wave speed and relative attenuation, respectively, is now suspect; however, together these two imaged quantities reproduced the shape of the target providing the desired discrimination.

The computational problems associated with imaging when k_0 is complex can avoided when using maximum likelihood estimation. In such cases, the imaging algorithm, Eqs. (7) thru (10), requires an inversion of an integral transform with a complex wavenumber; however, the maximum likelihood estimator, Eqs. (12) and (13), is a forward transform and can be implemented with simple and computationally-efficient numerical procedures. A further benefit of maximum likelihood estimation not previously discussed is that the vertical component of the wave vector \mathbf{K} , $K_z = \sqrt{4k_0^2 - \mu^2}$ is complex, not only when k_0 is complex, but also when $|\mu| > 2k_0$. This fact implies that there is little benefit to imaging using a fine spatial sampling since this sampling maps into large values of μ that must be discarded. Since the maximum likelihood estimator does not require an inverse integral transform, both complex k_0 and $|\mu| > 2k_0$ can be incorporated without any computational complications. When $|\mu| > 2k_0$, evanescent waves exist. These waves propagate horizontally and decay exponentially in the vertical direction. When targets are in the near-field, there can be significant energy in evanescent modes and, by exploiting this information, target discrimination can be enhanced⁶.

REFERENCES

1. Molyneux, J.E. and Witten, A.J., "Diffraction tomographic imaging in a monostatic measurement geometry," *IEEE Transaction on Geoscience and Remote Sensing*, **31**, pp. 507-511, 1993.
2. Morse, P.M. and Feshbach, H., *Methods of Theoretical Physics, Part 1*, McGraw-Hill, New York, 1953.
3. Devaney, A.J. and Tsihrintzis, G.A., "Maximum likelihood estimation of object location in diffraction tomography," *IEEE Transactions on Signal Processing*, **39**, pp. 672-685, 1991.
4. McKnight, S.W., DeMarzio, C., Li, W., Hogenboom, D.O., and Sauermann, G., "Laser-induced acoustic detection of buried objects," *SPIE Transactions on Detection of Mines and Mine-Like Objects*, **3392**, pp. 841-847, 1998.
5. Witten, A.J., Won, I.J., and Norton, S.J., "Subsurface imaging with broadband electromagnetic induction," *Inverse Problems*, **13**, pp. 1621-1639, 1997.
6. Witten, A.J., Schatzberg, A., and Devaney, A.J., "Vector radar wave diffraction tomography maximum likelihood estimation," *Journal of Environmental and Engineering Geophysics*, **0**, pp. 91-104, 1996.

Subsurface imaging with broadband electromagnetic induction

Alan Witten[†], I J Won[‡] and Stephen J Norton[§]

[†] School of Geology and Geophysics, The University of Oklahoma, 810 Energy Center, Norman, OK 73019-0628, USA

[‡] Geophex Ltd, 605 Mercury St, Raleigh, NC 27603-2348, USA

[§] Energy Division, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, TN 37831-6200, USA

Received 7 February 1997, in final form 11 August 1997

Abstract. Diffraction tomography (DT) is a quantitative technique for high-resolution subsurface imaging. In general, DT algorithms are used for subsurface imaging with propagating waves. In this study an imaging algorithm is developed and tested for use with broadband electromagnetic induction for a so-called 'multimonostatic' measurement geometry; a primary and secondary coil are coincident and move in unison over a uniformly spaced grid, on or above the ground surface. The algorithm is formulated in three dimensions and tested on simulated data for inhomogeneities that are both two and three dimensional. The algorithm is also applied, in two dimensions, to data acquired over a pair of parallel tunnels. One important finding is that good images can be reconstructed when the frequency band is limited to the case where all skin depths are greater than the depth of inhomogeneities.

1. Introduction

Since the early 1980s (Devaney 1982), work has been proceeding on the development of a class of high-resolution imaging algorithms that has become known as diffraction tomography (DT). The earliest concepts for geophysical diffraction tomography (GDT) (Devaney 1984) were for transmission-mode measurements—either offset vertical seismic profiling or cross-borehole—where tomographic images are reconstructed for a fixed frequency of illuminating wave.

In a parallel development, Won (1980) showed that broadband electromagnetic induction (EMI) could be used to perform vertical soundings of the Earth. By moving a coil pair along or above the ground surface and acquiring data over a broad range of frequencies at each coil location, a defocused image will result from displaying the secondary coil output as a function of skin depth and horizontal coil position. To illustrate this, consider an isolated conductive anomaly at some depth d in a homogeneous background and a coil pair fixed on the ground surface directly above the inhomogeneity. Increasing the skin depth δ (decreasing the driving frequency) from some small value to successively larger values, such that δ remains less than d , will not produce any significant change in the measured secondary field. This is because: (1) the background conductivity is constant; and (2) both the primary and secondary field decay exponentially with the skin depth so that the conducting anomaly is exposed to a very weak primary field that, in turn, produces a weak secondary field that undergoes substantial decay between the anomaly and the ground

surface. The secondary field strength is some weighted average of the vertical variations in conductivity over the skin depth. Consequently, as δ is increased to the point where it is comparable to d , the inhomogeneity produces a perturbation in the secondary field. As δ increases beyond d , the vertical extent of the anomaly occupies a decreasing fraction of the skin depth and the secondary field becomes progressively weaker. Since the field from a dipole coil diminishes by geometric spreading at a rate inversely proportional to the distance cubed, moving the coil pair laterally away from the inhomogeneity and repeating the sounding yields an attenuated and, since the field propagates as well as decays with distance, perhaps, phase-shifted replicate of the sounding taken directly over the inhomogeneity. Thus, the display of broadband EMI data as a function of measurement position and skin depth essentially produces a blurred image of the anomaly that peaks at the lateral location of the anomaly and at a skin depth comparable to the target depth.

More recently, DT algorithms have been developed for ground penetrating radar (Molyneux and Witten 1993, Witten *et al* 1994). This work considers broadband electromagnetic wave illumination in a multimonostatic measurement geometry. Since the multimonostatic geometry assumed in the above cited radar-based imaging algorithms is the same as that typically employed in EMI data acquisition, it would seem that these radar algorithms could be extended to the case of a complex background wavenumber and thereby allow imaging with broadband EMI. However, there exists decades of experience in processing wave-based data; and consequently, optimal spatial and temporal sampling is well known. Specifically, in wave-based methods, it is necessary to sample spatially at intervals no greater than one-half of a wavelength. In geophysical EMI measurements, the wavelength is quite large so that the wave characteristics can be resolved with a coarse spatial sampling. It is expected that sampling based on the Nyquist criterion will not yield well resolved images and it remains to establish rules for optimal spatial sampling for EMI imaging. Because most EMI tools operate at relatively low frequencies, up to tens of kHz, an issue of particular importance is whether inhomogeneities, having characteristic depths less than the skin depth associated with the highest frequency, can be imaged.

In this study, a DT-based imaging algorithm is developed for broadband, multimonostatic imaging with EMI. A forward model (Witten *et al* 1994) is first presented that serves as the basis for the subsequent imaging algorithm. This algorithm is then applied to both synthetic and actual field data.

2. The forward model

The background geometry considered (figure 1) is a uniform isotropic space characterized by constant dielectric ϵ_0 , susceptibility μ_0 , and conductivity σ_0 , in which is embedded some localized inhomogeneity characterized by a spatially-variable conductivity, $\sigma(\mathbf{r})$ (and, perhaps $\epsilon(\mathbf{r})$) having support in the half-space $z < 0$. The governing equation for an electric field \mathbf{E} driven at a frequency ω is derived from Maxwell's equations

$$(\nabla \times \nabla \times - k_0^2) \mathbf{E} = f(\mathbf{r}) \mathbf{E} \quad (1)$$

where $k_0^2 = i\omega\mu_0(-i\omega\epsilon_0 + \sigma_0)$ is the background square wavenumber and inhomogeneities are characterized by the object function $f(\mathbf{r}) = i\omega\mu_0(-i\omega[\epsilon(\mathbf{r}) - \epsilon_0] + [\sigma(\mathbf{r}) - \sigma_0])$. For simplicity, a low-frequency approximation is assumed where $\sigma(\mathbf{r}), \sigma_0 \gg \omega\epsilon(\mathbf{r}), \omega\epsilon_0$, so that

$$k_0 \approx \sqrt{i\omega\mu_0\sigma_0} \quad (2)$$

and

$$f(\mathbf{r}) \approx i\omega\mu_0[\sigma(\mathbf{r}) - \sigma_0] = k_0^2 \left[\frac{\sigma(\mathbf{r})}{\sigma_0} - 1 \right] = k_0^2 O(\mathbf{r}) \quad (3)$$

where $O(\mathbf{r})$ is referred to as the object function.

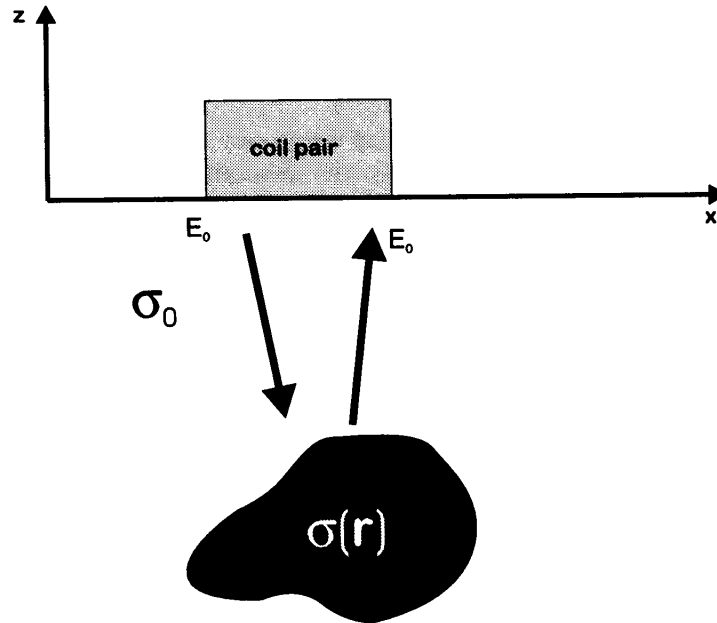


Figure 1. The host geometry under consideration. The y axis (not shown) is into the plane of the figure. A co-located transmitting and receiving coil is located in the box labelled coil pair. Each coil is assumed to have an arbitrary orientation.

Equation (1) can be expressed in the form of an integral equation for the secondary field E_s as

$$E_s(\mathbf{r}) = \int d\mathbf{r}' E(\mathbf{r}') \cdot \mathcal{G}(\mathbf{r} - \mathbf{r}') f(\mathbf{r}') \quad (4)$$

where \mathcal{G} is the dyadic Green function. Assuming that $z' < z$, which will be the case when measurements are made on or above the ground surface and the support of f is entirely in the lower half-space (figure 1), \mathcal{G} can be expressed as the Weyl-type expansion (Morse and Feshbach 1953)

$$\mathcal{G}(\mathbf{r} - \mathbf{r}') = \frac{i}{8\pi^2} \int d\mathbf{K} \frac{1}{m} (\mathcal{I} - \hat{\mathbf{k}}_+ \hat{\mathbf{k}}_+) e^{i\mathbf{k}_+ \cdot (\mathbf{r} - \mathbf{r}')} \quad (5)$$

where \mathcal{I} is the identity dyad, $\mathbf{k}_+ = \mathbf{K} + m(\mathbf{K})\hat{\mathbf{z}}$, $m(\mathbf{K}) = \sqrt{k_0^2 - |\mathbf{K}|^2}$, the imaginary part of $m \geq 0$, the hatted quantities are unit vectors, and \mathbf{K} is the two-dimensional horizontal wavevector.

By assuming that the electric field E changes slowly over the support volume of $f(r)$, the volume over which $f(r) \neq 0$, the field appearing under the integral in (4) can be replaced by the primary field E_0 . This is the Born approximation. The primary field at some r' with $z' < 0$ can be represented in terms of the primary field at $z = 0$ by

$$E_0(r') = \frac{1}{(2\pi)^2} \int dK' e^{ik'_- \cdot r'} \tilde{E}_0(K', z = 0) \quad (6)$$

where \tilde{E}_0 is the two-dimensional, horizontal Fourier transform of E_0

$$\tilde{E}_0(K', z) = \int dx' e^{-iK' \cdot x'} E_0(x', z) \quad (7)$$

$r' = (x', z')$ and $k'_- = K' - m(K')\hat{z}$.

Invoking the Born approximation, taking the primary coil to be located at point r_0 and using (5) and (6) in (4); the horizontally Fourier-transformed secondary field becomes

$$\begin{aligned} \tilde{E}_s(K, z; r_0) &= \int dx E_s(x, z) e^{-iK \cdot x} \\ &= e^{imz} (\mathcal{I} - \hat{k}_+ \hat{k}_+) \cdot \frac{i}{2m} \int dK' \tilde{E}_0(K', 0) e^{-ik'_- \cdot r_0} \int dr' f(r') e^{-i(k_+ - k_-) \cdot r'} \end{aligned} \quad (8)$$

and $\tilde{E}_0(K', 0)$ characterizes the coil used to generate the primary field.

The next steps in the development of the forward model formulation involve the representation of the transmitter and receiver coils in terms of their orientations and their input and output characteristics. The open terminal output voltage from the receiver coil resulting from its response to a secondary field E_s is given by

$$V_{oc}(r_0) = -\frac{1}{I_0} \int dr' J(r') \cdot E_s(r' + r_0) \quad (9)$$

where I_0 is the current across the coil terminals and $J(r')$ is the current distribution within the coil. It has been assumed that the receiver coil is coincident with the transmitter coil at r_0 .

The vector magnetic potential for a current distribution J_0 in the transmitter coil is given by

$$A(r) = \mu_0 \int dr' J_0(r') G(r - r') \quad (10)$$

where

$$G(r - r') = \frac{i}{8\pi^2} \int dK' \frac{1}{m(K')} e^{ik'_- \cdot (r - r')}. \quad (11)$$

Using the relationship between the magnetic potential and electric field

$$E = \frac{i}{\omega\mu_0\epsilon_0} \nabla \times \nabla \times A \quad (12)$$

the two-dimensional, horizontally Fourier-transformed primary field can be written as

$$\tilde{E}_0(K, z) = -\frac{\omega\mu_0}{2m} (\mathcal{I} - \hat{k}_- \hat{k}_-) \cdot \int dr' J_0(r') e^{-ik_- \cdot r'}. \quad (13)$$

The final introduction of coil-specific parameters is made by assuming that a coil (either the transmitter or receiver) is of small radius a , consisting of N turns with its axis pointing in the direction \hat{n} , and is located at some elevation above the ground surface $h \geq 0$. By taking the current to be constant within the coil,

$$(\mathcal{I} - \hat{k}_\pm \hat{k}_\pm) \cdot \int dr' J(r') e^{\pm ik_\pm \cdot r'} = \pm i\pi N a^2 e^{imh} \hat{n} \times \hat{k}_\pm. \quad (14)$$

Substituting (

$$\tilde{V}_{oc}(K) = -ia$$

where $\tilde{V}_{oc}(K)$

N_s is the number of turns of the coil, a_s is the radius of the coil,

$$H(K, K'; k_+)$$

and \hat{n}_s and \hat{n}_0 are the unit vectors in the direction of the transmitter coil and the receiver coil, respectively. The material having a two-layer system

3. Inversion and

The inner integral in the forward model is the form of the forward model of diffraction tomography. The transform of the forward model is a GPST, imaging the forward model. The forward model is a (K') integration. In order to derive

The procedure whereby the outer integral is approximated. This approximation varies slowly compared to the inner integral. This, in turn, requires that the inner integral be approximated by δ where $\delta = 1/\sqrt{\epsilon_0}$. This approximation was used by Witten (1993). It is hoped that this final step in the forward model. Performing the K''

$$\tilde{V}_{oc}(K)$$

where

$$\gamma(K) = \sqrt{4k_0^2 - |K|^2}$$

$$k = K + \gamma(K)\hat{z}$$

$$\tilde{O}_1(k) = \int dr \frac{O_1(r)}{(h - z)}$$

$$\tilde{O}(K, z) = \int dx C$$

Substituting (9), (13) and (14) into (8) yields the final form of the forward model

$$\tilde{V}_{oc}(K) = -i\omega\mu_0(\pi k_0)^4 \int dK' (e^{iQh}/Q) H(K, K'; k_+) \int dr' O(r') e^{-i(K \cdot x + Qz)} \quad (15)$$

where $\tilde{V}_{oc}(K)$ is the horizontally Fourier-transformed output voltage,

$$Q = m(K/2 + K') + m(K/2 - K')$$

N_s is the number of turns in the receiver coil, N_0 is the number of turns in the transmitter coil, a_s is the radius of the receiver coil, a_0 is the radius of the transmitter coil,

$$H(K, K'; k_+) = I_0 N_s N_0 a_s^2 a_0^2 [\hat{n}_s \times \hat{k}_+(K' + K/2)] \cdot [\hat{n}_0 \times \hat{k}_+(K' - K/2)] \quad (16)$$

and \hat{n}_s and \hat{n}_0 are the unit vectors defining the direction of the axis of the receiver and transmitter coils, respectively. Equation (15) assumes that the coils are imbedded in a material having the same conductivity as the material that hosts the support of $O(r)$. A two-layer system, such as air over soil or rock, is considered in appendix A.

3. Inversion and imaging

The inner integral in (15) is a spatial Fourier transform of the object function O . This form of the forward model is similar to a 'generalized projection slice theorem' (GPST) of diffraction tomography. GPSTs are analytic relationships between the spatial Fourier transform of the acquired data and the spatial Fourier transform of the object function. With a GPST, imaging can be accomplished by inversion using standard numerical techniques. The forward model defined by (15) is not suitable for such an inversion due to the outer (K') integration. Thus, it is necessary to seek some closed-form evaluation of this integral in order to derive a GPST for broadband EMI imaging.

The procedure employed here parallels that established by Molyneux and Witten (1993) whereby the outer integration in (15) is performed approximately by the stationary phase. This approximate representation is valid provided that $H/m(K/2 + K')m(K/2 - K')$ varies slowly compared to $[m(K/2 + K') + m(K/2 - K')](h - z)$ over some range of K . This, in turn, requires that there be a large coefficient in the exponential of the integrand and this will be the case when $(h - z)$ is much greater than one-half of the skin depth δ where $\delta = 1/\sqrt{\sigma_0\mu_0\omega}$. Rigorously, at least in the forward sense, this weak far-field approximation may limit the applicability of this imaging algorithm; however, Molyneux and Witten (1993) demonstrated that no such limitation exists in the inverse sense. It is hoped that this finding can be generalized to the case of EMI inversion considered here. Performing the K' integration in (15) by stationary phase yields the approximate relationship

$$\tilde{V}_{oc}(K) \approx \frac{1}{2}\omega\mu_0 k_0^2 \pi^5 \bar{H}(k) \gamma(K) e^{i\gamma h} \bar{O}_1(k) \quad (17)$$

where

$$\gamma(K) = \sqrt{4k_0^2 - |K|^2}$$

$$k = K + \gamma(K)\hat{z}$$

$$\bar{O}_1(k) = \int dr \frac{O(r)}{(h - z)} e^{-ik \cdot r} = \int dz \frac{e^{-i\gamma z}}{h - z} \int dx O(x, z) e^{-iK \cdot x} = \int dz \frac{e^{-i\gamma z}}{h - z} \bar{O}(K, z)$$

$$\bar{O}(K, z) = \int dx O(x, z) e^{-iK \cdot x} \quad (18)$$

and

$$\bar{H}(\mathbf{k}) = I_0 N_0 N_s a_0^2 a_s^2 (\hat{\mathbf{n}}_0 \times \mathbf{k}) \cdot (\hat{\mathbf{n}}_s \times \mathbf{k}) \quad (19)$$

is a coil-specific filter function.

Equation (17) is the final form of the GPST for broadband EMI imaging. This equation can be rewritten as

$$\hat{O}_1(\mathbf{k}) = \frac{2}{\omega \mu_0 \pi^2 k_0^2 \gamma(K)} e^{-i\gamma h} \bar{V}_{oc}(K) \quad (20)$$

to explicitly represent O_1 as a function of V_{oc} . An image of the spatial variations in O_1 and, hence, O or $\sigma(r)$ can be reconstructed by inverting the integral transform

$$\bar{O}_1(\mathbf{k}) = \int d\mathbf{x} e^{-i\mathbf{K} \cdot \mathbf{x}} \bar{O}_1(\mathbf{x}, \mathbf{k}) \quad (21)$$

where

$$\bar{O}_1(\mathbf{x}, \mathbf{k}) = \int dz O_1(\mathbf{x}, z) e^{-i\gamma z} \quad (22)$$

Equation (21) is a horizontal, two-dimensional, spatial Fourier transform that can be inverted by fast Fourier transform (FFT) techniques; however, (22) is a one-dimensional Fourier transform with a complex wavenumber γ and, therefore, cannot be inverted using an FFT. For this reason, the inversion is a two-step process where, in the first step, the integral transform

$$\bar{\bar{O}}_1(\mathbf{k}) = \int dz \bar{O}_1(K, z) \frac{e^{-i\gamma z}}{h - z} \quad (23)$$

is inverted and then $\bar{O}_1(K, z)$ is inverted by a standard two-dimensional FFT. Equation (23) is inverted by approximating the integral by the quadrature formula

$$\bar{\bar{O}}_1(K, \gamma_i) \approx \sum_j A_{ij} \bar{O}_1(K, z_j) \quad (24)$$

where

$$A_{ij} = e^{-i\gamma_j z_i} \quad (25)$$

$$\gamma_i = \sqrt{4i\delta_i^{-2} - |K|^2} \quad (26)$$

$$z_j = z_0 + (j - 1)\Delta z$$

and

$$\delta_i = \delta_0 + (i - 1)\Delta\delta.$$

Equation (24) represents a mapping from uniformly spaced sampling in z , beginning at a minimum depth z_0 , in increments of Δz , into a uniformly spaced sampling in skin depth from δ_0 , in increments of $\Delta\delta$.

Equation (24) could be solved by inverting the matrix A_{ij} ; however, the system of equations is ill-conditioned so that it becomes necessary to seek a less direct method of inversion. The method used here is regularization where, rather than considering (24), the equation

$$\bar{\bar{O}}_1(K, \gamma_i) = \sum_j (A_{ij} - \lambda I_{ij}) \bar{O}_1(K, z_j) \quad (27)$$

is inverted,
parameter.
parameter. If
will be excess
Following
computed by
be reconstruct
turn, be comp

4. Results

The forward m
are three dime
developments
surface. For s
slightly more
This model can
section 3. For
et al 1992) wa

4.1. Two-dimen

Images from th
and (20) and (
to be two dime
considered. Th
horizontal coils
(1, 0, 0); and h
the transect line

Extensive to
simple shapes. I
field and a large
coil orientations
coil orientation.
orientations and
those associated
images: these ar
For all results pr
a 16 by 4 rectan
at 32 horizontal
[-32, 30].

Figure 2 pro
configuration dis
that, for relative
however, the ima
small regularizati
of the simulated
in this figure, 16
established by W

is inverted, where I_{ij} is the identity matrix and λ is referred to as the regularization parameter. The inversion of (27) must be optimized with respect to the regularization parameter. If λ is too small, the inversion will be unstable and if λ is too large, the image will be excessively blurred.

Following the inversion of (27) by regularization for each value of K , $O_1(\mathbf{r})$ can be computed by inversion of the two-dimensional Fourier transform. The object function can be reconstructed from $O(\mathbf{r}) = (h - z)O_1$ and the spatial variations in conductivity can, in turn, be computed using $\sigma(\mathbf{r}) = \sigma_0(O(\mathbf{r}) + 1)$.

4. Results

The forward model derived in section 2 and the inversion algorithm derived in section 3 are three dimensional and allow arbitrary primary and secondary coil orientations. These developments are restricted to cases where both coils are in intimate contact with the ground surface. For situations where the coils are above the ground surface, the formulation is slightly more complicated and the forward model for this case is given in appendix A. This model can be used to develop an inversion algorithm following the procedure given in section 3. For all the results presented in this section, a standard numerical scheme (Press *et al* 1992) was used to implement the regularized inversion defined in (27).

4.1. Two-dimensional synthetic data

Images from broadband EMI are presented using (17) to simulate the broadband data, and (20) and (27) are used to reconstruct the image. For simplicity, the targets are taken to be two dimensional, varying only in the x, z plane, and only three coil orientations are considered. These are: vertical coils—both coil axes are vertical, $\hat{n}_0 = \hat{n}_s = (0, 0, 1)$; horizontal coils—both coil axes are horizontal and parallel to the transect line, $\hat{n}_0 = \hat{n}_s = (1, 0, 0)$; and horizontal coplanar coils—both coil axes are horizontal and perpendicular to the transect line, $\hat{n}_0 = \hat{n}_s = (0, 1, 0)$.

Extensive testing was performed on all coil orientations for a variety of targets of simple shapes. Since the vertical coil orientation produces a smaller gradient in the vertical field and a larger gradient in the horizontal field, as compared to either of the horizontal coil orientations, the forward model results exhibited noticeable differences with varying coil orientation. In contrast, reconstructed images were quantitatively similar for all coil orientations and target shapes. For this reason, the results presented here are limited to those associated with variations in the parameters that affect the quality of reconstructed images: these are the choice of regularization parameter λ (27) and the skin depth interval. For all results presented, a horizontal coil orientation was assumed, the inhomogeneity was a 16 by 4 rectangular conducting anomaly centred at $(0, -16)$, and data were synthesized at 32 horizontal measurement points at $z = 0$ and uniformly distributed over the interval $[-32, 30]$.

Figure 2 provides simulated results for the above described target and measurement configuration displaying the effects of varying the regularization parameter. It is expected that, for relatively large values of this parameter, the inversion should be quite stable; however, the image will be blurred as a result of excessive smoothing. For an extremely small regularization parameter, the inversion will be unstable. Figure 2(a) is the real part of the simulated data where the vertical axis is skin depth. For the simulations presented in this figure, 16 skin depths uniformly distributed on the interval $[2, 32]$ were used. As established by Won (1980), the simulated data appear as an out-of-focus image of the

rectangular target, with the target appearing as a negative anomaly. The reconstructed images, figures 2(b) through 2(e), exhibit the expected result of increased image sharpness with decreasing regularization parameter. An important finding is that the inversion is quite stable and reasonable images were reconstructed over a range of regularization parameter that spans six orders of magnitude.

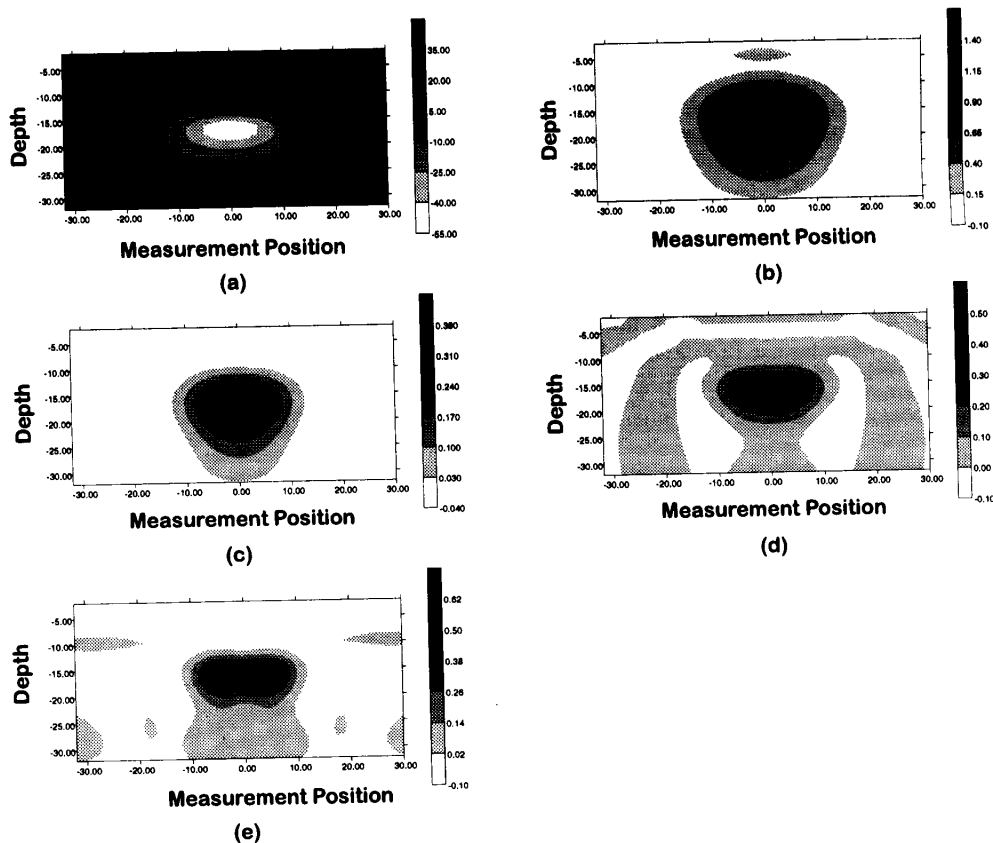


Figure 2. Data and images of a 16×4 conducting rectangle centred at $(0, -16)$ for 32 measurement points uniformly distributed over the interval $[-32, 30]$ along the ground surface, 16 skin depths uniformly distributed on $[2, 32]$ and a horizontal coil axis for (a) the real part of the simulated data and images for regularization parameters of (b) $1e-4$, (c) $1e-6$, (d) $1e-8$, and (e) $1e-10$.

Figures 3 and 4 are reconstructed images displaying the effect of varying the skin depth interval. For all reconstructions, data based on 16 skin depths were used. Figure 3 demonstrates the effect of broadening the range of skin depths and, in this figure, the minimum skin depth is always significantly less than the depth of the conducting rectangle, while the maximum skin depth is progressively increased from 8 (figure 3(a)) to 128 (figure 3(e)). Poor images are obtained when the maximum skin depth is less than the deepest extent of the inhomogeneity (figures 3(a) and (b)). For a maximum skin depth of 8 (figure 3(a)), the reconstructed rectangle has the proper shape, but is too shallow, and increasing the maximum skin depth to 16 (figure 3(b)) results in a reconstruction of the rectangle that is somewhat deeper; however, it is still too shallow and the contrast is reversed. Increasing the maximum skin depth to 24 (figure 3(c)) results in a reasonable reconstruction.

The reconstruction at a maximum skin depth of 128 (figure 3(e)) is a somewhat sharp image, but the contrast is reversed. When the skin depth is 128, the reconstructed image is quite good, but the contrast is reversed. It is clear that a reasonable reconstruction can be obtained for such a coarse set of skin depths.

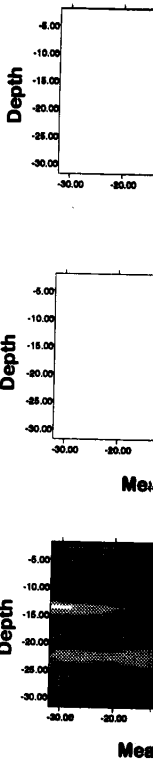


Figure 4. Data and images of a 16×4 conducting rectangle centred at $(0, -16)$ for 32 measurement points uniformly distributed over the interval $[-32, 30]$ along the ground surface, 16 skin depths uniformly distributed on $[2, 32]$ and a horizontal coil axis for (a) the real part of the simulated data and images for regularization parameters of (b) $1e-4$, (c) $1e-6$, (d) $1e-8$, and (e) $1e-10$.

The reconstructed image is excellent for a maximum skin depth of 48 (figure 3(d)) and even at a maximum skin depth of 128 (figure 3(e)) the reconstructed image is quite good. This is a somewhat surprising result, since for 16 skin depths uniformly distributed on the interval $[2, 128]$, the skin depth increment is almost 8, which is twice the thickness of the rectangular anomaly. While skin depth is somewhat analogous to time in wave-based methods, it is clear that a reflector may be entirely absent in the data for a coarse time sampling; however, such a coarse sampling in skin depths does not profoundly impact the quality of the EMI images.

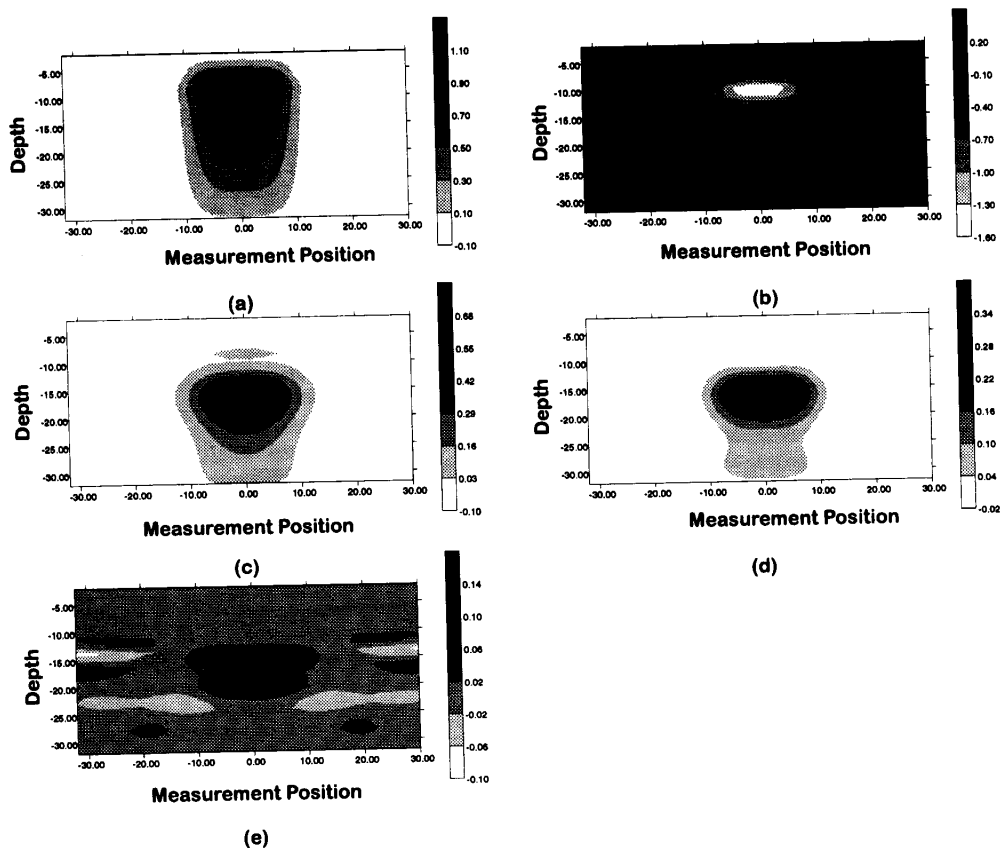


Figure 3. Images of a 16×4 conducting rectangle centred at $(0, -16)$ for 32 measurement points uniformly distributed on the interval $[-32, 30]$ along the ground surface, a regularization parameter of $1e-8$, a horizontal coil axis, and 16 skin depths uniformly distributed over (a) $[1, 8]$, (b) $[1, 16]$, (c) $[2, 24]$, (d) $[2, 48]$, and (e) $[2, 128]$.

Figure 4 demonstrates the effects of simultaneously varying the minimum skin depth and the skin depth interval. The purpose of these simulations is to investigate implications of using a minimum skin depth that is greater than the target depth. Figure 4(a) is a reconstructed image for a minimum skin depth of 8 and a maximum skin depth of 32. Here, the range of skin depths spans the depth interval of the conductor and the reconstructed image is quite good. When the skin depth does not completely span the vertical extent of the target (figure 4(b)), the image is still good. This is in complete contrast to figure 3(b) where the deeper skin depths, rather than the shallower skin depths, are absent. Increasing the minimum skin depth (figure 4(c)) and further increasing the minimum skin depth and

broadening the skin depth interval (figure 4(d)) results in a progressive blurring of the image. The reason that skin depths less than the target depth are not required to adequately image inhomogeneities is apparent in the simulated data (figure 2(a)). It is clear in this figure that the gradient with respect to skin depth is much steeper for skin depths less than target depth, than for skin depths greater than the target depth. Furthermore, for skin depths less than about one-half the target depth, the secondary field is almost zero. For skin depths substantially larger than the target depth, the secondary field decays slowly with increasing skin depth and there is information regarding target structure encoded in these large skin depths.

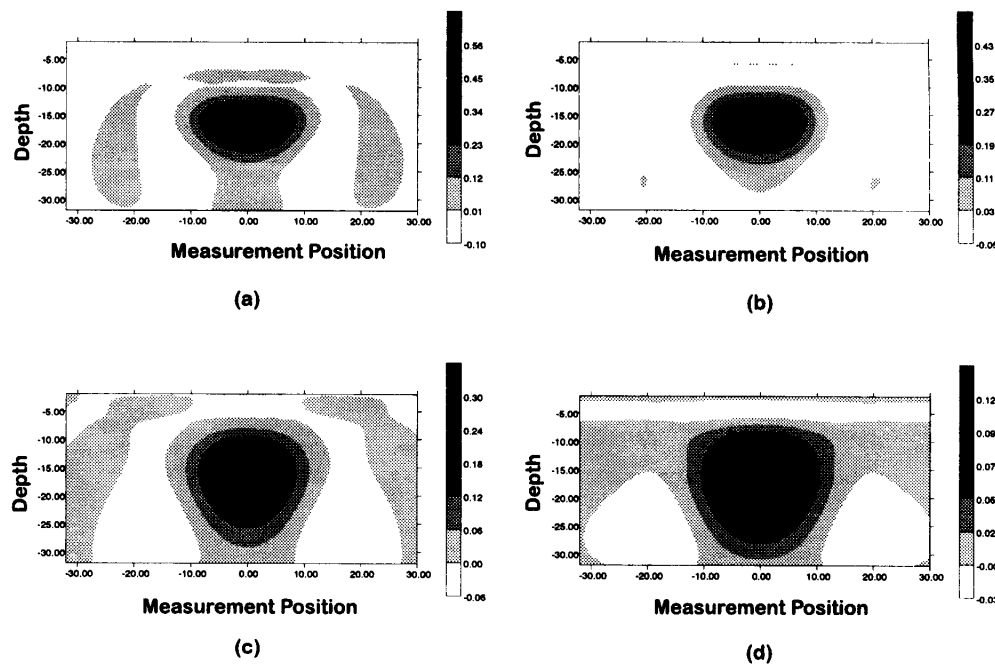
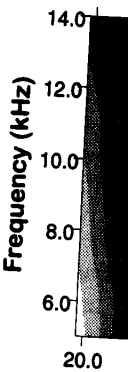


Figure 4. Images of a 16×4 conducting rectangle centred at $(0, -16)$ for 32 measurement points uniformly distributed on the interval $[-32, 30]$ along the ground surface, a horizontal coil axis, and 16 skin depths uniformly distributed over (a) $[8, 32]$, (b) $[16, 48]$, (c) $[32, 64]$, and (d) $[64, 128]$. The regularization parameter used for (a) and (b) is $1e-8$ and for (c) and (d) is $1e-6$.

While the images given above represent quite good reconstructions, they were obtained from noise-free simulated data. Since the inversion of (24) can be ill-conditioned, it can be expected that signal noise can have a profound impact on image quality, particularly when all skin depths are greater than the target depth. To address this problem, additional imaging was performed with noise added to the synthetic data. Two noise models were considered; the relative noise case where the signal-to-noise ratio was defined relative to the maximum signal at each skin depth, and the absolute noise case where the signal-to-noise ratio was defined relative to the maximum signal over all skin depths independent of the skin depths used in the image reconstruction. Noise was added to the signal by first generating a sequence of random numbers uniformly distributed over the interval $[-1, 1]$. After dividing by the signal-to-noise ratio, these random numbers were added to the simulated data. For both noise models, images were reconstructed for a variety of ranges of skin depths. The

results of



- (1) Decrease in the signal-to-noise ratio at a point, can be expected to have a profound impact on image sharpness.
- (2) There is a trade-off between the reconstructed image-to-noise ratio and the signal-to-noise ratio respectively.
- (3) Signal-to-noise ratio is greater than the target depth.

4.2. Two-dimensional

While the results of the two-dimensional imaging are good, the skin depths are small. As the skin depths increase, the signal-to-noise ratio decreases substantially. This is an important issue that needs to be addressed. Images can be reconstructed more importantly, given by (17) and (18), which are required to achieve a target depth of inhomogeneity.

results of these simulations are summarized below.

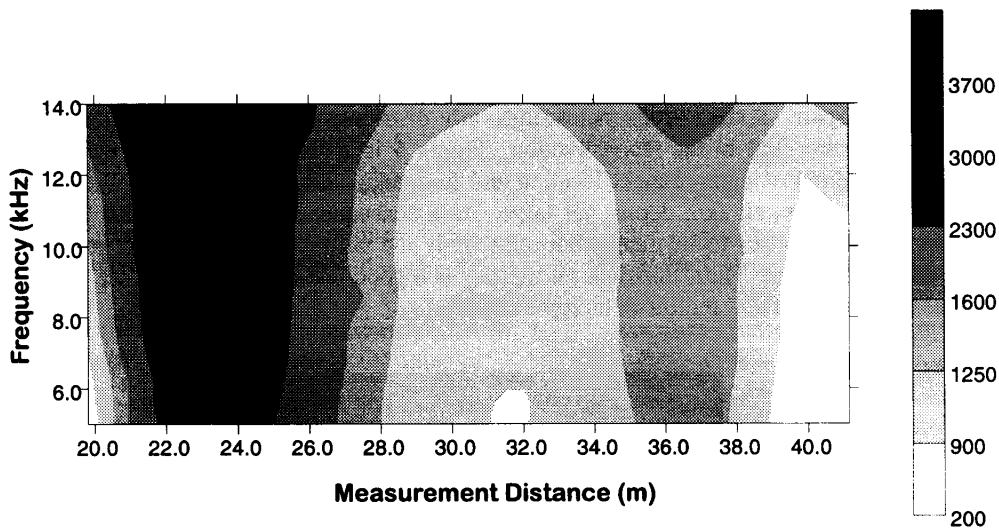


Figure 5. Plot of the real part of the ratio of primary to secondary field as a function of measurement location and frequency for EMI data acquired over the top of a pair of subway tunnels.

(1) Decreasing the signal-to-noise ratio produces a degradation in image quality that, to a point, can be offset by increasing the regularization parameter but with an associated loss of image sharpness.

(2) There is a signal-to-noise ratio lower limit below which the target cannot be reconstructed no matter how large the regularization parameter. These limiting signal-to-noise ratio values are about 100 and 20 for the absolute and relative noise models, respectively.

(3) Signal noise did not affect the images reconstructed for cases where all skin depths are greater than the target depth more than for cases where the skin depths span the target depth.

4.2. Two-dimensional field data

While the results based on synthetic data establish that images can be reconstructed when all skin depths are significantly greater than the target depth, it must be recognized that, as skin depths increase beyond the target depth, the magnitude of the secondary field decreases substantially. Since the signal-to-noise ratio for synthetic data is essentially infinite, an important issue related to the viability of geophysical EMI imaging is whether adequate images can be reconstructed for large skin depths when the signal-to-noise ratio is finite, or more importantly, small. The synthetic data examples considered here use the forward model given by (17) and, as such, do not address the issue of the weak far-field approximation required to achieve the GPST. The validity of this approximation will be tested provided that all, or some, of the frequencies acquired correspond to skin depths greater than the depth of inhomogeneities.

To address these concerns, broadband EMI data were acquired along a line on the ground surface over a parallel pair of subway tunnels near the Washington, DC, Anacostia Metro station. The broadband geophysical tool used was the GEM-2 (Won *et al* 1996). A subset of this data set was used for imaging. Sixteen measurement points were uniformly distributed over the interval [18 m, 41 m] on the ground surface perpendicular to the axes of the tunnel pair. Sixteen frequencies uniformly distributed over the interval [5010 Hz, 14 010 Hz] were used in the reconstruction. Figure 5 is a display of the real part of the data as a function of measurement location and frequency. Since the range of frequencies shown is quite low, all skin depths are large. As a consequence, the data as displayed do not appear as an out-of-focus image, but rather only the tail of the gradient with respect to skin depth is evident. Thus, this data set represents a good test of both the impact of a low signal-to-noise ratio and the far-field approximation.

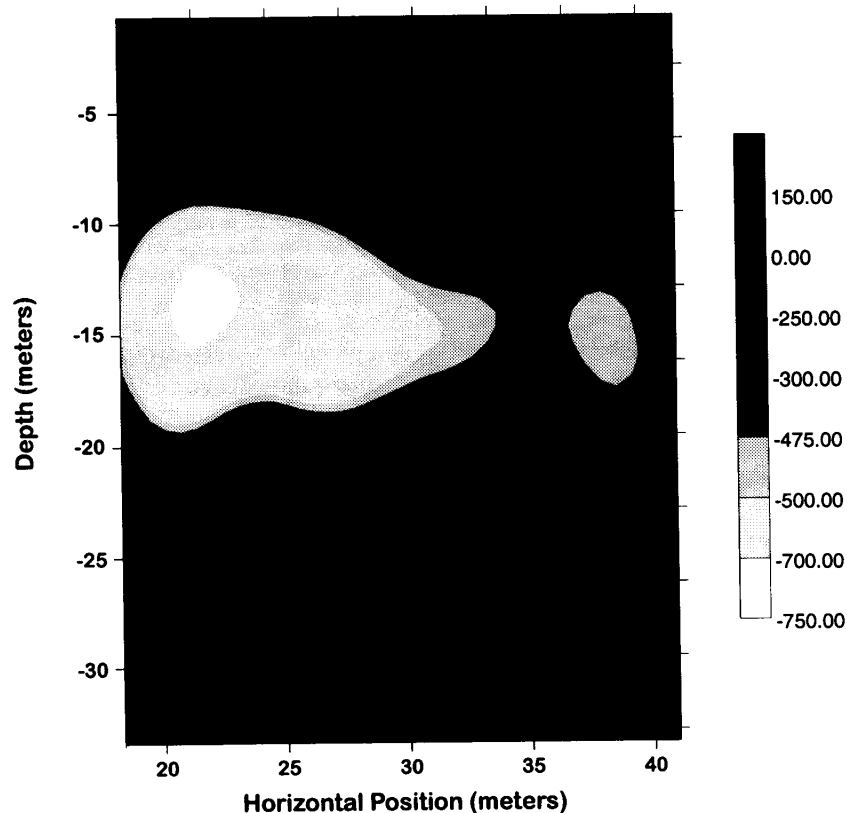


Figure 6. Reconstructed image of a pair of parallel subway tunnels near the Washington, DC, Anacostia Metro Station. The tunnel cross sections appear as the near-circular, low conductivity anomalies at a depth of about 12 m. The near-surface high conductivity layer is believed to be saturated soil.

Figure 6 is the image reconstructed from these data based on a regularization parameter of $1e-6$ and an assumed background conductivity of $0.1 \text{ Siemen m}^{-1}$. The cross sections of the two tunnels are clearly evident as the near-circular low conductivity anomalies at a depth of about 12 m. In addition, there is an apparent high conductivity layer in the upper 3 m of the image. There was standing water on the ground at the time the data were

acquired, getting (the right-hand) and near-surface, however, their

4.3. Three-dimensional

To demonstrate measurements was employed, centre at (0, 0, - depths uniform regularization p image through of the target loc reconstructions is based on the simulations.

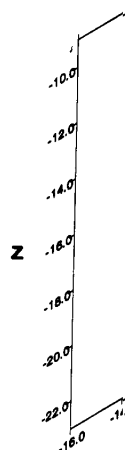


Figure 7. Reconstructed image of a pair of parallel subway tunnels near the Washington, DC, Anacostia Metro Station. The tunnel cross sections appear as the near-circular, low conductivity anomalies at a depth of about 12 m. The near-surface high conductivity layer is believed to be saturated soil.

acquired, getting progressively deeper to about 0.3 m at one end of the measurement line (the right-hand side of figure 5), and this layer is likely a result of the presence of surface and near-surface water. The depth of the tunnels at the measurement line is not known, however, their imaged depth is consistent with available estimates.

4.3. Three-dimensional synthetic data

To demonstrate EMI imaging in three dimensions, broadband EMI data was synthesized for measurements made over a two-dimensional grid on the ground surface. A 16×16 grid was employed, where $-16 \leq x, y \leq 14$. The target was a rectangular solid conductor with centre at $(0, 0, -16)$ and having dimensions $(8, 16, 4)$. Data were synthesized for eight skin depths uniformly distributed on the interval $[2, 32]$. The three-dimensional reconstruction was based on eight depth points uniformly distributed on the interval $[-9, -23]$ and a regularization parameter of $1e-8$. Figure 7 is a 'chair cut' of the three-dimensional image through the centre of the inhomogeneity. This figure is an accurate representation of the target location and aspect ratio. While this image is not as sharp as the synthetic reconstructions for the two-dimensional rectangle, it must be recognized that this image is based on the equivalent of one-quarter of the data points used in the two-dimensional simulations.

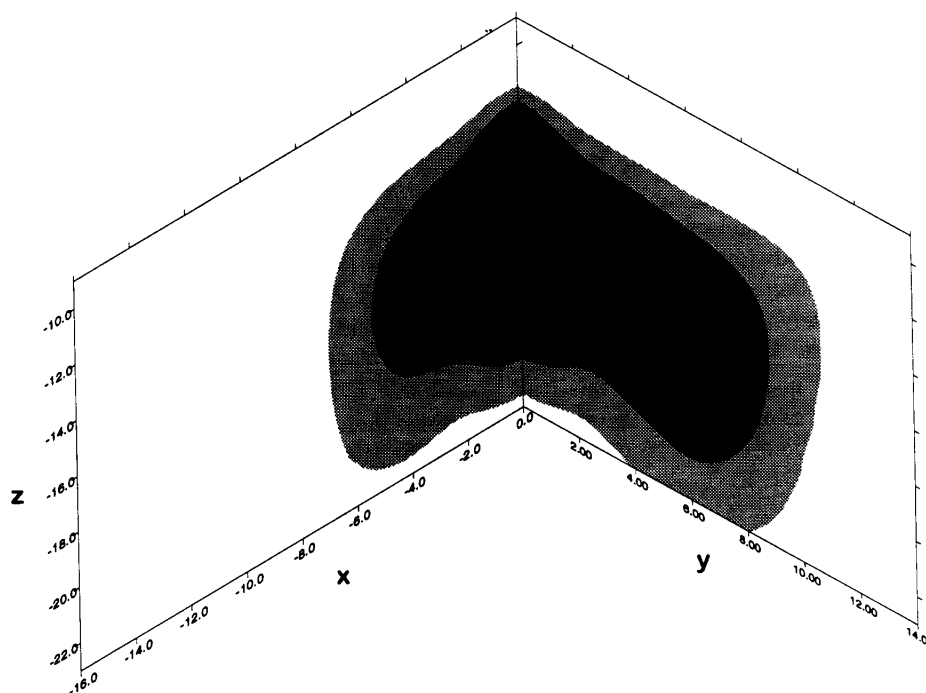


Figure 7. 'Chair cut' through a three-dimensional image of a reconstructed rectangular solid conductor.

5. Conclusions

A three-dimensional imaging algorithm for broadband EMI data has been developed and implemented. This algorithm has been tested on two- and three-dimensional synthetic data and on two-dimensional field data. It has been shown that EMI imaging can yield good reconstructions of both conductive and resistive targets, even for a relatively small number of frequencies. Perhaps the most important finding is that adequate images can be obtained when the highest frequency is so low that all skin depths are greater than the target depth. This is exemplified by the reconstructed image of the tunnels (figure 6) where the tunnels are manifested in the data as only a gradient in frequency (figure 5), rather than a well defined peak in response at a skin depth comparable to the target depth (figure 2(a)). This suggests that there is some hope for imaging relatively small and shallow targets with the relatively low-frequency EMI tools.

The imaging approach is based on the concept of diffraction tomography where an analytic relationship is established between the spatial Fourier transform of the acquired data and a spatial integral transform of the variations in conductivity. For EMI imaging, the integral transform is Fourier (real wavenumber) in the horizontal directions while the vertical component of the integral transform is characterized by a complex wavenumber. For this reason, the full integral transform cannot be inverted by the computationally efficient fast Fourier transform procedure. The mapping from depth into skin depth must be inverted by matrix inversion at every horizontal wavevector point, which is far less efficient. Despite this limitation, the imaging algorithm is computationally efficient. For example, the three-dimensional image reconstruction required only several seconds on a Pentium-based processor. For production-mode processing where image depths and skin depths are fixed, the algorithm can be rendered more computationally efficient by employing the 'pseudo-inverse' concept developed by Deming and Devaney (1996). Using this approach, the matrix inversion becomes independent of the acquired data and, as a consequence, can be pre-computed and stored. Furthermore, the pseudo-inverse explicitly accounts for the band-limited nature of the acquired data and, for this reason, promises better images when a limited bandwidth and a small number of frequencies are used.

In the derivation of the imaging algorithm, it is assumed that the host formation is completely characterized by its conductivity and is manifested in the background wavenumber $k_0 \approx \sqrt{i\omega\mu_0\sigma_0}$. This assumption was made as a matter of convenience rather than necessity. Since this background wavenumber is complex, no complications arise from defining $k_0 = \sqrt{i\omega\mu_0(-i\omega\epsilon_0 + \sigma_0)}$ and thereby characterizing the host formation in terms of both its dielectric and conductivity. Similarly, the object function was assumed real, where $O(\mathbf{r}) = [(\sigma(\mathbf{r})/\sigma_0) - 1]$. The object function is, in general, complex, so that spatial variations in conductivity can be reconstructed from the real part of O , while spatial variations in dielectric can be reconstructed from $\text{Im}(O)$.

Appendix. Introduction of ground-air planar interface effects

In section 2 the formulation for imaging a finite support heterogeneity from a suite of multimonostatic experiments with loop antennas scanning the $z = h, h \geq 0$ was presented, and it was assumed that the background conductivity (and dielectric) was the same for both half-spaces $z \geq 0$ and $z < 0$. The fact that the medium is a good conductor affects the formulation only in modifying the plane wavenumber from purely real in dielectric media to complex with a 45° phase. The 'propagation' effects are now called electromagnetic induction or diffusion, but other than terminology, nothing has essentially

Subj

changed compared is that the general scattered fields to the frequencies, a fact

In this appendi embedded in a con is limited to the ad weak, such that its themselves are neg loop antennas that distance from the i

A.1. Dyadic Fresnel an air-ground inte

The geometry consi by a wavenumber i wavenumber $k_1 =$

The scalar Fres the continuity con boundary at $z = 0$

$$\tilde{T}_+(K) = \frac{2m_0}{m_0 + m_{11}}$$

$$\tilde{T}_-(K) = \frac{2m_1}{m_0 + m_{11}}$$

such that the tange

 $(\mathcal{I} -$
 $(\mathcal{I} -$

The full vector elec Fresnel transmission

$$\tilde{E}_1(K) = \tilde{T}_+(K) \cdot$$

$$\tilde{E}_0(K) = \tilde{T}_-(K) \cdot$$

where the z comp the electric field w $(K - m_0\hat{z}) \cdot \tilde{E}_0(K)$

A.2. Scattered elect boundary

In section 2, it was due to a source ant

$$\tilde{E}^s(K'; r_0) = (\mathcal{I} -$$

$$\tilde{E}_0(K; r_0) = \tilde{E}_0(K)$$

changed compared to the electromagnetic dielectric case. A significant effect on the problem is that the generalized projection slice theorem (GPST) now relates plane wave spectra of scattered fields to the Fourier transform of the scatterer profile evaluated at complex spatial frequencies, a fact that poses special difficulties for the inverse (imaging) problem.

In this appendix, the work of section 2 is extended to the case where the scatterer is embedded in a conducting half space at $z < 0$, where the access for interrogating the object is limited to the adjacent free space at $z \geq 0$. It is assumed that the scatterer is sufficiently weak, such that its coupling to the $z = 0$ interface is negligible, and also that the antennas themselves are negligibly coupled to this interface (which is always true in the case of loop antennas that are sufficiently small compared to the wavelength and compared to their distance from the interface).

A.1. Dyadic Fresnel coefficients for transmission of plane electromagnetic waves through an air-ground interface

The geometry considered is shown in figure A.1 where the ground at $z < 0$ is characterized by a wavenumber $k_0 = (i\omega\mu_0\sigma_0)^{1/2}$ and the air at $z \geq 0$ is characterized by a purely real wavenumber $k_1 = \omega(\mu_0\epsilon_0)^{1/2}$.

The scalar Fresnel transmission coefficients are readily obtained through application of the continuity condition of the tangential electric field and its normal derivative across the boundary at $z = 0$

$$\begin{aligned}\tilde{T}_+(K) &= \frac{2m_0}{m_0 + m_1} & m_0(K) &= \sqrt{k_0^2 - K^2} & \text{Im}\{m_0\} \geq 0, \text{Re}\{m_0\} \geq 0 \\ \tilde{T}_-(K) &= \frac{2m_1}{m_0 + m_1} & m_1(K) &= \sqrt{k_1^2 - K^2} & \text{Im}\{m_1\} \geq 0, \text{Re}\{m_1\} \geq 0\end{aligned}\quad (\text{A.1})$$

such that the tangential electric fields are related through

$$\begin{aligned}(\mathcal{I} - \hat{z}\hat{z}) \cdot \tilde{E}_1(K) &= \tilde{T}_+(K)(\mathcal{I} - \hat{z}\hat{z}) \cdot \tilde{E}_0^{\text{inc}}(K) \\ (\mathcal{I} - \hat{z}\hat{z}) \cdot \tilde{E}_0(K) &= \tilde{T}_-(K)(\mathcal{I} - \hat{z}\hat{z}) \cdot \tilde{E}_1^{\text{inc}}(K).\end{aligned}\quad (\text{A.2})$$

The full vector electric fields across the $z = 0$ boundary are thus related through a dyadic Fresnel transmission coefficient

$$\begin{aligned}\tilde{E}_1(K) &= \tilde{T}_+(K) \cdot \tilde{E}_0^{\text{inc}}(K) & \tilde{T}_+(K) &= \left[\mathcal{I} - \hat{z} \left(\hat{z} + \frac{K}{m_1(K)} \right) \right] \tilde{T}_+(K) \\ \tilde{E}_0(K) &= \tilde{T}_-(K) \cdot \tilde{E}_1^{\text{inc}}(K) & \tilde{T}_-(K) &= \left[\mathcal{I} - \hat{z} \left(\hat{z} - \frac{K}{m_0(K)} \right) \right] \tilde{T}_-(K)\end{aligned}\quad (\text{A.3})$$

where the z components have been derived by invoking the transversality condition of the electric field with respect to the plane wavevector $(K + m_1\hat{z}) \cdot \tilde{E}_1(K) = 0$ and $(K - m_0\hat{z}) \cdot \tilde{E}_0(K) = 0$.

A.2. Scattered electric field within the Born approximation in the presence of a planar boundary

In section 2, it was shown (8) that the plane wave spectrum of the scattered field at $z = 0$, due to a source antenna positioned at r_0 in a homogeneous background k_0 , is given by

$$\begin{aligned}\tilde{E}^s(K'; r_0) &= (\mathcal{I} - \hat{k}'_0 \hat{k}'_0) \cdot \frac{i}{2m'_0} \int dK \tilde{E}_0(K; r_0) \int dr' f(r') e^{-i(k'_0 - k_0) \cdot r'} \\ \tilde{E}_0(K; r_0) &= \tilde{E}_0(K) e^{-ik_0 \cdot r_0}\end{aligned}\quad (\text{A.4})$$

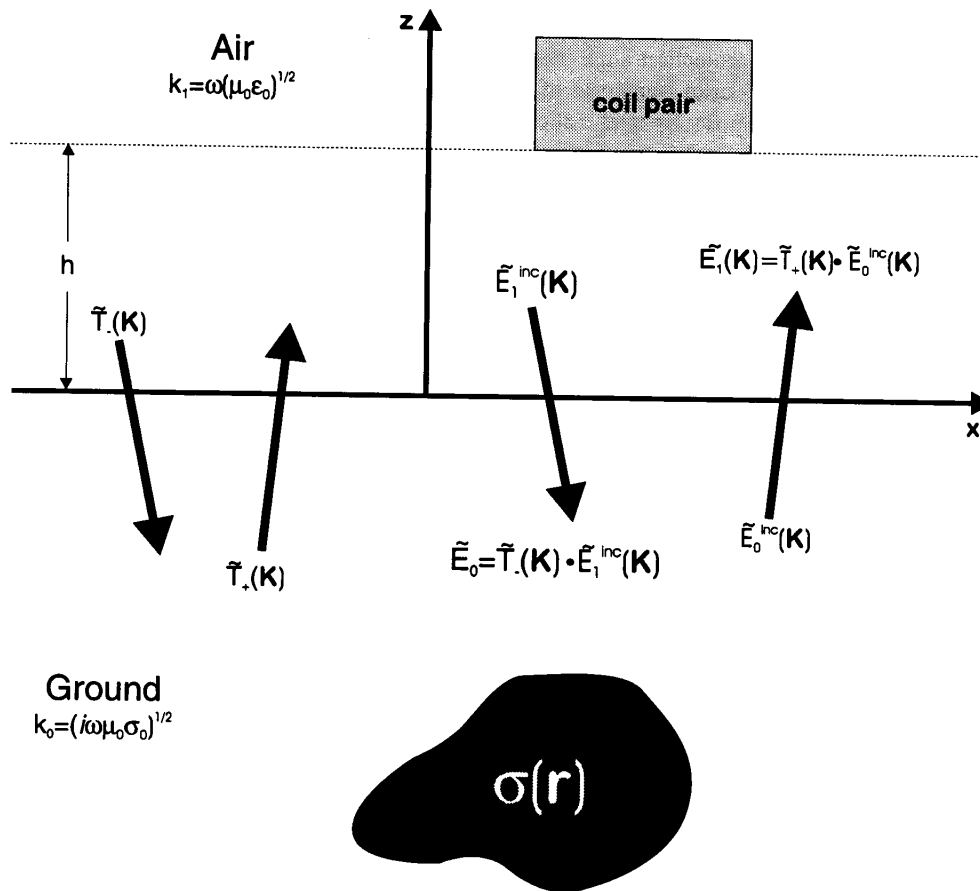


Figure A.1. Plane wave spectral relations at an interface between air and ground for the geometry under consideration.

where

$$\begin{aligned} k_{0\pm}(K) &= K \pm m_0(K)\hat{z} & m_0(K) &= \sqrt{k_0^2 - K^2} & \text{Im}\{m_0\} &\geq 0 & K &= |K| \\ k'_{0\pm} &= k_{0\pm}(K') & m'_0 &= m_0(K') \end{aligned} \quad (\text{A.5})$$

and $\tilde{E}_0(K)$ is the plane wave spectrum of the electric field emitted by the source antenna when positioned at reference location $r_0 = 0$. $\tilde{E}^s(K'; r_0)$ is the plane wave spectrum of the scattered field at $z = 0$ when the source antenna is located at r_0 .

If it is assumed that the upper half space $z \geq 0$ is free space $k_1 \neq k_0$, and also assumed that the scatterer is negligibly coupled to the $z = 0$ interface, then the above equation provides an expression for the scattered field incident on the $z = 0^-$ boundary with \tilde{E}_0 being the source field transmitted into the lower domain at $z = 0^-$. In terms of the electric fields at $z = 0^+$,

$$\begin{aligned} \tilde{E}_1^s(K'; r_0) &= \tilde{T}_+(K')(\mathcal{I} - \hat{k}'_{0+}\hat{k}'_{0+}) \cdot \frac{i}{2m'_0} \int dK \tilde{T}_-(K) \tilde{E}_1^{inc}(K; r_0) \\ &\times \int dr' f(r') e^{-i(k'_{0+} - k_{0-}) \cdot r'} \end{aligned}$$

$$\tilde{E}_1^{inc}(K; r_0) =$$

where the plan

For later use, according to

to obtain the d

where the c -dim

The doubly tran given by

As found in sect 0) field is relate

such that, in ter by the following electric field

$$\tilde{E}_1^s(K'; K) =$$

Note that thi and does not com boundary as well into the formulati arrival of these c

A.3. Measured sig

In this section we $z = z_0$ in synchr current I_0 .

$$\tilde{E}_1^{\text{inc}}(\mathbf{K}; \mathbf{r}_0) = \tilde{E}^{\text{inc}}(\mathbf{K}) e^{-i\mathbf{k}_1 \cdot \mathbf{r}_0} \quad (\text{A.6})$$

where the plane wavevector in the upper half space is defined by

$$\mathbf{k}_{1\pm} = \mathbf{K} \pm m_1(\mathbf{K})\hat{z} \quad m_1(\mathbf{K}) = (k_1^2 - K^2)^{1/2} \quad \text{Im}\{m_1\} \geq 0. \quad (\text{A.7})$$

For later use, it is desirable to Fourier transform \mathbf{x}_0 , the lateral spatial variable of \mathbf{r}_0 , according to

$$\tilde{\tilde{E}}_1^s(\mathbf{K}'; \mathbf{K}, z_0) = \int d\mathbf{x}_0 e^{-i\mathbf{K} \cdot \mathbf{x}_0} \tilde{E}_1^s(\mathbf{K}'; \mathbf{x}_0, z_0) \quad (\text{A.8})$$

to obtain the doubly Fourier-transformed multimonostatic scattered field

$$\begin{aligned} \tilde{\tilde{E}}_1^s(\mathbf{K}'; \mathbf{K}, z_0) &= (2\pi)^{c-1} \frac{i}{2m'_0} e^{im_1 z_0} \tilde{\mathcal{D}}(\mathbf{K}'; \mathbf{K}) \cdot \tilde{E}_1^{\text{inc}}(-\mathbf{K}) \tilde{f}(\mathbf{k}'_{0+} + \mathbf{k}_{0+}) \\ \tilde{\mathcal{D}}(\mathbf{K}'; \mathbf{K}) &= \tilde{T}_+(\mathbf{K}')(\mathcal{I} - \hat{\mathbf{k}}'_{0+} \hat{\mathbf{k}}_{0+}) \tilde{T}_-(-\mathbf{K}) \end{aligned} \quad (\text{A.9})$$

where the c -dimensional Fourier transform of the object function is defined according to

$$\tilde{f}(\mathbf{q}) = \int d\mathbf{r} e^{-i\mathbf{q} \cdot \mathbf{r}} f(\mathbf{r}). \quad (\text{A.10})$$

The doubly transformed multimonostatic scattered field referenced to $z_0 = 0$ is therefore given by

$$\tilde{\tilde{E}}_1^s(\mathbf{K}'; \mathbf{K}) = (2\pi)^{c-1} \frac{i}{2m'_0} \tilde{\mathcal{D}}(\mathbf{K}'; \mathbf{K}) \cdot \tilde{E}_1^{\text{inc}}(-\mathbf{K}) \tilde{f}(\mathbf{k}'_{0+} + \mathbf{k}_{0+}). \quad (\text{A.11})$$

As found in section 2, the plane wave spectrum of the transmitting antenna (subscripted by 0) field is related to its vector effective height $\tilde{\mathbf{h}}_0$ through

$$\tilde{E}_1^{\text{inc}}(\mathbf{K}) = -I_0 \frac{\omega\mu_0}{2m_1} \tilde{\mathbf{h}}_0(-\mathbf{K}) \quad (\text{A.12})$$

such that, in terms of the vector effective height of the transmitting antenna, it is given by the following expression for the doubly Fourier-transformed multimonostatic scattered electric field

$$\tilde{\tilde{E}}_1^s(\mathbf{K}'; \mathbf{K}) = -(2\pi)^{c-1} I_0 \frac{i\omega\mu_0}{4m_1 m'_0} \tilde{\mathcal{D}}(\mathbf{K}'; \mathbf{K}) \cdot \tilde{\mathbf{h}}_0(\mathbf{K}) \tilde{f}(\mathbf{k}'_{0+} + \mathbf{k}_{0+}). \quad (\text{A.13})$$

Note that this is only the component of the field scattered off the interior of $z < 0$ and does not contain the undesired (strong) component of the field reflected off the $z = 0$ boundary as well as the direct field. The undesired components can be easily incorporated into the formulation or, alternatively, if illuminating the ground with short pulses, the early arrival of these components can be effectively time gated.

A.3. Measured signal in a multimonostatic measurement configuration above ground

In this section we obtain the terminal signal at a receiving antenna, that sweeps a plane $z = z_0$ in synchronization with the transmitting antenna that is excited with a terminal current I_0 .

Assuming that the scattered signal $\tilde{E}_1^s(K; r_0)$ is received by a receiving antenna that is positioned at r_1 such that if it were to transmit, its current distribution would be $J_1(r - r_1)$, with terminal current I_1 . By the Lorentz reciprocity theorem, it is known that the open circuit terminal voltage is given by

$$\begin{aligned} V(r_1; r_0) &= -\frac{1}{I_1} \int d\mathbf{r}' J_1(\mathbf{r}' - r_1) \cdot \mathbf{E}_1^s(\mathbf{r}'; r_0) = -\frac{1}{I_1} \int d\mathbf{r}' J_1(\mathbf{r}') \cdot \mathbf{E}_1^s(\mathbf{r}' + r_1; r_0) \\ &= -(2\pi)^{1-c} \frac{1}{I_1} \int d\mathbf{r}' J_1(\mathbf{r}') \cdot \int d\mathbf{K}' e^{i\mathbf{K}' \cdot (\mathbf{r}' + r_1)} \tilde{E}_1^s(\mathbf{K}'; r_0) e^{im'_1(z' + z_1)} \\ &= -(2\pi)^{1-c} \int d\mathbf{K}' \tilde{h}_1(\mathbf{K}') \cdot \tilde{E}_1^s(\mathbf{K}'; r_0) e^{i\mathbf{k}'_{1+} \cdot r_1} \\ &= -(2\pi)^{2(1-c)} \int d\mathbf{K}' \tilde{h}_1(\mathbf{K}') \cdot \int d\mathbf{K} \tilde{E}_1^s(\mathbf{K}'; \mathbf{K}) e^{i(\mathbf{k}'_{1+} \cdot r_1 + \mathbf{k}_{1+} \cdot r_0)} \end{aligned} \quad (\text{A.14})$$

where the vector effective height of the antenna for fields arriving from $z < 0$ is defined in the usual way (13)

$$\tilde{h}_1(\mathbf{K}') = (\mathcal{I} - \hat{\mathbf{k}}'_{1+} \hat{\mathbf{k}}'_{1+}) \cdot \frac{1}{I_1} \int d\mathbf{r}' J_1(\mathbf{r}') e^{i\mathbf{k}'_{1+} \cdot \mathbf{r}'}. \quad (\text{A.15})$$

To this point, the measured signal is at r_1 due to a source located at r_0 . In the multimonostatic case, $r_1 = r_0$, such that

$$V(r_0; r_0) = V_{\text{mm}}(r_0) = -(2\pi)^{2(1-c)} \int d\mathbf{K}' \tilde{h}_1(\mathbf{K}') \cdot \int d\mathbf{K} \tilde{E}_1^s(\mathbf{K}'; \mathbf{K}) e^{i(\mathbf{k}'_{1+} + \mathbf{k}_{1+}) \cdot r_0} \quad (\text{A.16})$$

where V_{mm} denotes the multimonostatic signal (now dependent on only one set of coordinates). Finally, Fourier transforming with respect to the x_0 variable gives

$$\tilde{V}_{\text{mm}}(\mathbf{K}, z_0) = -(2\pi)^{1-c} \int d\mathbf{K}' e^{i[m'_1 + m_1(\mathbf{K} - \mathbf{K}')z_0]} \tilde{h}_1(\mathbf{K}') \cdot \tilde{E}_1^s(\mathbf{K}'; \mathbf{K} - \mathbf{K}'). \quad (\text{A.17})$$

Now, substituting the expression for the doubly Fourier-transformed field (A.13) yields

$$\begin{aligned} \tilde{V}_{\text{mm}}(\mathbf{K}, z_0)/I_0 &= \frac{i\omega\mu_0}{4} \int d\mathbf{K}' \frac{e^{i[m'_1 + m_1(\mathbf{K} - \mathbf{K}')z_0]}}{m_1(\mathbf{K} - \mathbf{K}')m'_0} \tilde{h}_1(\mathbf{K}') \cdot \tilde{\mathcal{D}}(\mathbf{K}'; \mathbf{K} - \mathbf{K}') \\ &\quad \cdot \tilde{h}_0(\mathbf{K} - \mathbf{K}') \tilde{f}[\mathbf{k}'_{0+} + \mathbf{k}_{0+}(\mathbf{K} - \mathbf{K}')]. \end{aligned} \quad (\text{A.18})$$

This result is in agreement with (15) when z_0 is set to zero and the Fresnel transmission dyads are set to unity. Compared to the result in section 2 (equation (15)), here the effect of the $z = 0$ interface is embodied in the dyadic denoted by $\tilde{\mathcal{D}}$.

This result relates $\tilde{V}(\mathbf{K})$ to the Fourier transform of the object function at spatial frequencies $\mathbf{k}_{0+}(\mathbf{K}') + \mathbf{k}_{0+}(\mathbf{K} - \mathbf{K}')$. It is the \mathbf{K}' integral that prevents this last equation from being an ordinary projection slice theorem relation.

References

- Deming R and Devaney A J 1996 A filtered backpropagation algorithm for GPR *J. Env. Eng. Geophys.* **0** 113-24
 Devaney A J 1982 A filtered backpropagation algorithm for diffraction tomography *Ultrasonic Imaging* **4** 336-50
 —1984 Geophysical diffraction tomography *IEEE Trans. Geosci. Remote Sensing* **GE-22** 3-13
 Molyneux J E and Witten A J 1993 Diffraction tomographic imaging in a monostatic measurement geometry *IEEE Trans. Geosci. Remote Sensing* **31** 507-10
 Morse P M and Feshbach H 1953 *Methods of Theoretical Physics* (New York: McGraw-Hill)

Norton S J and B
 Press W H, Teuk
 Cambridge U
 Witten A J, Moly
 IEEE Trans.
 Witten A J, Schat
 detection Na
 Won I J 1980 A wi
 45 928-40
 Won I J, Keiswett
 sensor J. Enw

- Norton S J and Bowler J R 1993 The theory of eddy current inversion *J. Appl. Phys.* **73** 501-12
- Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes* 2nd edn (Cambridge: Cambridge University Press)
- Witten A J, Molyneux J E and Nyquist J 1994 Ground penetrating radar tomography: algorithms and case studies *IEEE Trans. Geosci. Remote Sensing* **32** 461-7
- Witten A J, Schatzberg A and Devaney A J 1994 Signal processing of electromagnetic induction data for tunnel detection *National Security Programs Office Report K/NSP-208*
- Won I J 1980 A wide-band electromagnetic exploration method: some theoretical and experimental results *Geophys.* **45** 928-40
- Won I J, Keiswetter D A, Fields G R A and Sutton L C 1996 GEM-2: a new multifrequency electromagnetic sensor *J. Env. Eng. Geophys.* **1** 129-37

Imaging Underground Structures Using Broadband Electromagnetic Induction

Alan Witten
School of Geology and Geophysics, University of Oklahoma
Norman, Oklahoma

I.J. Won
Geophex, Inc.
Raleigh, North Carolina

Stephen Norton
Energy Division, Oak Ridge Nation Laboratory
Oak Ridge, Tennessee

ABSTRACT

Broadband electromagnetic induction is a promising technique for the detection and location of underground structures. Along with being simple to use and allowing rapid data acquisition, underground structures can be detected by this method either because the large volume of air necessary for human occupancy will be manifested as an electrical conductivity low relative to the host formation or because there could exist a large metallic mass associated with either structural support or the mission of the facility that would constitute a relative conductivity high. The use of broadband information literally offers an additional dimension, that being depth, to the level of information that can be gleaned from acquired data since, by varying the operating frequency, the probing depth changes. A three-dimensional tomographic imaging algorithm is applied to broadband data acquired at two underground structure sites to reconstruct the relative spatial variations in electrical conductivity. A pair of parallel subway tunnels exist at one site and a cloud chamber at the other. At both sites, quite good images of the structures were obtained and the structures appeared as relative conductivity lows.

Introduction

Under ideal conditions, almost any geophysical tool can be used to detect and located subsurface structures suitable for human occupancy. In practice, however, the choice of geophysical method will be dictated by the anticipated size and depth of the underground structure as well as site conditions including the mechanical and electrical properties of the host geology, whether data can be acquired directly over the underground structure, and the presence of surface or near surface cultural features.

Beyond the above-cited issues that may influence the choice of geophysical methods, there may be operational constraints that must additionally be considered. There are a number of scenarios under which site access time is severely limited and data must be acquired by operators with no geophysical training. In such cases, electromagnetic induction (EMI) is a strong candidate since EMI tools are relatively simple to operate and data can be acquired quite rapidly. Broadband EMI is particularly attractive since, by its nature, it offers information about target depth with no additional complications in data acquisition.

It has been shown (Won, 1980) that a pseudo cross-sectional image can be obtained from broadband EMI measurements made at intervals along a line. By displaying this

data as a vertical slice having a depth axis inversely proportional to the square root of the acquisition frequency, a crude form of an image of spatial variations in electrical conductivity within the cross-section is produced. This technique is referred to as frequency sounding. In a more recent development, Witten, Won, and Norton (1997) established that frequency sounding produced an out-of-focus image of conductivity anomalies and presented a generalized form of diffraction tomography (Devaney, 1984) whereby focused images may be reconstructed.

In this paper, the general concept of frequency sounding is discussed and a three-dimensional broadband EMI imaging algorithm is summarized. This algorithm is applied to broadband EMI data acquired at two sites known to contain underground structures.

EMI Imaging Concepts

The use of EMI in geophysics exploits the fact that a current will be induced in an electrically conductive anomaly, when immersed in a time-varying electromagnetic field. In turn, the current induced in conductive anomalies will induce a (secondary) time-varying electromagnetic field. Thus, EMI measurements can be used to detect and locate subsurface, electrically conductive features by creating a time-varying field

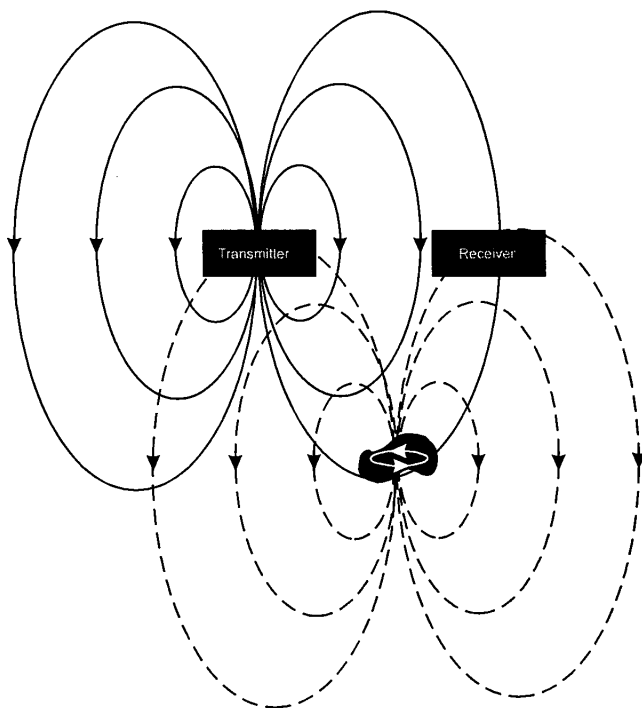


Figure 1. Illustration showing transmitting coil producing a primary field (solid contours). This field induces a current in a conductor (black mass) that, in turn, produces a secondary field (dashed contours) detected by the receiving coil.

using a transmitting coil and measuring the induced secondary field with a receiving coil (fig. 1). Because rock and soil are somewhat conductive and will produce a secondary field when exposed to a primary field, EMI can also be used to detect anomalous features that are more resistive than the background.

In general, there are two methods by which depth information can be extracted from EMI measurements. These are geometrical soundings and broadband soundings. In geometrical soundings, measurements are made at a fixed frequency for a range of separation distances between the transmitting and receiving coils while, for broadband soundings, decreasing the frequency is associating with deeper probing. This study deals only with broadband soundings. In a conducting medium, electromagnetic waves will both oscillate and decay with distance from the source and both the wavenumber and the decay rate are proportional to the frequency. Figure 2 depicts two conducting targets at two different depths below a transmitting coil generating a primary field with conductor 1 being the shallower target. Since this primary field decays with distance, conductor 1 is immersed in a stronger field than conductor 2 and, hence, a stronger current and, consequently, a stronger secondary field is induced in this conductor. Furthermore, there will be greater decay of the secondary field from conductor 2 since it is farther away from a receiving coil taken

to be located alongside the transmitting coil. The receiving coil will measure a secondary field that is a composite of the response of both conductors with this measured response being a depth-integrated function of the conductivity inversely weighted by the depth.

Figure 3 is a sequence of illustrations designed to show how broadband soundings can be used to quantify target depth. In fig. 3a, the driving frequency is sufficiently high that the primary field decays to a negligibly small value before it reaches the depth of the conductive anomaly. Decreasing the frequency (fig. 3b), decreases the rate of decay of the primary field such that a measurable response to the anomaly is produced at the receiver coil. When the frequency is so low that the primary field penetrated to a depth much greater than the conducting target (fig. 3c), the anomaly is only weakly manifested in the measured secondary because, as noted above, this field is depth average and the anomaly occupies only a small portion of the penetration depth. Thus, decreasing the frequency below that shown in fig. 3b will produce a gradual decay in the measured secondary field. Figure 3d shows an idealized magnitude of the measured secondary field as a function of frequency illustrating the broadband sounding concept.

A measure of the penetration depth of a primary field is the skin depth, δ , defined to be

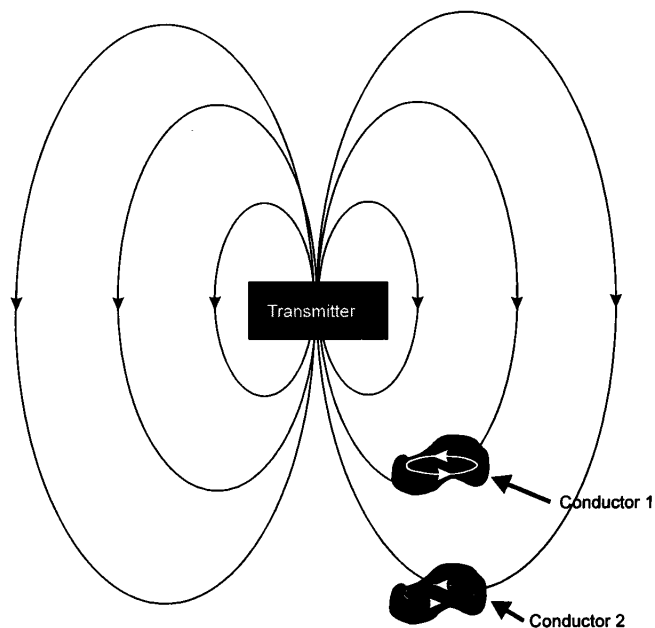


Figure 2. Illustration showing two conductors at different depths below a transmitting coil. Since conductor 2 is further away from the transmitting coil than conductor 1, a weaker current (indicated by the narrower white line depicting the current loop) is induced in this feature.

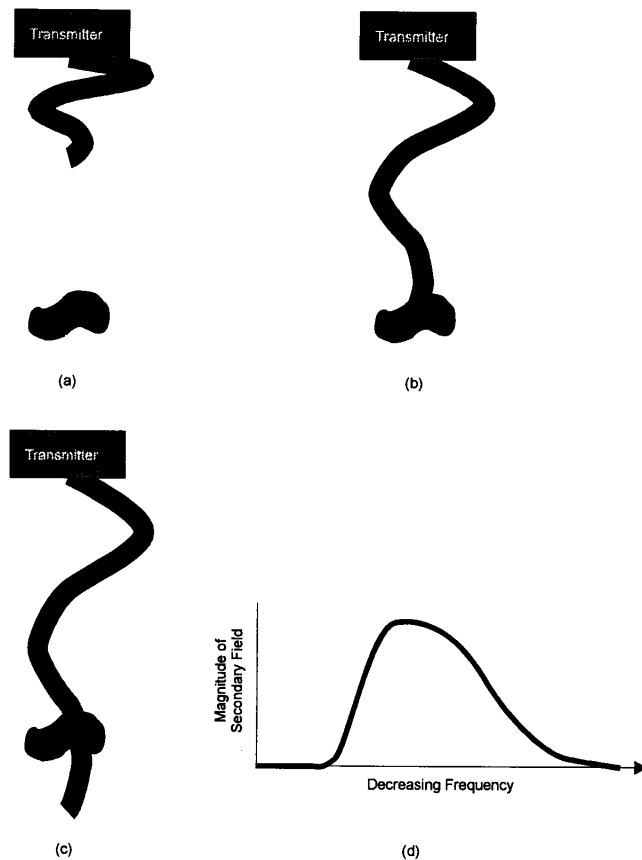


Figure 3. Illustration showing the oscillation and decay of a time-varying electromagnetic field and a conductor below the transmitting coil for a) a high frequency, b) a moderate frequency, and c) a low frequency. The variation in secondary field strength as a function of frequency is shown in d).

$$\delta = \frac{\sqrt{2}}{\sqrt{\omega \mu_0 \sigma_0}} \quad (1)$$

where ω is the frequency, μ_0 is the background magnetic susceptibility, and σ_0 is the background conductivity. Won (1980) demonstrated that broadband soundings such as that shown in fig. 3d will peak at a skin depth corresponding to the target depth. Figure 4 is simulated data for measurements made by traversing the ground surface with a co-located transmitting/receiving coil pair making measurements every meter for a range of frequencies corresponding to uniformly-spaced skin depths between 1 and 32 meters. The target assumed here is a rectangular conductor dimensioned 12 horizontally by 8 vertically centered at (15.5, -16.5). This figure displays the in-phase component of the output voltage from the secondary coil as gray levels as a function of measurement location and skin depth. It is clear that the data, so displayed, is an out-of-focus image of the rectangular target with the target appearing at a skin depth near the target depth. The phase changes appearing in the

pseudo-image are a result of the fact that both the primary and secondary field are spatially oscillatory. The next section will describe tomographic imaging concepts whereby out-of-focus images, such as that shown in fig. 4 can be sharpened and the phase changes removed.

Broadband EMI Imaging

One means to describe the EMI imaging concept is through the related problem of imaging using seismic reflection and specifically migration. For a co-located seismic source and receiver deployed over the plane $\mathbf{r} = (x, y)$, the spatially and temporally Fourier transformed received signal

$$\tilde{u}(\kappa, \omega) = \int dt e^{i\omega t} \int d\mathbf{r} e^{-i\kappa \cdot \mathbf{r}} u(\mathbf{r}, t) \quad (2)$$

is related to the three-dimensional reflectivity by

$$R(x, y, z) = \frac{1}{8\pi^3} \int d\omega \int d\kappa \exp\{i(\sqrt{4k^2 - \kappa^2} z + \kappa \cdot \mathbf{r})\} \tilde{u}(\kappa, \omega) \quad (3)$$

(Cheng and Coen, 1984), where $k = \omega/c_0$, c_0 is the background wave speed and κ is the magnitude of the two-dimensional vector κ . Spatial variations in reflectivity are recon-

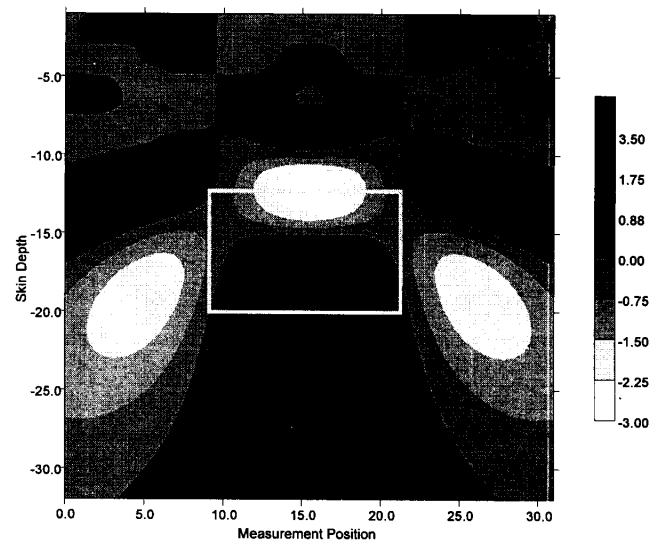


Figure 4. Simulated broadband data for measurements made at uniformly spaced points on the ground surface over the top of 12 by 8 m rectangular conductor buried at an assumed depth of -16.5 m. The horizontal axis is the measurement location and the vertical axis is skin depth. In this form, the data appears as an out-of-focus image. The gray rectangular outline displays the size, shape, and depth of the target.

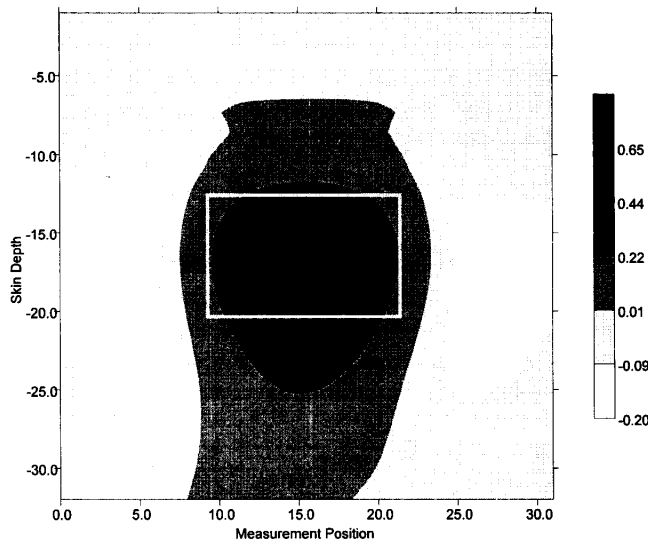


Figure 5. Vertical cross-sectional image reconstructed from the simulated broadband data shown in Fig. 4. The outline of the actual assumed target is outlined in gray.

structed in Eq. (3) by backpropagating (migrating) the data u downward to some plane $z < 0$ and evaluating this data at time zero. In a similar approach within the framework of diffraction

tomography, the spatial variations of the object function, $O(x, y, z) = 1 - c_0^2/c^2(x, y, z)$, are reconstructed by

$$O(x, y, z) = \frac{1}{8\pi^3} \int d\omega \int d\kappa H(\kappa, \omega) \exp\{i(\sqrt{4k^2 - \kappa^2}z + \kappa \cdot \mathbf{r})\} \tilde{u}(\kappa, \omega) \quad (4)$$

where $c(x, y, z)$ are the spatial variations in wave speed about c_0 and H is measurement geometry-specific filter function. The three-dimensional spatial Fourier transform of Eq. (4) is

$$\tilde{O}(\kappa, \sqrt{4k^2 - \kappa^2}) = H(\kappa, \omega) \tilde{u}(\kappa, \omega) \quad (5)$$

and it is clear that the object function can be reconstructed by applying an inverse Fourier transform to Eq. (5). Equations relating the spatial Fourier transform of data to the spatial Fourier transform of the object function, such as Eq. (5), are known as generalized projection slice theorems (Devaney, 1984).

A generalized projection slice theorem for broadband EMI imaging (Witten, Won, and Norton, 1997) relating the spatially Fourier transformed output voltage from the receiving coil, $V(x, y)$, to an integral transform of the object function, is

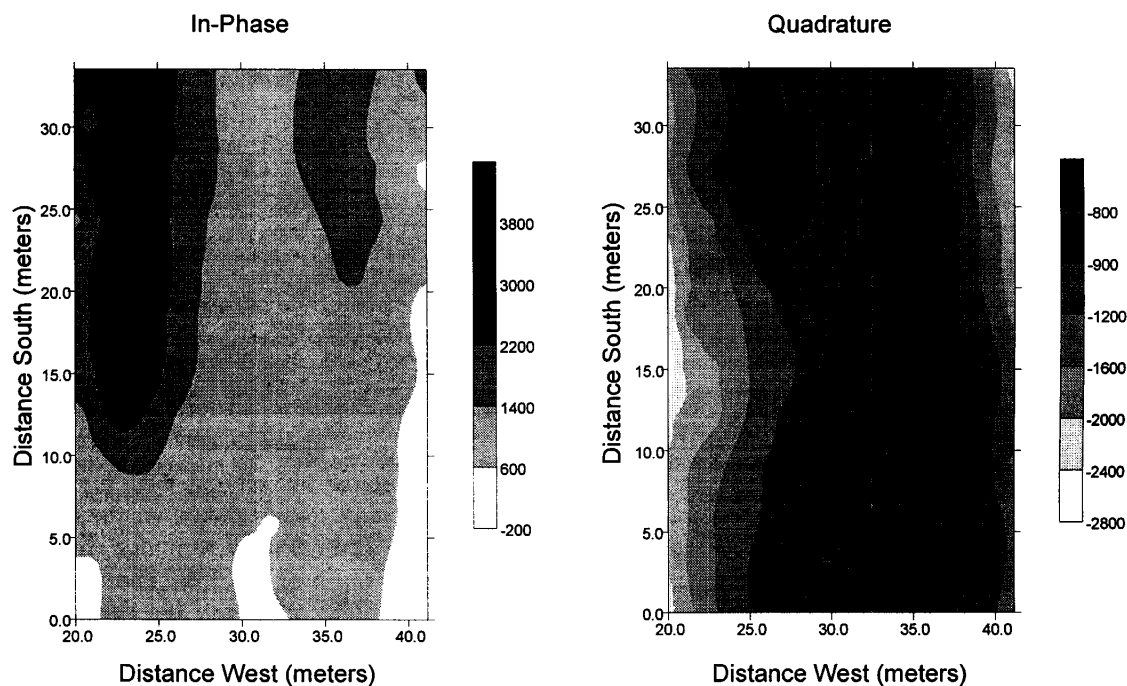


Figure 6. In-phase and Quadrature data at 9810 Hz as measured at the Anacostia tunnel site.

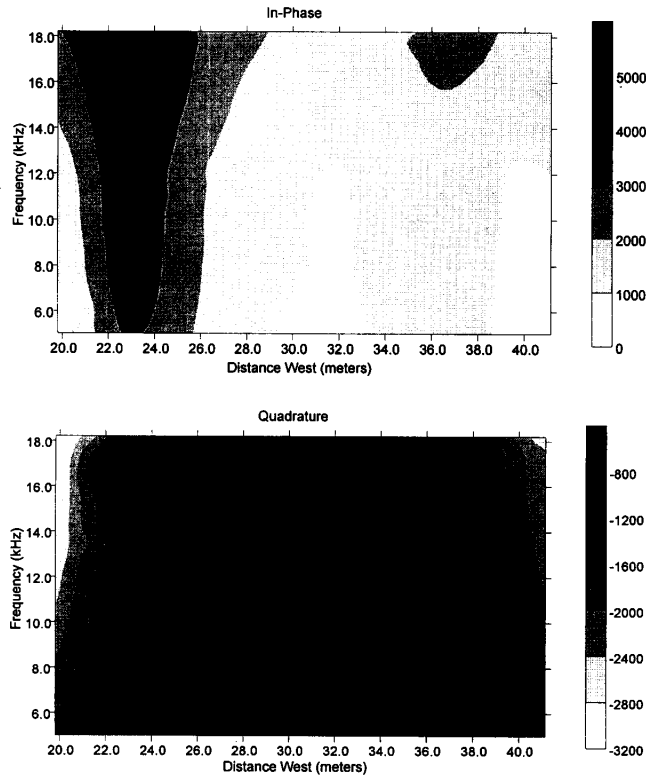


Figure 7. Pseudo image of the Anacostia data presented as measurement location along a line at 21 m north as a function of frequency (the vertical axis).

$$\tilde{O}(\kappa, \sqrt{(32i\pi^2 / \delta^2) - \kappa^2}) = H(\kappa, \omega) \tilde{V}(\kappa, \omega), \quad (6)$$

where $O(x, y, z) = (\sigma(x, y, z) / \sigma_o) - 1$ and $\sigma(x, y, z)$ are the spatial variations in conductivity relative to a background σ_o . The most important difference between this relationship and that for wave-based imaging, Eq. (5), is that the vertical component of the wave vector in Eq. (5) is real while that for EMI imaging, Eq. (6), is complex and cannot be completely inverted by Fourier transform. The inversion of Eq. (6) can be expressed in the form

$$O(x, y, z) = \frac{1}{2\pi^2} \int d\kappa e^{i\kappa \cdot r} \bar{O}(\kappa, z), \quad (7)$$

where the object function is reconstructed by performing an inverse two-dimensional (horizontal) Fourier transform of \bar{O} and \bar{O} is the inverse of the integral transform

$$\begin{aligned} \tilde{O}(\kappa, \sqrt{(32i\pi^2 / \delta^2) - \kappa^2}) = \\ \int dz H(\kappa, \omega) \exp\{-i\sqrt{(32i\pi^2 / \delta^2) - \kappa^2} z\} \tilde{V}(\kappa, \omega). \end{aligned} \quad (8)$$

Equation (8) is a mapping from depth into skin depth and consequently its inverse is a mapping from skin into depth. Since this integral transform is characterized by a complex wave number, it cannot be inverted by Fourier transform. In the algorithm developed here, Eq. (8) is inverted by matrix inversion, however, this procedure is ill-conditioned necessitating the use of a regularization procedure.

Figure 5 is an image resulting from the application of this reconstruction procedure to the broadband data presented in fig. 4. It is clear that this imaging algorithm serves to focus the pseudo-image and removes phase changes evident in fig. 4.

Case Studies

The broadband EMI imaging algorithm has been applied to data acquired at two underground structure sites. At both sites, data were acquired with the GEM-2 sensor (Won, Keiswiter, Fields, and Sutton, 1996). This broadband tool consists of a transmitting and receiving coil located at opposite ends of a boom that is approximately two meters long. A third coil, the "bucking" coil, is used to nullify the primary field at the receiving coil so that this coil responds only to the much weaker secondary field. The GEM-2 can operate at frequencies up to 22kHz.

There are two notable differences between GEM-2's operation and configuration and assumptions made in the imaging algorithm. First, the imaging algorithm assumes that the transmitting and receiving coils are co-located while these coils are separated in the GEM-2. This inconsistency can potentially produce errors in the reconstructed images where

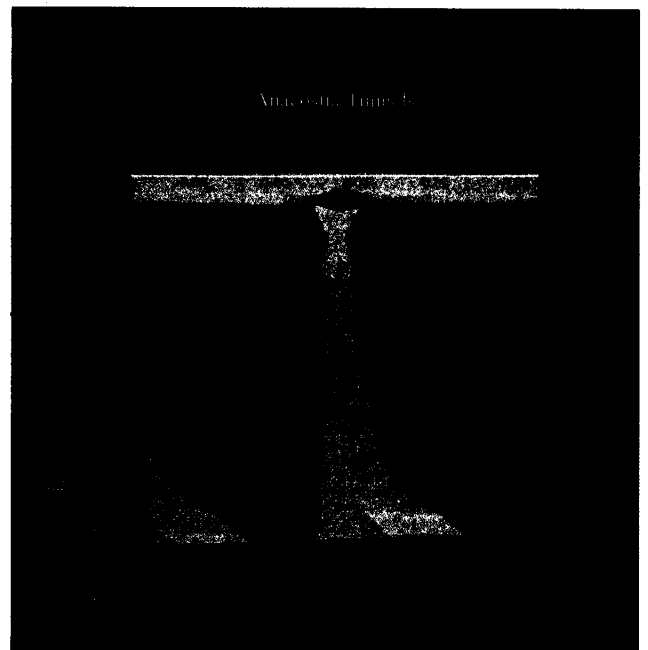


Figure 8: Three-dimensional rendering of the reconstructed image of the subway tunnels at the Anacostia site.

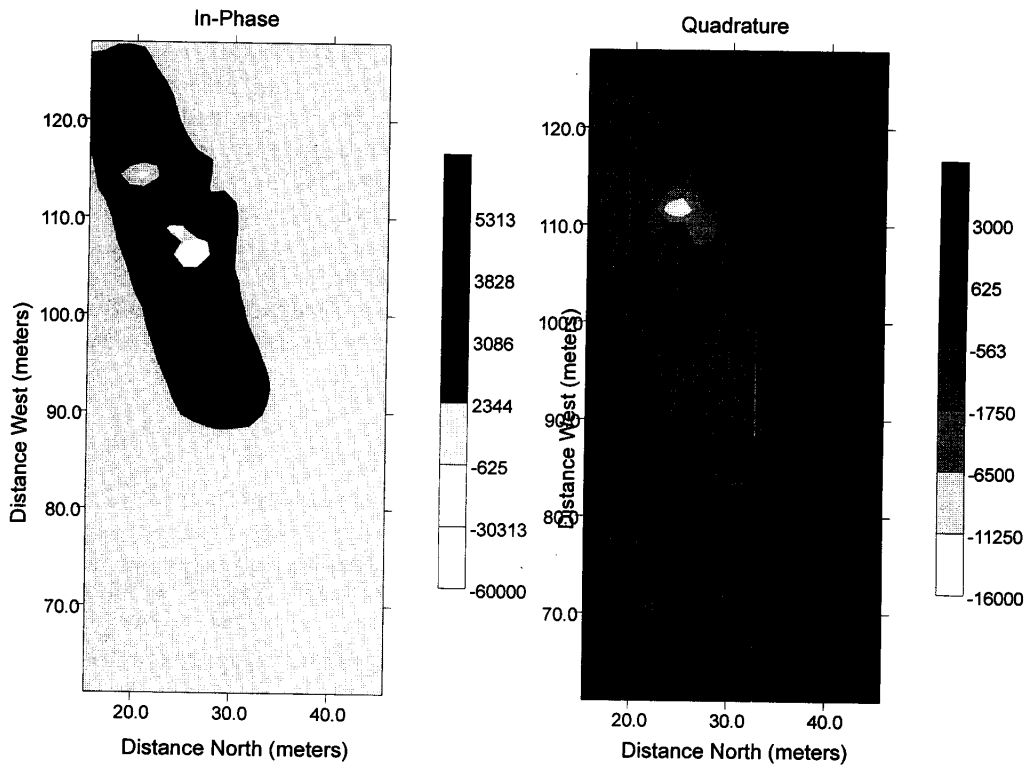


Figure 9: In-phase and Quadrature data at 9270 Hz as measured at the cloud chamber site.

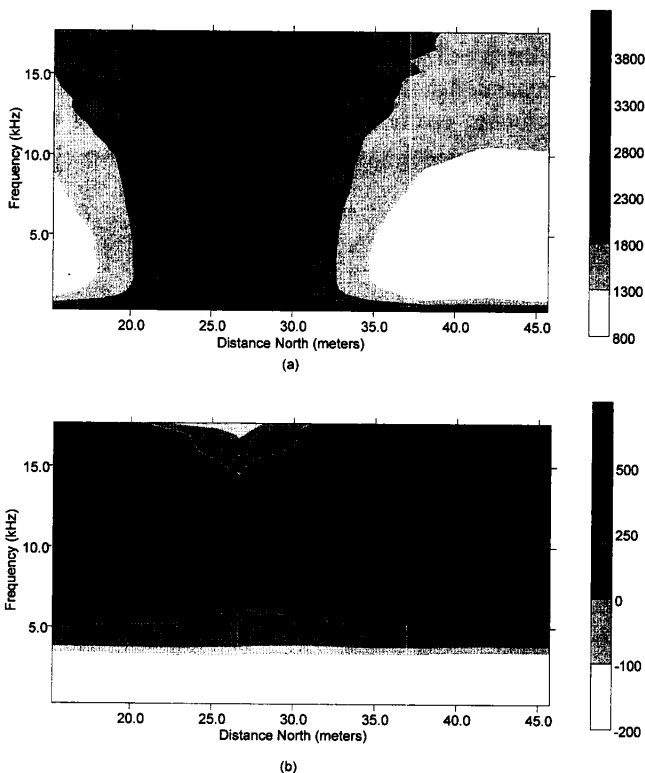


Figure 10: Pseudo image of the cloud chamber data presented as measurement location along a line at 99 m west as a function of frequency (the vertical axis).

the magnitude of such errors will depend on the target depth, the coil separation, and frequencies used in the reconstruction. For the parameters appropriate to the case studies presented here, errors associated with the coil separation can be shown to be negligible. The imaging algorithm described in the previous section is based on output voltage at the receiving coil associated with secondary field. The instrument-specific parameters that influence this voltage are the diameter, number of turns, and orientation of both the transmitting and receiving coils as well as the amperage and frequency applied to the transmitting coil. These parameters appear explicitly in the filter H appearing in Eq. (8). The GEM-2 output is the ratio of the secondary to primary field at the receiving coil in parts per million which is linearly related to receiving coil output voltage. Because of the difference between the actual output from the GEM-2 and that assumed in the imaging algorithm, absolute conductivities cannot be reconstructed, however, reconstructed values of the object function are expected to be correct in a relative sense.

The Anacostia Site

The Anacostia site is vacant lot about 152 m from the Washington, D.C. Anacostia metro station. GEM-2 broadband EMI data were acquired over an area approximately 44 m by 30 m. Part of this area was marshy with standing water varying in depth up to about 0.35 m. The target of interest at this site is a pair of parallel subway tunnels having a center-to-center

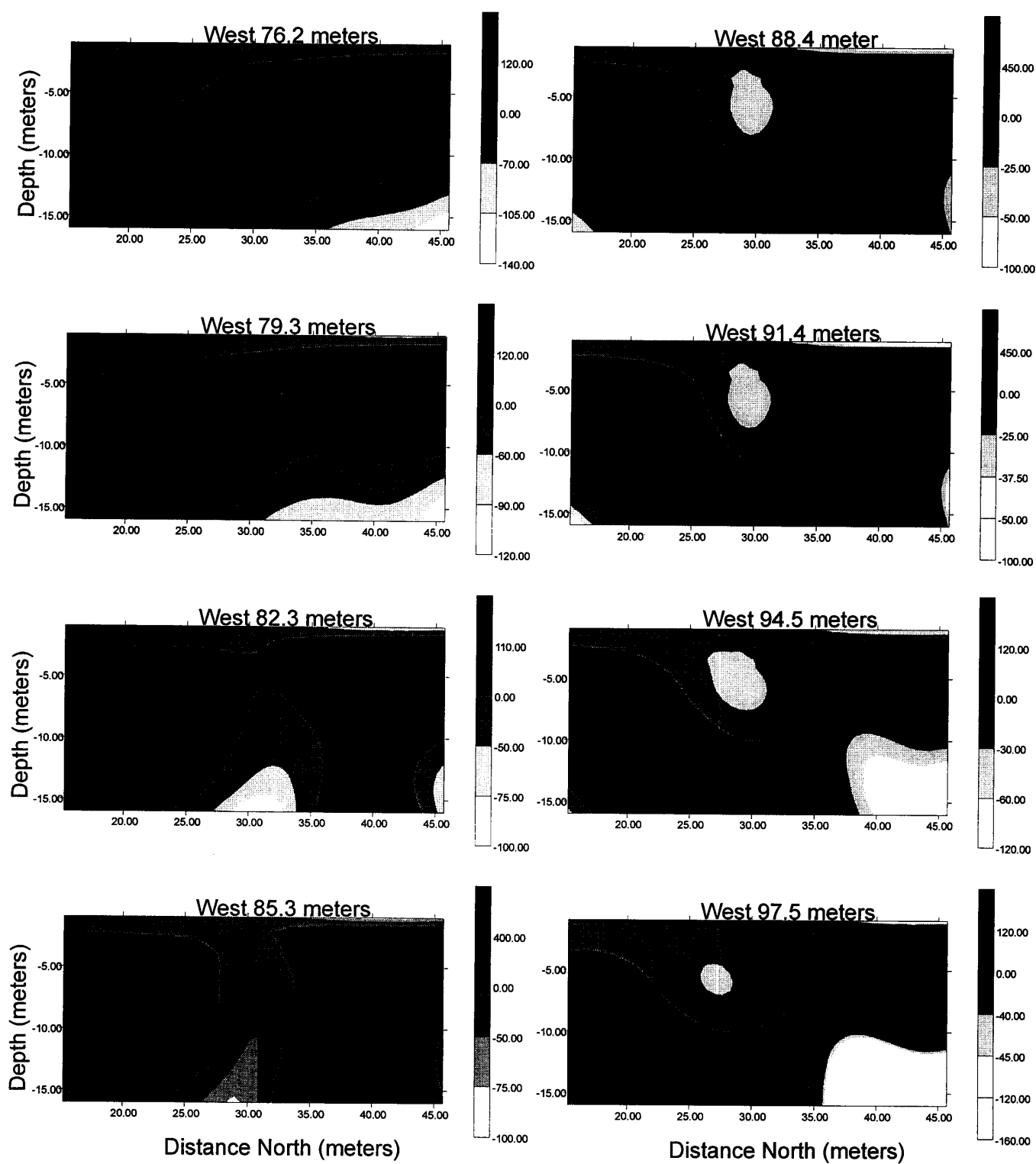


Figure 11a. Three-dimensional image of the cloud chamber displayed as 16 uniformly spaced vertical cross-sections at 76.2 through 97.5 m west.

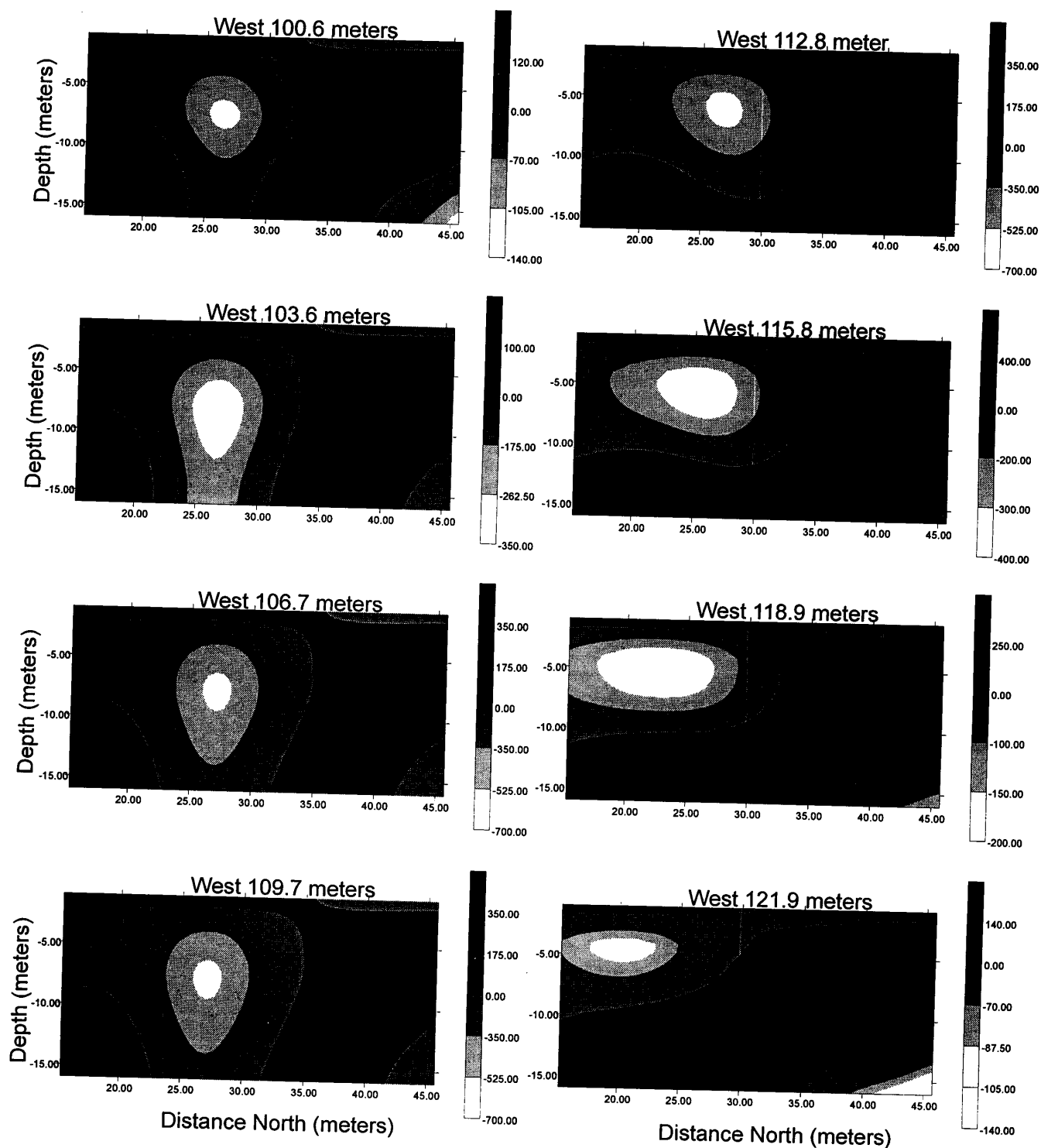


Figure 11b. Three-dimensional image of the cloud chamber displayed as 16 uniformly spaced vertical cross-sections at 100.6 through 121.9 m west.

separation of about 12 m. The axes of these tunnels are oriented in a north-south direction. Each tunnel is approximately 4.5 m in diameter and the tunnel ceilings are believed to be 10 to 13 m below the ground surface. GEM-2 measurements were made at 28 points spaced at 1.524 m in the east-west direction and at 11 points spaced at 3.048 m in the north-south direction. At every measurement point data were acquired with a vertical coil axes orientation for 23 discrete frequencies ranging from 5010 Hz to 18210 Hz in 600 Hz increments.

Figure 6 shows the in-phase and quadrature measurements at 9810 Hz from the Anacostia site. Over the northern half of the survey area both tunnels are manifested in the in-phase data with the western tunnel producing a stronger anomaly. In the southern half, however, there is little evidence of the either tunnel presumably because of an increase in tunnel depth or a change in background conductivity. The existence of the tunnel pair is evident over the entire north-south extent of the quadrature data, however, in this case the individual tunnels are indistinct. This general pattern is typical of data acquired at all frequencies. Figure 7 is a pseudo cross-sectional image of the in-phase component of a slice occurring at 21 m north similar to the synthetic data shown in fig. 4. Here the vertical axis is decreasing frequency and it is expected that there should be a localized anomaly occurring at a frequency associated with a skin depth, Eq. (1), corresponding to the target depth. A localized anomaly does occur for the eastern tunnel at a frequency of about 17 kHz, however, there is no relative high or low for the western tunnel.

Figure 8 is a three-dimensional rendering of the image reconstructed from the GEM-2 data using the imaging algorithm described in the previous section. Sixteen frequencies ranging from 5010 Hz to 14010 Hz were used to compute the object function at 16 depths ranging from 1 to 23.5 m and an assumed background conductivity of 0.1 Siemen/m. In this image, rendered with EarthVision software¹, the lowest values of object function (those associated with the lowest relative conductivity) were removed to give the subway tunnels a hollow appearance. In this image, the tunnels have the correct separation and horizontal diameter. The tunnels are vertically elongated in the image resulting in a vertical dimension that is slightly too large and a depth to tunnel ceilings that is slightly too small, however, the depth to the tunnel centers is approximately correct.

The Cloud Chamber Site

The cloud chamber site exists on the Department of Energy's Nevada Test Site in Nye County, Nevada. As its name implies, the target of interest at this site is cloud chamber original built to measure certain properties of ionized particles. The cloud chamber has not been used since 1968. This underground structure is shaped like a quonset hut, 42.7 m long, 9.75 m wide at its base, with a maximum height of 4.9 m. The depth to the roof of the cloud chamber is about 5 m.

GEM-2 measurements were over a rectangular region approximately 30 m north-south by 67 m east-west. Data were acquired at uniformly spaced intervals of 1.524 m in both directions. Thirty frequencies ranging from 270 Hz to 17670 Hz in intervals of 600 Hz were used for both vertical and horizontal coil axis orientation. Figure 9 shows the acquired in-phase and quadrature data at 9270 Hz for the vertical coil axis orientation. The cloud chamber is clearly manifested in both components of this data as the rectangular anomaly extending diagonally from about 30 m north, 100 m west to about 20 m north, 120 m west. Smaller anomalous features evident in the data are associated with vents and an access shaft. The data presented in this figure is typically of the entire frequency range. Figure 10 displays the entire acquired vertical coil axis frequency band along a line at 99 m west as a pseudo cross-sectional image. As in fig. 7, the vertical axis is decreasing frequency (increasing skin depth). The lateral location of the cloud chamber is clearly resolved, however, the highest frequencies used were too low to allow the resolution of the top of this feature.

A subset of the broadband, vertical coil axis EMI data was input to the above-described imaging algorithm. Here, 16 uniformly-spaced frequencies ranging from 8670 Hz to 17670 Hz were used. The full range of measurement points in the north-south direction were used but data in the east-west direction were limited to a sub-region extending from 76 m to 122 m west at 3.048 m intervals. Figure 11 displays the three-dimensional image reconstructed at a background conductivity of 0.004 Siemen/m as 16 parallel vertical cross-sections. Although the contrast scale changes from cross-section to cross-section, the cloud chamber appears in these images as a relative conductivity low in approximately the proper horizontal location and over the correct approximate depth interval. The width of the cloud chamber is about 50% too small in many of the cross-sections.

Conclusions

Underground structures at two sites have been imaged with broadband EMI data. Based on this experience and in the absence of any horizontally-broad, near-surface high conductivity anomalies, it is possible to image other underground structures of similar size and depth. Furthermore, it is anticipated that deeper underground structures can similarly be imaged since the fundamental limiting parameter in EMI imaging is the penetration depth. At least with the GEM-2 EMI tool, it is possible to acquire data down to the DC level suggesting that great penetration depths can be achieved allowing the imaging of much deeper structures, albeit with an attendant loss of spatial resolution.

The ability to image shallow targets within the frequency range considered here is somewhat speculative. This is because the lack of high frequency information will, in turn, limit mapping between target depth and skin depth to the point

where the inversion of Eq. (8) is no longer robust. Although the bandwidth used here yielded skins depths that were, at best, no less than the target, it is unclear at this time how much shallower the imaging algorithm can be stressed.

References

- Cheng, G. and Coen, S., 1984, The relationship between Born inversion and migration for common-midpoint stacked data, *Geophysics*, **49**, 2117-2131.
- Devaney, A.J., 1984, Geophysical diffraction tomography, *IEEE Trans. Geosci. And Remote Sensing*, **GE-22**, 3-13.
- Witten, A.J., Won, I.J., and Norton, S.J., 1997, Subsurface imaging with broadband electromagnetic induction, submitted to *Inverse Problems*.
- Won, I.J., 1980, A wide-band electromagnetic induction method: some theoretical and experimental results, *Geophysics*, **45**, 928-940.
- Won, I.J., Keiswiter, D.A., Fields, G.R.A., and Sutton, L.C., 1996, A new multifrequency electromagnetic sensor, *J. Env. and Eng. Geophysics*, **1**, 129-137.

¹ EarthVision is a registered trademark of Dynamic Graphic, Inc.

Estimation of object location from short pulse scatter data

George A. Tsihrintzis^a, Anthony J. Devaney^a, and Ehud Heyman^b

^aCenter for Electromagnetics Research

Department of Electrical and Computer Engineering

Northeastern University

Boston, MA 02115, USA.

^bDepartment of Electrical Engineering and Physical Electronics

Tel-Aviv University

Tel-Aviv, 69978, ISRAEL.

ABSTRACT

We present an efficient algorithm for computation of the maximum likelihood estimate of the location of a known target from short pulse scatter data measured in a suite of tomographic experiments. The algorithm consists of a three step procedure: (i) data filtering, (ii) time-domain backpropagation, and (iii) coherent summation and employs of a number of forward and inverse Radon transforms integrated in a tomographic scheme. A computer simulation is included for illustration purposes.

Keywords: Inverse Scattering, Tomography, Filtered Backpropagation, Radon Transform.

1. INTRODUCTION

Inverse Scattering (IS) is the scientific discipline in which an object (scatterer) is probed with waves in an attempt to estimate (reconstruct) its internal structure from scattered field measurements. The scattering object is assumed embedded in a known, non-attenuating background medium and the objective of IS is to quantitatively estimate the spatial distribution of its complex-valued refraction index by inverting the mathematical mapping relating the probing wave, the refraction index, and the measurable total wave [1, 2, 13, 21, 3, 4]. This objective is non-trivial to achieve due to the non-uniqueness of the mapping in any finite set of experiments and its non-linear and non-local character [6, 15, 12]. The non-uniqueness issue can be addressed by employing a multiplicity of experiments, whereby the object is probed from several incident wave directions with pulses of broad bandwidth. However, the issue of non-linearity (and non-locality) is significantly harder to address and, to date, research has only produced mathematical results or

computationally intensive iterative algorithms as opposed to practically implementable inversion algorithms.

Over the past fifteen years, an alternative approach to the IS problem has been followed based on certain linearizing approximations to it [7, 9, 10, 8, 11, 13] (see also [14] for a review of the state-of-the-art in linearized scattering models). This approach has led to an expanded discipline within the regime of tomography, known as *Diffraction Tomography (DT)*, which has reached today the stage of building prototype commercial tomographic scanners for biomedical [17, 18, 29, 28, 19, 27, 30, 32] and underground [35, 37, 25, 36, 34, 33] imaging systems based on the algorithms of the linearized version of the IS problem. Despite the success of the linearized algorithms in several practical applications, their success depends critically on the two assumptions of linearity and availability of multiple experiments. In many cases, the linearity assumption fails, while different constraints (economic, safety, operating, geometric, or physical) [26, p.21] limit the number of scattering experiments that can be performed and/or provide low signal-to-noise ratio data. As a result, practical scatterer reconstruction algorithms often suffer from high noise levels, poor resolution, and artifacts, such as streaking and geometric distortion.

To overcome these limitations, a more modest inverse problem was addressed, originally within the framework of linearized [16] and later exact [31] scattering theory. This inverse problem attempted to *detect* and *classify* a *known* target scatterer and *estimate* its *location* from wave scatter data. It was found that, for monochromatic plane-wave probing, the optimum (in the maximum likelihood sense) location estimate could be obtained via a filtered backpropagation algorithm, in which partial images formed by filtering and backpropagating scatter data for different probing directions were coherently summed. The algorithm is optimum (in the maximum likelihood sense) for additive Gaussian noise and an arbitrary number of scattering experiments and can yield an estimate of all three coordinates of the target even from a single experiment as long as the wavelength of the probing radiation is comparable with the typical dimensions of the target [16].

The work in [31] can be readily extended to the case of probing with multiple frequencies via frequency-domain synthesis. However, a derivation of a time-domain algorithm may be computationally more efficient in the case of probing with short pulses. A time-domain algorithm was recently presented for the classical IS reconstruction problem of DT for cases where the usual frequency-domain algorithms were not computationally efficient [23]. In this paper, we are concerned with the problem of optimally processing (limited and noisy) short pulse scatter data to decide whether a certain target scatterer is present in a region of interest and, if so, determine its location. The problem arises in a number of inverse scattering applications, such as locating subsurface targets in seismic and electromagnetic surveys, detecting tumors in ultrasonic medical imaging, and detecting and identifying radar targets. The main goal of this paper is to extend the single frequency maximum likelihood DT algorithms derived in [16, 31] to the time domain in a similar manner as was done in [23] for the classical DT reconstruction problem.

The paper is organized as follows: Section 2 contains a review of the data measurement configuration, the wave scattering equations, and frequency-domain and time-domain plane-wave spectra of waves. Section 3 is concerned with location estimation algorithms and, including a short review of the frequency-domain solution, derivation of a time-domain filtered backpropagation algorithm. Computer simulations are included in Section 4, while Section 5 contains a summary, a statement of conclusions, and a list of possible research avenues to be followed in the future.

computationally intensive iterative algorithms as opposed to practically implementable inversion algorithms.

Over the past fifteen years, an alternative approach to the IS problem has been followed based on certain linearizing approximations to it [7, 9, 10, 8, 11, 13] (see also [14] for a review of the state-of-the-art in linearized scattering models). This approach has led to an expanded discipline within the regime of tomography, known as *Diffraction Tomography (DT)*, which has reached today the stage of building prototype commercial tomographic scanners for biomedical [17, 18, 29, 28, 19, 27, 30, 32] and underground [35, 37, 25, 36, 34, 33] imaging systems based on the algorithms of the linearized version of the IS problem. Despite the success of the linearized algorithms in several practical applications, their success depends critically on the two assumptions of linearity and availability of multiple experiments. In many cases, the linearity assumption fails, while different constraints (economic, safety, operating, geometric, or physical) [26, p.21] limit the number of scattering experiments that can be performed and/or provide low signal-to-noise ratio data. As a result, practical scatterer reconstruction algorithms often suffer from high noise levels, poor resolution, and artifacts, such as streaking and geometric distortion.

To overcome these limitations, a more modest inverse problem was addressed, originally within the framework of linearized [16] and later exact [31] scattering theory. This inverse problem attempted to *detect* and *classify* a *known* target scatterer and *estimate* its *location* from wave scatter data. It was found that, for monochromatic plane-wave probing, the optimum (in the maximum likelihood sense) location estimate could be obtained via a filtered backpropagation algorithm, in which partial images formed by filtering and backpropagating scatter data for different probing directions were coherently summed. The algorithm is optimum (in the maximum likelihood sense) for additive Gaussian noise and an arbitrary number of scattering experiments and can yield an estimate of all three coordinates of the target even from a single experiment as long as the wavelength of the probing radiation is comparable with the typical dimensions of the target [16].

The work in [31] can be readily extended to the case of probing with multiple frequencies via frequency-domain synthesis. However, a derivation of a time-domain algorithm may be computationally more efficient in the case of probing with short pulses. A time-domain algorithm was recently presented for the classical IS reconstruction problem of DT for cases where the usual frequency-domain algorithms were not computationally efficient [23]. In this paper, we are concerned with the problem of optimally processing (limited and noisy) short pulse scatter data to decide whether a certain target scatterer is present in a region of interest and, if so, determine its location. The problem arises in a number of inverse scattering applications, such as locating subsurface targets in seismic and electromagnetic surveys, detecting tumors in ultrasonic medical imaging, and detecting and identifying radar targets. The main goal of this paper is to extend the single frequency maximum likelihood DT algorithms derived in [16, 31] to the time domain in a similar manner as was done in [23] for the classical DT reconstruction problem.

The paper is organized as follows: Section 2 contains a review of the data measurement configuration, the wave scattering equations, and frequency-domain and time-domain plane-wave spectra of waves. Section 3 is concerned with location estimation algorithms and, including a short review of the frequency-domain solution, derivation of a time-domain filtered backpropagation algorithm. Computer simulations are included in Section 4, while Section 5 contains a summary, a statement of conclusions, and a list of possible research avenues to be followed in the future.

The mapping between the scattered wave ψ^s and the object function O is non-linear and non-local, since the scattered wave appears at both sides of Eq.(2-6). Linearizing approximations of the Born type approximate Eq.(2-6) by replacing the total field in the right hand side has been replaced with the incident field component. A related approximation is the Rytov approximation which has been found more accurate than the Born approximation in several applications [20, 29, 14]. In the present work, we will not make any approximations to Eqs.(2-2) through (2-6) and proceed with exact scattering models.

Our starting point is the “source translation property,” originally derived in [31]. To state the source translation property, consider an object function of the form:

$$O(\mathbf{r}; \mathbf{R}_c) = O_0(\mathbf{r} - \mathbf{R}_c), \quad (2-7)$$

i.e., an object function O_0 that has been shifted in space by a vector \mathbf{R}_c . The induced source corresponding to illumination with a certain incident wave $\hat{\psi}_0(\mathbf{r}, \omega)$ can be expressed as

$$\begin{aligned} \hat{\rho}(\mathbf{r}', \omega; \mathbf{R}_c) &\equiv O(\mathbf{r}'; \mathbf{R}_c) \hat{\psi}(\mathbf{r}', \omega; \mathbf{R}_c) \\ &= e^{ik\mathbf{s}_0 \cdot \mathbf{R}_c} \hat{\rho}(\mathbf{r}' - \mathbf{R}_c, \omega; 0), \end{aligned} \quad (2-8)$$

where \mathbf{s}_0 is the unit vector in the direction of incidence of the illuminating wave and $\hat{\psi}(\mathbf{r}', \omega; \mathbf{R}_c)$ and $\hat{\rho}(\mathbf{r}', \omega; \mathbf{R}_c)$ denote the field and induced source resulting from the object function $O(\mathbf{r}; \mathbf{R}_c)$. Eq.(2-8), which is established in [31], relates the source induced by a plane-wave incident on the scattering object $O_0(\mathbf{r} - \mathbf{R}_c)$ to the source induced by the same illuminating wave but incident on the (centered) object $O_0(\mathbf{r})$. Eqs.(2-6) and (2-8) allow us to express the field scattered by the object $O_0(\mathbf{r} - \mathbf{R}_c)$ to the field scattered by the (centered) object $O_0(\mathbf{r})$. We find that [31]

$$\hat{\psi}^s(\mathbf{r}, \omega; \mathbf{R}_c) = e^{ik\mathbf{s}_0 \cdot \mathbf{R}_c} \hat{\psi}^s(\mathbf{r} - \mathbf{R}_c, \omega; 0). \quad (2-9)$$

Eqs.(2-8) and (2-9) lead to a frequency-domain filtered backpropagation algorithm for the object location estimation problem within exact scattering theory [31]. They will be employed here to derive the time-domain counterpart of that algorithm. For completeness, however, we present their time-domain version, obtained via inverse Fourier transforming Eqs.(2-8) and (2-9). The result is:

$$\rho(\mathbf{r}, t; \mathbf{R}_c) = \rho(\mathbf{r} - \mathbf{R}_c, t - \mathbf{s}_0 \cdot \mathbf{R}_c / c_0; 0) \quad (2-10)$$

$$\psi^s(\mathbf{r}, t; \mathbf{R}_c) = \psi^s(\mathbf{r} - \mathbf{R}_c, t - \mathbf{s}_0 \cdot \mathbf{R}_c / c_0; 0). \quad (2-11)$$

B. Time-Domain Plane-Wave Spectra of Wavefields

Consider a time-domain wavefield $\psi(\mathbf{r}, t)$ propagating into the half-space $z > 0$ and having the boundary value $\psi(\mathbf{x}, t)$ ($\mathbf{x} = x\hat{x} + y\hat{y}$), on the $z = 0$ plane, with equivalent frequency-domain representation

$$\hat{\psi}(\mathbf{x}, \omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} \psi(\mathbf{x}, t). \quad (2-12)$$

The frequency-domain plane-wave spectrum of the wavefield is defined as [24]

$$\hat{\hat{\psi}}(\xi, \omega) = \int d^2x e^{-ik\xi \cdot \mathbf{x}} \hat{\psi}(\mathbf{x}, \omega), \quad (2-13)$$

where $k = \frac{\omega}{c_0}$ and ξ is a frequency-independent wave vector [24]. From Eq.(2-13), one can obtain the frequency-domain representation of the wavefield at an arbitrary point $\mathbf{r} = \mathbf{x} + z\hat{\mathbf{z}}$ in space as

$$\hat{\psi}(\mathbf{r}, \omega) = \left(\frac{k}{2\pi}\right)^2 \int d^2\xi e^{ik(\xi \cdot \mathbf{x} + \zeta z)} \hat{\tilde{\psi}}(\xi, \omega), \quad (2-14)$$

where

$$\zeta = \begin{cases} \sqrt{1 - |\xi|^2}, & \text{if } |\xi| \leq 1 \\ i\sqrt{|\xi|^2 - 1}, & \text{else.} \end{cases}$$

Eq.(2-14) is the, so-called, angular spectrum expansion of $\hat{\psi}(\mathbf{r}, \omega)$ and decomposes the wavefield $\hat{\psi}(\mathbf{r}, \omega)$ into a superposition of propagating (corresponding to $|\xi| \leq 1$) and evanescent (corresponding to $|\xi| > 1$) plane waves.

Eq.(2-13) has a time-domain equivalent, the time-domain plane-wave spectrum of the wavefield:

$$\tilde{\psi}(\xi, \tau) = \frac{1}{2\pi} \int d\omega e^{-i\omega\tau} \hat{\tilde{\psi}}(\xi, \omega). \quad (2-15)$$

Substitution of Eqs.(2-12) and (2-13) into Eq.(2-15) gives

$$\tilde{\psi}(\xi, \tau) = \int d^2\mathbf{x} \psi(\mathbf{x}, \tau + \frac{\xi \cdot \mathbf{x}}{c_0}), \quad (2-16)$$

as the time-domain relation between the wavefield and its plane-wave spectrum. Eq.(2-16) is recognized as a Radon transform [5] of the wavefield $\psi(\mathbf{x}, t)$ in the three-dimensional space (\mathbf{x}, t) and has, thus, been termed a slant-stack transform [23]. Eq.(2-16) can be inverted to give the time-domain equivalent of Eq.(2-14). The inversion formula that includes the evanescent modes requires use of the analytic signal (see [22] for details). Here we will assume that the evanescent modes have been sufficiently attenuated to not contribute to the inversion formula. With this in mind, the result is

$$\psi(\mathbf{r}, t) = -\frac{1}{(2\pi c_0)^2} \int_{|\xi| \leq 1} d^2\xi \frac{\partial^2}{\partial \tau^2} \tilde{\psi}(\xi, \tau = t - \frac{\xi \cdot \mathbf{x} + \zeta z}{c_0}), \quad (2-17)$$

and can be recognized as a bank of inverse Radon transforms, each corresponding to a different z [5].

3. ESTIMATION OF OBJECT LOCATION

A. Data Model

With the definitions and results of Section 2, we are now in position to proceed with the object location estimation problem. Consider a scatterer described by the object function:

$$O(\mathbf{r}; \mathbf{R}_c) = O_0(\mathbf{r} - \mathbf{R}_c), \quad (3-1)$$

where O_0 is a known function and \mathbf{R}_c is an unknown scatterer location. The object is probed with short plane-wave pulses propagating in the direction of unit vectors \mathbf{s}_0 , i.e., plane-wave pulses of the form $p_{\mathbf{s}_0}(t - \frac{\mathbf{s}_0 \cdot \mathbf{r}}{c_0})^1$. We assume

¹The subscript \mathbf{s}_0 is used to indicate that the form of the pulses may, in general, be different from experiment to experiment.

that scattered pulse data are measured over planes perpendicular to the direction \mathbf{s}_0 of propagation of the incident pulses. More specifically, we assume the measurement (data) model

$$y(\mathbf{r}_p, t, \mathbf{s}_0) = r_{s_0}(\mathbf{r}_p, t) \otimes \psi_{s_0}^s(\mathbf{r}_p + l\mathbf{s}_0, t; \mathbf{R}_c) + n_{s_0}(\mathbf{r}_p, t); \quad \mathbf{r}_p \in \mathbf{R}^2, -\infty < t < \infty, \mathbf{s}_0 \text{ in some set of unit vectors,} \quad (3-2)$$

where $r_{s_0}(\mathbf{r}_p, t)$ is a convolutional space-time measurement filter and $\psi_{s_0}^s(\mathbf{r}_p + l\mathbf{s}_0, t; \mathbf{R}_c)$ is the scattered field over the plane $\mathbf{r} = \mathbf{r}_p + l\mathbf{s}_0$. Additionally, $n_{s_0}(\mathbf{r}_p, t)$ is zero-mean Gaussian noise, white² in the variables \mathbf{r}_p , t , and \mathbf{s}_0 , i.e.,

$$\mathcal{E}\{n_{s_0}(\mathbf{r}_p, t)n_{s'_0}(\mathbf{r}'_p, t')\} = \sigma_n^2 \delta(\mathbf{r}_p - \mathbf{r}'_p) \delta(t - t') \delta_{s_0, s'_0}. \quad (3-3)$$

The inverse problem is that of estimating the unknown parameter (object location) \mathbf{R}_c from the measurements $y(\mathbf{r}_p, t, \mathbf{s}_0)$ in Eq.(3-2).

B. Likelihood Function

We obtain an estimate $\hat{\mathbf{R}}_c$ of the unknown \mathbf{R}_c by minimizing the following data functional with respect to a test object location \mathbf{r}_c :

$$\mathcal{J}(\mathbf{r}_c) = \sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} dt \int d^2 r_p |y(\mathbf{r}_p, t, \mathbf{s}_0) - \alpha(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c)|^2, \quad (3-4)$$

where

$$\alpha(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c) = r_{s_0}(\mathbf{r}_p, t) \otimes \psi_{s_0}^s(\mathbf{r}_p + l\mathbf{s}_0, t; \mathbf{r}_c) \quad (3-5)$$

is the scattered pulse on the measurement plane (filtered by the measurement filter) for the object located at \mathbf{r}_c . The estimate $\hat{\mathbf{R}}_c$ returned by the above minimization is also the maximum likelihood estimate (MLE) under the Gaussianity assumption for the measurement noise in Eq.(3-2) [?].

To obtain the MLE $\hat{\mathbf{R}}_c$ directly from Eq.(3-4) is an intensive process, since numerical minimization is required of the functional $\mathcal{J}(\mathbf{r}_c)$. However, expansion of the square term in Eq.(3-4) shows that the MLE $\hat{\mathbf{R}}_c$ is equivalently obtained by maximization of the "log likelihood function" [?]

$$L(\mathbf{r}_c) = \sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} dt \int d^2 r_p y(\mathbf{r}_p, t, \mathbf{s}_0) \alpha(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c) - \frac{1}{2} \sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} dt \int d^2 r_p |\alpha(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c)|^2. \quad (3-6)$$

Next, we will further simplify the expression for the log likelihood function and obtain an algorithm that can be efficiently implemented on the computer.

Theorem 1 *If evanescent plane-wave spectra are ignored, the term $\frac{1}{2} \sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} dt \int d^2 r_p |\alpha(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c)|^2$ in the log likelihood function in Eq.(3-6) is constant with respect to \mathbf{r}_c .*

Proof: From Parseval's theorem:

$$\int_{-\infty}^{\infty} dt \int d^2 r_p |\alpha(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c)|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \int d^2 r_p |\hat{\alpha}(\mathbf{r}_p, \omega, \mathbf{s}_0; \mathbf{r}_c)|^2.$$

We apply now the fact [31] that, for each temporal frequency ω , $\int d^2 r_p |\hat{\alpha}(\mathbf{r}_p, \omega, \mathbf{s}_0; \mathbf{r}_c)|^2$ is constant with respect to \mathbf{r}_c , as long as the corresponding evanescent spectra are ignored, and the theorem follows. •

²Gaussian noise of arbitrary color can be handled by expanding the algorithm of this section to include a proper whitening filter.

Theorem 2 Ignoring evanescent plane-wave spectra, the term $\sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} dt \int d^2 r_p y(\mathbf{r}_p, t, \mathbf{s}_0) \alpha(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c)$ in the log likelihood function in Eq.(3-6) is equal to

$$\sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} dt \int d^2 r_p y(\mathbf{r}_p, t, \mathbf{s}_0) \alpha(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c) = -\frac{1}{(2\pi c_0)^2} \sum_{\mathbf{s}_0} \int_{|\xi| \leq 1} d^2 \xi \frac{\partial^2}{\partial \tau^2} [\tilde{y}(\xi, \tau, \mathbf{s}_0) \bar{\alpha}(\xi, \tau, \mathbf{s}_0; 0)]|_{\tau = -\frac{\xi \cdot \mathbf{r}_{cp} + \xi \cdot \mathbf{s}_0 \cdot \mathbf{r}_c}{c_0}}, \quad (3-7)$$

where $\mathbf{r}_c = \mathbf{r}_{cp} + \mathbf{s}_0 \cdot \mathbf{r}_c$.

Proof: We start again with Parseval's theorem:

$$\begin{aligned} \int_{-\infty}^{\infty} dt \int d^2 r_p y(\mathbf{r}_p, t, \mathbf{s}_0) \alpha(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \int d^2 r_p \hat{y}(\mathbf{r}_p, \omega, \mathbf{s}_0) \bar{\alpha}(\mathbf{r}_p, \omega, \mathbf{s}_0; \mathbf{r}_p) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \int d^2 r_p \hat{y}(\mathbf{r}_p, \omega, \mathbf{s}_0) \bar{\alpha}(\mathbf{r}_p - \mathbf{r}_c, \omega, \mathbf{s}_0; 0) e^{-i \frac{\omega}{c_0} \mathbf{s}_0 \cdot \mathbf{r}_c}, \end{aligned}$$

where we used the translation property in Eq.(2-9). We now apply Eq.(2-14) and obtain

$$\begin{aligned} \int_{-\infty}^{\infty} dt \int d^2 r_p \hat{y}(\mathbf{r}_p, \omega, \mathbf{s}_0) \bar{\alpha}(\mathbf{r}_p, \omega, \mathbf{s}_0; \mathbf{r}_c) &= \\ \frac{1}{(2\pi)^3 c_0^2} \int_{-\infty}^{\infty} d\omega \omega^2 \int d^2 r_p \hat{y}(\mathbf{r}_p, \omega, \mathbf{s}_0) \int_{|\xi| \leq 1} d^2 \xi e^{-i \frac{\omega}{c_0} \xi \cdot \mathbf{r}_p} e^{i \frac{\omega}{c_0} [\xi \cdot \mathbf{r}_{cp} + (\xi - 1) \mathbf{s}_0 \cdot \mathbf{r}_c]} \bar{\alpha}(\xi, \omega, \mathbf{s}_0; 0) &= \\ \frac{1}{(2\pi)^3 c_0^2} \int_{|\xi| \leq 1} d^2 \xi \int_{-\infty}^{\infty} d\omega e^{i \frac{\omega}{c_0} [\xi \cdot \mathbf{r}_{cp} + (\xi - 1) \mathbf{s}_0 \cdot \mathbf{r}_c]} \omega^2 \bar{\alpha}(\xi, \omega, \mathbf{s}_0; 0) \int d^2 r_p e^{-i \frac{\omega}{c_0} \xi \cdot \mathbf{r}_p} \hat{y}(\mathbf{r}_p, \omega, \mathbf{s}_0) &= \\ \frac{1}{(2\pi)^3 c_0^2} \int_{|\xi| \leq 1} d^2 \xi \int_{-\infty}^{\infty} d\omega e^{-i \frac{\omega}{c_0} [\xi \cdot \mathbf{r}_{cp} + (\xi - 1) \mathbf{s}_0 \cdot \mathbf{r}_c]} [\omega^2 \bar{\alpha}(\xi, \omega, \mathbf{s}_0; 0) \hat{y}(\xi, \omega, \mathbf{s}_0) e^{i \omega \tau}]|_{\tau=0} &= \\ -\frac{1}{(2\pi c_0)^2} \int_{|\xi| \leq 1} d^2 \xi \frac{\partial^2}{\partial \tau^2} [\tilde{y}(\xi, \tau, \mathbf{s}_0) \bar{\alpha}(\xi, \tau, \mathbf{s}_0; 0)]|_{\tau = -\frac{\xi \cdot \mathbf{r}_{cp} + (\xi - 1) \mathbf{s}_0 \cdot \mathbf{r}_c}{c_0}} \end{aligned}$$

from which the theorem follows.

Eq.(3-7) can be interpreted as follows: For each pulse, the scattered pulse data are Radon-transformed with respect to their space-time coordinates,³ filtered in Radon space, and inverse Radon-transformed (one inverse Radon transform per value of $\mathbf{s}_0 \cdot \mathbf{r}_c$) into object space to form partial images of the log likelihood function. The Radon-space filter consists of the complex conjugate of the time-domain plane-wave spectra of the field scattered by the centered object $O_0(\mathbf{r})$. Finally, the partial images are coherently superimposed.

4. COMPUTER ILLUSTRATION

For simulation purposes, we consider a single scattering experiment in a two-dimensional geometry in which a target lies in the (x, z) -plane and is infinitely long and uniform along the y -axis. The probing pulse is incident from the direction of the positive z -axis and data are measured along the line $(x, z = l)$. The target signature (scattered pulse $\psi_0^s(\mathbf{r}, t)$ corresponding to the target located at the origin) is a pulse

$$\psi_0^s(\mathbf{r}, t) = \frac{\sigma^2}{c_0^2} \frac{\ddot{p}(t - \frac{|\mathbf{r}|}{c_0})}{\sqrt{|\mathbf{r}|}}, \quad \mathbf{r} = (x, z), \quad (4-1)$$

³Equivalently, the scatter data are measured in the far field of the scattering object.

where σ^2 is the scatterer cross section,

$$p(t) = \begin{cases} \frac{1}{T} \cos(\beta t^2), & \text{if } 0 < t < T \\ 0, & \text{else} \end{cases} \quad (4-2)$$

and \ddot{p} denotes second derivative, i.e.

$$\ddot{p}(t) = \begin{cases} \frac{2\beta}{T} [\cos(\beta t^2)[1 + 2\beta t^2] - 1], & \text{if } 0 < t < T \\ 0, & \text{else.} \end{cases} \quad (4-3)$$

Eq.(4-1) gives the far field of the Born approximation to the field scattered by a point scatterer of cross section σ^2 when probed with the plane-wave pulse corresponding to $p(t)$. The scattered field for arbitrarily located target is subsequently generated according to Eq.(2-11).

We assumed the scatterer to be located at the origin of the (x, z) -plane and chose the parameters $\sigma^2 = 1$, $c_0 = 1$, $\beta = 1$, $T = 1$, and $l = 5$. For the additive Gaussian noise, we examined three different cases with corresponding variances 0, 0.25, and 1. Figs. 2 show the field at $(x = 0, z = 5)$ for the time interval $0 < t < 16$ for three different noise levels. Finally, the likelihood functions computed via Eq.(3-7) are shown in Figs. 3 for the three noise levels.

5. SUMMARY, CONCLUSIONS, AND FUTURE WORK

In this paper, we derived an efficient algorithm for computation of the log likelihood function for estimation of the location of a target object from noisy short pulse scatter measurements. We assumed that the target object signature, i.e., the wave scattered by the object centered at the origin of the coordinate system was known and established that the log likelihood function could be computed in a three step procedure in which the measured data were firstly convolutionally filtered, then time-domain backpropagated, and finally the partial likelihood functions were coherently summed. The convolutional filter of the first step was the filter matched to the object signature and the procedure was implemented via a sequence of direct and inverse Radon transforms of the space-time measurements. Target identification can also be performed via a similar algorithm, in which a bank is employed of filters matched to various target signatures.

A computer simulation of a single scattering experiment was performed to illustrate the procedure, in which rather than the target object, a target signature was assumed. The simulation revealed very high performance even in the case of very low signal-to-noise ratio. Further performance improvements can be achieved if multiple scattering experiments are utilized.

Further research in this area seems appropriate. Issues to be addressed in the future include the derivation of proper location estimation algorithms for the cases of measurement planes that remain fixed from scattering experiment to experiment. This is the case in geophysical surveys in which the sensor array is fixed in space and several scattering experiments are performed, each with a different probing plane-wave pulse. Another avenue of future research seems to lead to the derivation of nonparametric algorithms for detection, location estimation, and classification of stochastic scattering objects. This problem appears in several underwater surveys. This and related research is currently pursued and its results will be announced shortly.

ACKNOWLEDGEMENTS

This work was supported in part by ARO Grant DAAG55-97-1-0013 and by AFOSR Grant F49620-96-1-0028.

References

- [1] Z. S. Agranovich and V. A. Marchenko. *The Inverse Problem of Scattering Theory*. Gordon and Breach, New York, 1963.
- [2] A. P. Anderson and M. F. Adams. Holographic and tomographic imaging with microwaves and ultrasound. In Wolfgang-M. Boerner, editor, *Inverse Methods in Electromagnetic Imaging, Part 2*. D. Reidel Publishing Company, Boston, 1983.
- [3] M. Cheney. A review of multidimensional inverse potential scattering. In D. Colton, R. Ewing, and W. Rundell, editors, *Inverse Problems in Partial Differential Equations*. SIAM, Philadelphia, 1990.
- [4] D. Colton and P. Monk. The inverse scattering problem for acoustic waves in an inhomogeneous medium. In D. Colton, R. Ewing, and W. Rundell, editors, *Inverse Problems in Partial Differential Equations*. SIAM, Philadelphia, 1990.
- [5] S. Deans. *The Radon Transform and Some of its Applications*. Wiley, New York, 1983.
- [6] A. J. Devaney. Nonuniqueness in the inverse scattering problem. *J. Math. Phys.*, 19:1526, 1978.
- [7] A. J. Devaney. Inverse scattering within the Rytov approximation. *Optics Letters*, 6:374, 1981.
- [8] A. J. Devaney. A computer simulation study of diffraction tomography. *IEEE Trans. Bio. Eng.*, BME-30:377, 1982.
- [9] A. J. Devaney. A filtered backpropagation algorithm for diffraction tomography. *Ultrasonic Imaging*, 4:336, 1982.
- [10] A. J. Devaney. Inversion formula for inverse scattering within the Born approximation. *Optics Letters*, 7:111, 1982.
- [11] A. J. Devaney. Geophysical diffraction tomography. *IEEE Trans. Geosci. and Remote Sensing*, GE-22:3, 1984.
- [12] A. J. Devaney. Fundamental limitations in inverse source and scattering problems in nondestructive evaluation. In D. O. Thompson and D. E. Chimenti, editors, *Review of Progress in Quantitative Nondestructive Evaluation*. Plenum, New York, 1986.
- [13] A. J. Devaney. Reconstructive tomography with diffracting wavefields. *Inverse Problems*, 2:161, 1986.
- [14] A. J. Devaney. Approximate scattering models in inverse scattering: Past, present, and future. In L. B. Felsen and H. Bertoni, editors, *Directions in Electromagnetic Modeling*. Plenum Press, New York, 1990.
- [15] A. J. Devaney and G. C. Sherman. Nonuniqueness in inverse source and scattering problems. *IEEE Trans. Ant. and Propag.*, AP-30:1034, 1982.
- [16] A. J. Devaney and G. A. Tsihrintzis. Maximum likelihood estimation of object location in diffraction tomography. *IEEE Trans. Signal Processing*, 39(3):672, 1991.
- [17] B. Duchene, D. Lesselier, and W. Tabbara. Experimental investigation of a diffraction tomography technique in fluid ultrasonics. *IEEE Trans. Bio. Eng.*, BME-35:437, 1988.
- [18] W. Tabbara *et al.* Diffraction tomography: Contribution to the analysis of some applications in microwaves and ultrasonics. *Inverse Problems*, 4:305, 1988.
- [19] L.-J. Gelius, I. Johansen, N. Sponheim, and J. J. Stamnes. A generalized diffraction tomography algorithm. *J. Acoust. Soc. Am.*, 89(2):523, 1991.
- [20] J. B. Keller. Accuracy and validity of the Born and Rytov approximation. *J. Opt. Soc. Am.*, 59:1003, 1969.

- [21] K J Langenberg. Wave modeling for inverse problems with acoustic, electromagnetic, and elastic waves. In H L Bertoni and L B Felsen, editors, *Directions in Electromagnetics Wave Modeling*. Plenum Press, New York, 1990.
- [22] T. Melamed. Phase-space beam summation: A local spectrum analysis of time-dependent radiation. *J. EM Waves Appl.*, 11:739-773, 1997.
- [23] T. Melamed, Y. Ehrlich, and E. Heyman. Short-pulse inversion of inhomogeneous media: a time-domain diffraction tomography. *Inverse Problems*, 12:977-993, 1996.
- [24] P. M. Morse and H. Feshbach. *Methods of Theoretical Physics, Parts I,II*. McGraw-Hill, New York, 1953.
- [25] R. B. Pratt and M. H. Worthington. The application of diffraction tomography to cross-hole seismic data. *Geophysics*, 53:1284, 1988.
- [26] D. J. Rossi and A. S. Willsky. Reconstruction from projections based on detection and estimation of objects, Parts I and II: Performance analysis and robustness analysis. *IEEE Trans. Acous., Speech, and Signal Process.*, ASSP-32:886, 1984.
- [27] N. Sponheim, L.-J. Gelius, I. Johansen, and J. J. Stamnes. Quantitative results in ultrasonic tomography of large objects using line sources and curved detector arrays. *IEEE Trans. Ultra., Ferroelec., and Freq. Con.*, UFFC-38:370, 1991.
- [28] N. Sponheim and I. Johansen. Experimental results in ultrasonic tomography using a filtered backpropagation algorithm. *Ultrasonic Imaging*, 13:56, 1991.
- [29] N. Sponheim, I. Johansen, and A. J. Devaney. Initial testing of a clinical ultrasound mammograph. In Hua Lee and Glen Wade, editors, *Acoustical Imaging, Vol. 18*. Plenum Press, New York, 1990.
- [30] J. J. Stamnes, L.-J. Gelius, I. Johansen, and N. Sponheim. Diffraction tomography applications in seismics and medicine. In *NATO advanced workshop on inverse problems in scattering and imaging*, North Falmouth, Massachusetts, 1991.
- [31] G. A. Tsihrintzis and A. J. Devaney. Maximum likelihood estimation of object location in diffraction tomography, Part II: Strongly scattering objects. *IEEE Trans. Signal Processing*, 39(6):1466, 1991.
- [32] G. A. Tsihrintzis and A. J. Devaney. Application of a maximum likelihood estimator in an experimental study of ultrasonic diffraction tomography. *IEEE Trans. Med. Imag.*, MI-12:545-554, 1993.
- [33] A. Witten. Application of a maximum likelihood estimator to tunnel detection. *Inverse Problems*, 7:L49-L55, 1991.
- [34] A. Witten, D. Gillette, W. C. King, and J. Sypniewski. Geophysical diffraction tomography at a dinosaur site. (submitted).
- [35] A. Witten and E. Long. Shallow applications of geophysical diffraction tomography. *IEEE Trans. Geosci. and Remote Sensing*, GE-24:654, 1986.
- [36] A. Witten and J. Tuggle. A practical approach to ultrasonic imaging using diffraction tomography. *J. Acoust. Soc. Am.*, 83:1645, 1988.
- [37] R. Wu and M. N. Toksoz. Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, 52(1):11, 1987.

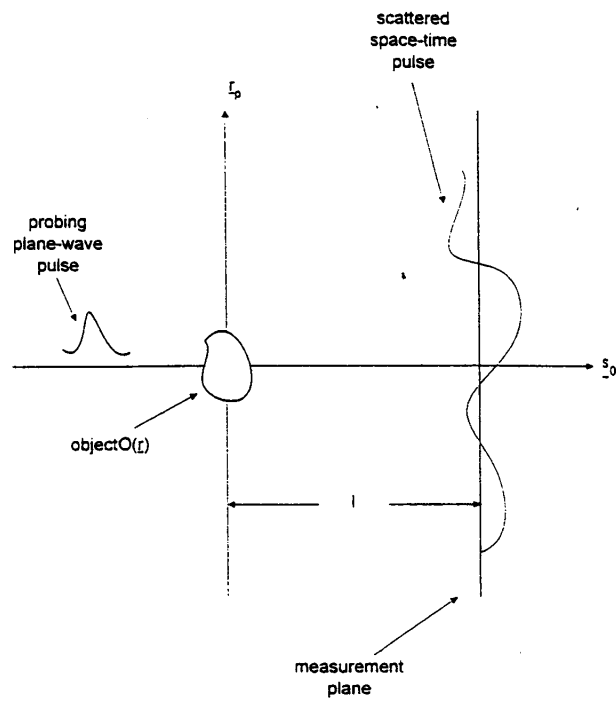


FIGURE 1: Experimental configuration of scattering experiments

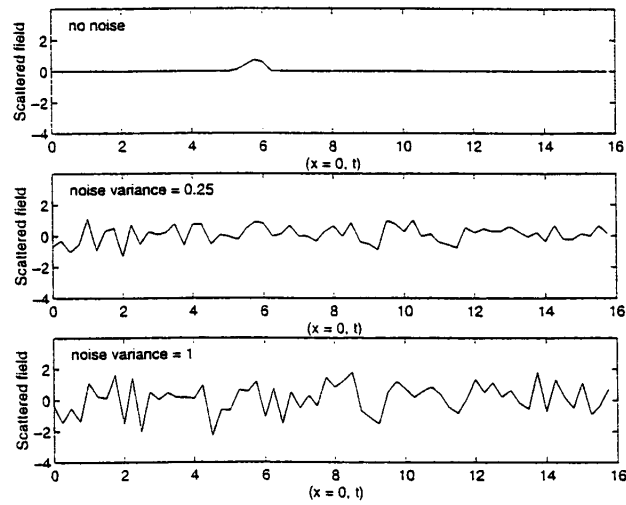


FIGURE 2: Measured scattered field at $(x=0, z=5)$ for t in $[0, 16]$

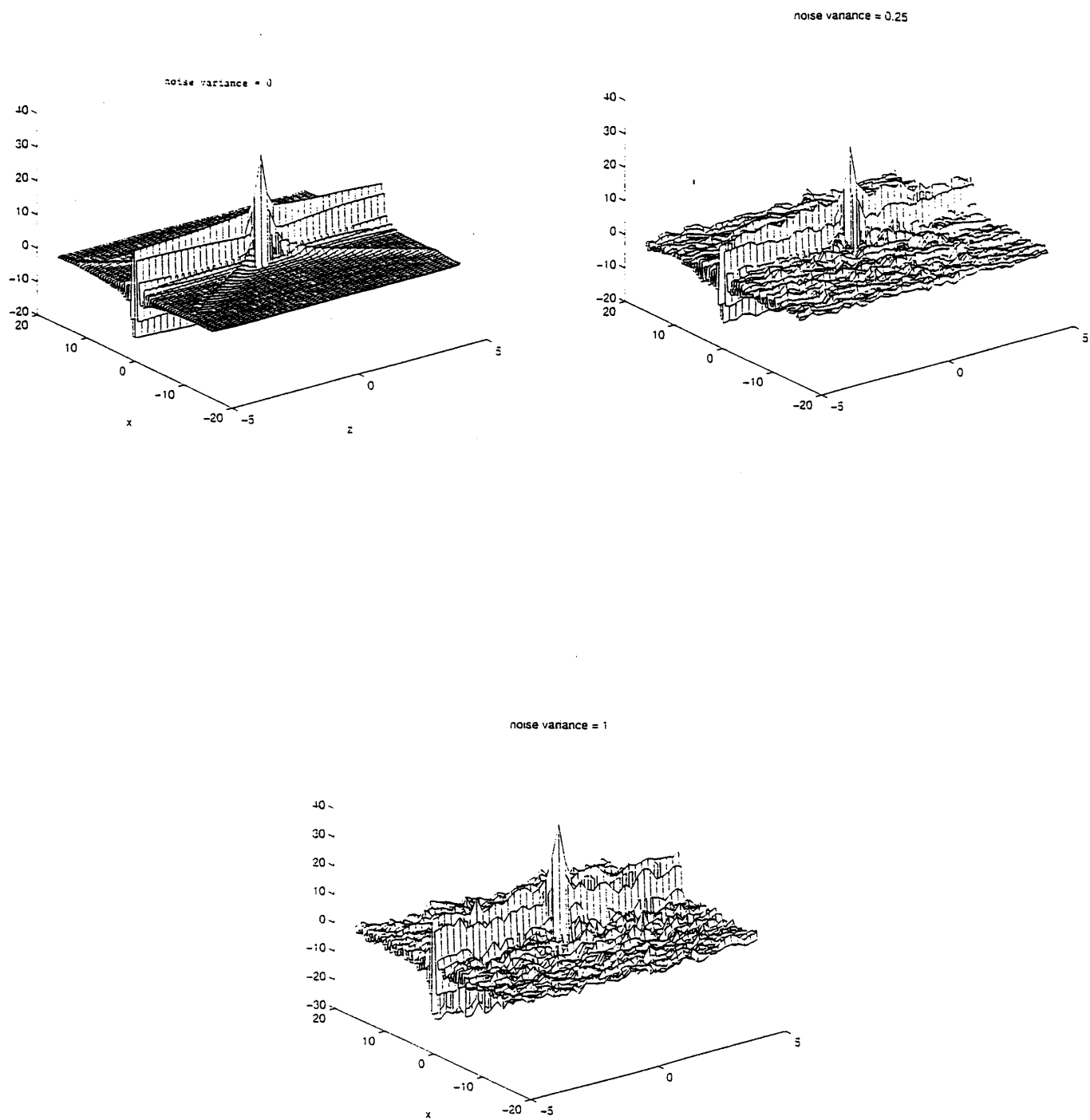


FIGURE 3: Log likelihood function for estimation of object location

Hyperspectral IR polarimetry with applications in demining and unexploded ordnance detection

Herman E. Scott, Stephen H. Jones, Frank Iannarilli, and Kurt Annen
Aerodyne Research, Inc., 45 Manning Road, Billerica, MA 01821-3976¹

ABSTRACT

Several years of effort in IR polarimetry have brought us convincing evidence of its effectiveness in differentiating man made objects from natural backgrounds. Adding modern focal plane array (FPA) technology (either cooled or uncooled) makes it possible to combine the benefits of polarimetry with the power of hyperspectral imaging. Aerodyne Research is embarked on a stepwise, controlled-risk development program with the objective of fielding an innovative and affordable hyperspectral imaging IR polarimeter. Proof-of-concept demonstrations are conducted for each significant technology increment as part of the prototype development effort. These steps, two demonstrated and two yet to be demonstrated, are: (1) LWIR (non-imaging) Spectral Polarimeter to demonstrate the effectiveness of combined polarimetric and hyperspectral discrimination capabilities in observations on static scenes; (2) LWIR Uncooled FPA Imaging (broadband) Polarimeter to test the sensitivity of an affordable Uncooled FPA in a broadband configuration against static scenes; (3) Multispectral Imaging Polarimeter to quantify clutter rejection performance improvements to be realized in multispectral polarimetry; and (4) Hyperspectral Imaging IR Polarimeter designed with optimal spatial and spectral resolution and sufficient throughput to achieve the reliable performance required in surface mine and UXO detection applications. Results from the ongoing proof-of-concept demonstrations in simulated surface mine detection will be presented.

Keywords: Infrared polarimetry, hyperspectral imaging, demining, unexploded ordnance

1. INTRODUCTION

1.1 Background

For more than 15 years Aerodyne Research has worked systematically toward the development of passive IR spectral and polarimetric discrimination techniques. In the first 10 years of these activities Aerodyne focussed on the development and validation of a first principles computer model called POLAR¹⁻³ that computes the spectrally dependent Stokes vector and user-selected in-band IR polarimetric images of targets given the geometry, optical properties⁴, and thermal description of the target. Support for the POLAR model came from Northrop⁵, the NADC/Warminster⁶⁻⁹, and internal R&D. Then in recent years, emphasis has turned to the systems applications of polar attributes such as non-cooperative target identification^{9,10} and the discrimination of targets in highly cluttered natural backgrounds.

In 1993 - 95 several important steps were accomplished: Aerodyne and the Boeing Defense and Space Group with the support of NAWC/Warminster used the modified IRAMMP sensor, POLIRAMMP, to collect polar IR imagery of aircraft on the ground to validate the POLAR model.¹¹ Results showed a strong correlation between the analytic model predictions and measured data. Subsequently in 1994 Aerodyne produced a seminal paper¹⁰ on a polarimeter sensor concept (3-Channel Simultaneous Acquisition Polarimeter) specifically designed for the long range Counter-Air Target Identification application. Shortly thereafter, Boeing and Aerodyne applied the Improved POLIRAMMP sensor in the 1994 Navy Long Jump Tests at China Lake to collect surface-to-air polarization imagery of aircraft and thereby further validate the POLAR model.¹² It was at this test site that we conducted our first primitive simulation of surface mine detection using the two-channel POLIRAMMP sensor. Even though this test¹³ simulated only surface mine detection, it provided quite convincing evidence of the clutter suppression capability of IR polarimetry.

In 1997 under a Phase I SBIR for the Air Force Research Laboratory (AFRL), Aerodyne invented a Hyperspectral Imaging IR Polarimeter concept capable of providing the co-registered, polarimetric imagery desired for locating polarized objects in a dynamic scene and cluttered background. We are in the process of building the prototype hyperspectral IR

¹ Further author information -

Email: scott@aerodyne.com; www.aerodyne.com; TEL: (978) 663-9500; FAX (978) 663-4918

polarimeter under Phase II SBIR support from the AFRL. In a parallel project sponsored by the Army Research Office (ARO) Humanitarian Demining MURI / Northeastern University (NEU) Team, we are taking incremental steps to demonstrate the value and reliability of LWIR polarimetry in the detection of antipersonnel (AP) mines, especially plastic flush-buried and scatterable types. Our stepwise progress and the LWIR polarimetry proof-of-concept demonstrations in the demining application are the topics of this paper.

1.2 Scope

Many researchers have investigated LWIR and multiband passive infrared sensing techniques for applications in demining;¹⁴ a few investigations have considered passive IR polarimetry and costly hyperspectral imaging methods separately for locating land mines;¹⁵ yet insofar as we have found, no previous investigators have pursued the combination of hyperspectral IR polarimetry and affordable LWIR uncooled focal plane arrays. Effort in this Demining MURI is devoted to deliberate, controlled-risk, proof-of-concept demonstrations that verify our particular innovative combination of LWIR hyperspectral polarimetry and affordable uncooled FPA technology can be reliable and effective in surveying and locating exposed surface mines of any material type. The exploitation of polar attributes that are demonstrated in this project effort were in most cases suggested and supported by our earlier analytic investigations into systems applications of IR polarimetry.^{9, 10, 16} In addition to the proof-of-concept tasks, we are investigating surface optical properties of the relevant classes of materials and coatings for mine fabrication; this is a concurrent laboratory effort to fill the important gaps remaining from prior optical properties measurements.

The present effort is focused on antipersonnel mines, many of which are partially exposed on the surface or are buried in a very shallow manner. Therefore, we are placing much greater emphasis on the optical properties of materials, emissivity and reflectance, and less emphasis than usual on the material thermal properties. We make no claims to be dealing with the full complexity of thermal conduction issues associated with buried mines, consisting of a wide range of materials in all manner of soils and terrain. Many investigators^{14,17} have examined the conductance and differential heating/cooling effects for opportunities to exploit in locating buried mines using passive IR techniques. Our premise is simple; IR Polarimetry is only one of a large arsenal of affordable and reliable technologies required to deal with the extensive, worldwide demining and unexploded ordnance problems. Our objective is to make IR Polarimetry an affordable and reliable technique to locate mines having any significant exposure on the surface, regardless of the material composition and including common types of paints and coatings. If we are successful, our method simultaneously positions us to include and exploit significant thermal effects as well since these effects exhibit themselves in the same LWIR wavelength regime. We must give close attention to the magnitude and impact of the heat transfer effects even as we examine the optical properties of materials and coatings.

Since we are implementing a controlled-risk, step-by-step effort to demonstrate a new combination of technologies including the uncooled focal plane array (uncooled FPA) microbolometer, it is important that we establish the terminology for these steps and the technologies to be demonstrated:

- Step (1) - LWIR Spectral Polarimeter
(8 - 12 μm Spectral Region, Non-Imaging).
- Step (2) - LWIR Uncooled FPA Imaging Polarimeter
(8 - 12 μm Broadband, No Spectral Resolution, Static Scenes).
- Step (3) - LWIR Uncooled FPA Multispectral Imaging Polarimeter
(Multiple Bands in 8 - 12 μm Spectral Region, Static Scenes).
- Step (4) - Uncooled Hyperspectral Imaging IR Polarimeter
(8 - 12 μm Spectral Region, Spectral Resolution $\sim 0.02 \mu\text{m}$, Imaging, Registration and Simultaneity in Polar Channels for Dynamic Scenes).

2. LWIR SPECTRAL POLARIMETER DEMONSTRATION TEST

The data reported in this first proof-of-concept demonstration were collected for a variety of mine-like materials and coatings, including Army CARC (Chemical Agent Resistant Coating, both tan and green), a plastic frisbee, asphalt, and quartz sand. The backgrounds consisted of sand or sandy asphalt in all cases. The asphalt is a particularly challenging background, having a degree of polarization in the infrared that is comparable to or higher than many paints and plastics. Table 1 summarizes the data collection geometries and simulated mine target areas for each location. It is particularly

important to note the ranges and the grazing angles between the sensor (polarimeter) line-of-sight and the target normal. These geometries closely approximate a reasonable deployment geometry for an LWIR spectral polarimeter mounted on either a land vehicle with boom or a low flying UAV platform; however, the modified FTIR spectrometer used in this proof-of-concept demonstration was in no way related to our concept for a field deployable multi-channel polarimeter.

The LWIR spectral polarimeter consisted of an FTIR spectrometer with a wire grid analyzer adapted to it. For this demonstration we used 4 cm^{-1} resolution at 10 scans/sec and collected 200 scans per data point per analyzer angle. The detector was MCT with $D^*_{\text{peak}} = 4 \times 10^{10}\text{ cm Hz}^{1/2}\text{ W}^{-1}$. The wire grid analyzer was 1200 lines/mm on a ZnSe substrate with antireflection coatings on both sides.

As illustrated in Fig 1, reflections from a flat surface produce polarization perpendicular to the plane of incidence (the plane of the page). Emission is polarized parallel to this plane.¹⁸ Wire grid polarizers transmit radiation when the E-vector is perpendicular to the wire and reflect radiation when the E-vector is parallel to the wire. We use the convention of 0-degree polarizer angle to mean the wire grid is oriented vertically. Also, we use the spectroscopic notation $\sigma(\sigma)$ for wavenumber (cm^{-1}), and our plots will appear reversed to many readers. The materials used in this demonstration do not transmit in the spectral region of interest, so the simple relationship between the reflectance and emissivity of the materials applies, $\epsilon(\sigma) = 1 - \rho(\sigma)$.

This proof-of-concept demonstration was set up locally in a parking area (see Fig 2) surrounded by several large white pine trees and on the northwest side by a steep embankment of rock outcroppings, rocky soil and scrubs. The parking surface was asphalt, and in the winter months when these data were collected, the asphalt was lightly covered with a residue of sand. The data of interest were collected at locations 4 and 5 as shown in Fig 2 and consisted of fifty polarimeter spectra for the various combinations of target, background, target angle, wire grid analyzer angle, ambient temperature and cloud cover. Again, the setup geometry details are listed in Table 1. Most of the data for these locations were collected in the late afternoon, both before and after sunset, on two successive days with the ambient temperatures between 0- and 15-C. Each spectrum required 20 sec to collect. Since the data collection time was not a primary driver in this demonstration, we made no effort to minimize the time, but it is quite clear that the SNR achieved will allow the collection time to be less than 2 sec per data point per analyzer angle with no important degradation of the results.

2.1 Interpretation of the LWIR spectral polarimeter data

For a view of the spectral data and an understanding of the analysis procedures, we show several overlays of the spectral data in Figs 3 and 4, each at two angles of the wire grid analyzer. As called for in our analysis procedure discussed below, data were normally collected for wire grid analyzer angles of 0-deg (vertical wire grid) and 60-deg; for verification and redundancy we also collected data intermittently at wire grid orientations of 90-deg and 120-deg. Our convention in this analysis is to use the 0-deg and 60-deg wire grid angles as the orthogonal components to differentiate the reflected and thermal emitted radiation components coming from the target and background sources within the field-of-view. The spectra are plotted with the envelope of the shaded area being the spectrum for 0-deg analyzer angle; this spectrum represents the polarized reflection component in the 'horizontal earth' geometry selected for the demonstration. Relative to this the emission component, passed by the wire grid analyzer in the 60- or 90-deg position, is plotted as a thin solid line lying close to and just above the envelope of the shaded area. The band, or the difference, between these two envelopes is a measure of the degree of linear polarization (DoLP) of the spatially unresolved scene in the field-of-view of the spectral polarimeter. When the source radiation is totally unpolarized, the orthogonal polar component spectra are coincident over the full spectral range of sensitivity for the polarimeter. Here the spectral region of interest is $800\text{ - }1250\text{ cm}^{-1}$ (or $8\text{ - }12.5\text{ }\mu\text{m}$). The general procedure for extracting DoLP from the spectra is given below, and the results are plotted in Fig 5.

Alternatively, the narrow bands in between the shaded area envelope and the line plot above represents the spectrally dependent, available polarimetric signal to be exploited. For any given broadband polarimeter the integrated area under the appropriate greybody plot such as Fig 3(a) is a measure of the total signal while the difference of these integrated areas for the orthogonal polarizations is a measure of the polarization signal or the fraction of target radiation that is linearly polarized. From the integrated areas measured for the materials in Figs 3(c), 3(d) and 4(a), we find the fraction of target radiation that is linearly polarized when weighted to fill the IFOV of the sensor to be of the order of 0.1 or 10%.

Upon closely examining the polar component spectra in Figs 3(a) and 3(b), we find that the 0-deg and 60-deg spectra are coincident over the $800\text{ - }1250\text{ cm}^{-1}$ region for each of these materials - soil and sand near normal incidence, so we may

Table 1. Summary of the Data Collection Geometries and Results of the LWIR Spectral Polarimeter Measurements for the Target and Background Materials at Locations 4 and 5.

Target Name and Location	Sandy Asphalt Location 4	Tan CARC Location 4	Plastic Frisbee Location 5	Green CARC Location 5
Plane of Target Relative to Vertical (deg)	86	66	85	79
Background Type	Sandy Asphalt	Sandy Asphalt	Sand Mound	Sand Mound
Range (ft)	30	30	28	28
Target Total Area (sq in)	737	64	33 Exposed	56
Angle Between Sensor LOS and Target Surface Normal (deg)	79	59	75	69
Target Area Projected onto FTIR FOV (sq. in.)	140	33	9	20
FTIR Weighted FOV Footprint at Target (sq. in.)	140	140	130	130
Fractional Area of Total Weighted FOV Occupied by Projected Area of Target	Do Not Know Frac of Asphalt Covered by Sand	0.24	0.07	0.15
Percent of Target Material Radiation that is Linearly Polarized When Weighted to Fill the IFOV of Sensor, for the Indicated Angle Between Sensor LOS and Target Surface Normal	----	8%	14%	13%
DoLP of Material with DoLP of Sand Removed, for the Spectral Region 800-1250 cm^{-1}	0.02	0.02	0.01	0.02

expect the degree of linear polarization (DoLP) for the materials to be very small or zero. Below, we compute the DoLP spectra and indeed show these materials are substantially unpolarized.

This is not the case, however, when we examine the polar component spectra for the plastic frisbee on sand, CARC on sand and sandy asphalt in Figs 3(c), 3(d) and 4(a), respectively. In Figs 3(d) and 4(a) we find large separations in the polar component spectra indicating that CARC and asphalt exhibit strong LWIR polarization effects. Furthermore, the separations in CARC and asphalt show an interesting spectral dependence, being very strong near the greybody curve peak at 925 cm^{-1} and gradually narrowing until the gap closes at about 1250 cm^{-1} . While the general shape of the spectra are determined by the greybody temperature and the spectral polarimeter response function (including the optics, analyzer, antireflection coatings and detector), the sharp spectral features in the $1100 - 1200 \text{ cm}^{-1}$ region are attributed to quartz sand. Notice these features are not present in the soil spectrum, Fig 3(a). These spectra, rich in features, demonstrate very clearly our interest in the LWIR and our rationale for pursuing hyperspectral polarimetry as a powerful technique to discriminate man made target materials against natural background clutter.

2.2 Analysis of polarimetric measurements

The degree and angle of linear polarization (DoLP and AoLP) of a *partially linearly polarized* wave can be obtained from measurements made at three different analyzer angles. The intensity measured at the output of the analyzer oriented at angle α is given by^{10, 18}:

$$I(\alpha) = \frac{1}{2} I_u + I_p \cos^2(\theta_p - \alpha) \quad (1)$$

where

$$\begin{aligned} I_u &= \text{Polarized light from target} \\ I_{u-igt} &= \text{Unpolarized light from target} \\ I_{u-bkg} &= \text{Unpolarized light from background} \\ I_u &= I_{u-igt} + I_{u-bkg} \\ \text{DoLP} &= I_p / (I_{u-igt} + I_p) = \text{Degree of linear polarization} \\ \theta_p &= \text{Angle of linear polarization.} \end{aligned}$$

As is frequently the case with simultaneous trigonometric equations, general solutions in terms of well known functions are difficult to find; a numerical solution can be obtained by searching for it using Newton's method or, alternatively, minimizing an objective function such as the sum of the squared differences between the measurements and the theoretical value (i.e. left hand side minus right hand side of Eq (1)). The field measurements obtained at Aerodyne were performed such that the surface normal vector of the targets was in the plane of incidence, thereby constraining the emitted light to a 90-degree analyzer angle and the reflected light to a 0-degree analyzer angle. Careful attention was paid to orienting the spectral polarimeter such that cold sky was reflected. Under these circumstances, it is reasonable to make the simplifying assumption that $\theta_p = 90$ degrees. Equation (1) then becomes

$$I(\alpha) = \frac{1}{2} I_u + I_p \sin^2(\alpha) \quad (2)$$

By choosing an analysis angle of 0-degrees, the unpolarized intensity is immediately obtained. The polarized intensity can then be obtained by using any other angle (except 180-degrees). In the present case of 90-degree angle of polarization, the greatest signal to noise ratio is obtained by using 90-degrees for the second analysis angle. A 1.25 dB loss in SNR results from the use of 60-degrees instead of 90 in this case.

2.3 Results

Using the simplified analysis approach represented in Equation (2), we extracted a DoLP spectrum for each of the materials observed in this demonstration. Since the quartz sand background was present in the field of view for all of these data, it was subtracted in each case. Otherwise, its region of low emissivity, 1100 - 1250 cm^{-1} , introduces a substantial step function into the DoLP spectra. Other features in the DoLP spectra deserving comment are:

1. Soil is effectively unpolarized and provides a point of reference for the other materials plotted in Fig 5.
2. CARC and sandy asphalt exhibit essentially the same DoLP, indicating they will not be easily distinguished using a broadband polarimeter. Refer ahead to view the CARC panel on sandy asphalt in Fig 9.
3. In the 1000 - 1100 cm^{-1} region the trace effect of ozone remains visible in the spectrum.
4. Fringes apparent in the spectra are attributed to a polarization sensitive beam splitter effect in the FTIR spectrometer.

2.4 The special case of CARC target on sandy asphalt background

The combination of the CARC target imbedded in a sandy asphalt background is a key piece of data from our proof-of-concept demonstration. The reason for this can be seen in Fig 5. Over the spectral interval 800 - 1250 cm^{-1} the measured DoLPs are the same for CARC on sandy asphalt and sandy asphalt alone; CARC and sandy asphalt cannot be differentiated solely by the linear polarization attribute, i.e. by a single broadband imaging polarimeter. In Section 4 we show image data to confirm this. However, this case provides an excellent example of the powerful tool obtained when the polarimetric and spectral capabilities are combined in an LWIR spectral polarimeter. One set of the data for tan CARC on sandy asphalt collected with our basic combination of FTIR spectrometer and wire grid polarizer are shown on a magnified scale in Figs

4(a), 4(b) and 4(c). Figure 4(c) is an overlay of four plots - two for the 'CARC present' and two for 'only sandy asphalt'. By inspection of their spectral dependence we can easily distinguish the presence of CARC in the sandy asphalt background. At 900 cm^{-1} the 0-deg and the 60-deg analyzer plots are identical with and without the CARC present; no differentiation here. However, as we move along the wave number axis toward 1200 cm^{-1} , two effects are evident. First, the separation by analyzer angle for each pair of curves (with and without the CARC present) is becoming less and washes out completely by 1250 cm^{-1} (see Fig 4(c) for this to be apparent in the 'CARC present' case). Secondly, the 'CARC present' traces are decidedly higher than 'without CARC' plots over the spectral region $1050 - 1250\text{ cm}^{-1}$. This effect is shown clearly in Fig 4(c). Therefore, in the spectral dimension (at $1050 - 1250\text{ cm}^{-1}$, not 900 cm^{-1}) we have a very strong differentiation of the CARC present in the sandy asphalt background even though the DoLP for the combined CARC and sandy asphalt background afforded no differentiation. This is true even though separately both the CARC and the sandy asphalt exhibit strong polarization effects. These data make a clear case for the importance of adding the spectral dimension to the LWIR polarimeter. Our goal in the demining application is to demonstrate an affordable and effective polarimeter having both spectral and imaging capabilities.

3. LABORATORY MEASUREMENT OF GRAZING ANGLE IR SPECULAR REFLECTANCE OF ANTIPERSONNEL MINE CASINGS

In order to exploit unique spectral features (particularly low emissivity regions) of land mine casings and similar man made objects (typical plastics such as ABS and blended polymers) we must know the spectral features of both the man made materials and the natural backgrounds in these spectral regions. Our laboratory measurement of grazing angle specular (80-deg) reflectance of four AP mine casings and several related, simulant materials of interest confirm the availability of many LWIR spectral features to be exploited in multi- or hyper-spectral polarimetry. These reflectance spectra exhibit the characteristic dispersion curve shape near the centers of absorption features in the materials.

Our laboratory grazing angle reflectance spectra were collected with an FTIR spectrometer at a resolution of 4 cm^{-1} and generally covering the range of 3 - 15 microns ($666 - 3333\text{ cm}^{-1}$). In each case the sample spectrum has been ratioed to the spectrum of a front surface gold mirror that has a reflectance greater than 0.99 over the entire spectral range covered here. The inerted AP mine casings that were available for study are listed in Table 2 and pictured in Fig 6. In addition, Table 2 includes several related or simulant materials of interest for which we collected reflectance spectra, but not all of these will be discussed here.

Sample grazing angle reflectance spectra are plotted in Figs 7(a), (b) and (c). To make the examination of these spectra easy it is important first to eliminate from further consideration all spectral features attributed to the incomplete removal of water vapor and carbon dioxide from the optical path in the laboratory spectrometer; even though the path was constantly under purge with dry nitrogen, this did not completely remove all atmospheric gases from the sample compartment of the instrument. The spectral features to be ignored are water vapor bands at 2.7 and 6.3 microns and carbon dioxide bands at 4.3 and 15 microns. Then it is important to remember that the dispersion features in the reflectance spectra mark the locations of absorption features for the material. Since the material does not transmit radiation at these characteristic absorption wavelengths, the grazing angle reflectance, like the index of refraction, shows a strong change across the width of an absorption feature.

Our hyperspectral polarimetry approach to developing a reliable detection method for partially exposed scatterable land mines exploits differences in the emissivity and polar attributes of mine materials and coatings relative to these

Table 2. List of Anti-Personnel Mine Casings and Related Materials Included in Laboratory Measurement of Grazing Angle IR Reflectance Spectra.

Anti-Personnel Mine Component	Related Materials or Simulants
M-14 Plastic	CARC (Chemical Agent Resistant Coating), Both Green and Tan
PFM-1 (Butterfly)	
TS-50 Rubber Center	RTV 3110 Silicone Rubber
TS-50 and VS-50 Plastic	Floppy Diskette Plastic Shell
VS-50 Rubber Center	Rubber Hockey Puck
	Asphalt
	Sand
	Soil

same properties for natural backgrounds. Typically, we will examine the spectral reflectance, and thereby the spectral emittance, of the mine materials for unique spectral features and high grazing angle reflectance compared to the backgrounds. With this in mind the measured reflectance spectra presented in Fig 7 have a number of features of interest. First, examining the spectra in Fig 7(c) for similarities, we find the M-14 and Butterfly (PFM-1) mine casings very similar in having very little spectral structure and a difference in overall reflectance level that is accounted for simply by the difference in surface roughness. Although not shown here, the one band at 14.3 microns is common to both the M-14 mine and the floppy diskette shell, making the latter a very sensible simulant for M-14 surface properties. Also in Fig 7(c) the plastic shells of the TS-50 and VS-50 (not over plotted to avoid confusion) have identical spectral features at 8.1, 8.9 and 10.7 microns and are apparently the same polymer blend. We believe this to be an ABS (acrylonitrile butadiene styrene) plastic, but we have more work to do in matching spectra to polymer blends before claiming to have identified the materials.

Second, examining the spectra of the rubber materials in Figs 7(a) and (c), we find that the TS-50 rubber center and RTV 3110 have the same spectral features at 7.87, 8.96, and 12.2 microns. We conclude the TS-50 center element is actually an RTV silicone rubber. Comparing these same features to the reflectance spectrum for the VS-50 center element in Fig 7(c), we find them to be wholly different and learn by comparison to another spectral database that the VS-50 center element is natural rubber.

In Fig 7(b), we examine the spectrum of an Army camouflage paint of interest, green and tan CARC, because it is a low reflectance coating purposefully formulated to have the optical properties of the natural background. The grazing angle spectral reflectance features measured for tan CARC are located at 9 and 12.5 microns, with the 9 micron feature being very robust. The infrared spectral features of CARC have been studied extensively in the Joint Multispectral Program (JMSP)¹⁹ for the purpose of identifying spectral bands that provide a dependable two-color discriminant for CARC in a variety of backgrounds. Using apparent spectral contrast data collected over a wide range of backgrounds and local meteorological conditions, this study concluded that the 9 micron feature provides a persistent two band spectral contrast discriminant for CARC (8.73 / 9.23 microns). This is a successful illustration of our effort to use narrow band spectral features for discrimination of mines within clutter when broadband polarimetry alone cannot eliminate all of the clutter.

In general, our laboratory measurements of grazing angle specular reflectance for the materials listed in Table 2 compare well to the more comprehensive and dedicated measurement efforts of Salisbury and D'Aria²⁰ on the spectral reflectance of terrain background materials, of Johnson and coworkers²¹ on the spectral emissivity variations in disturbed soils, and of Cederquist and coworkers^{22, 23} on the spectral reflectance of target materials and coatings. For insight into the spectral features of blended polymers we found the book by Garton²⁴ to be very useful. Our purpose in making comparisons to these dedicated studies is to confirm our approach and to put a quantitative foundation under our efforts to apply spectral polarimetry to mine detection.

Given that reliable mine detection has proven to be a difficult and frustrating challenge for many qualified investigators, we stand firmly by our stepwise, controlled risk approach as outlined above, and note here again that our laboratory measurement of mine materials spectral reflectance features is a key element in our approach. These data tell us if and where we can expect to apply spectral polarimetry to antipersonnel mine detection successfully. Overall we are encouraged by the availability and robustness of spectral features for plastic mine casing materials found in the LWIR region of interest.

4. LWIR IMAGING POLARIMETER DEMONSTRATION MEASUREMENTS

A series of demonstration measurements were performed on a sampling of inerted non-metallic antipersonnel (AP) mines in various natural backgrounds. The images were collected using an uncooled microbolometer imager operating in the 8 - 12 μm band. A manually rotated wire grid polarizer (1200 lines/mm) was mounted to the camera in front of a 25 mm focal length f/1.0 germanium lens. Although complete characterization of the polarization state requires four independent measurements (three if partial linear polarization is assumed), only two measurements were made in this proof-of-concept demonstration. The justifications for this are twofold. First, it can be assumed that the elliptical polarization is small. Second, the geometry of the scene was chosen such that the angle of polarization for the objects of interest was either vertical or horizontal, in which case the maximum image contrast will be seen in the difference between images made with the polarizer in these two orientations. A sequence of 10 frames was co-added for each polarization angle before subtraction.

The Fresnel reflectance relations indicate that the angle of polarization of reflected radiation is perpendicular to the plane defined by the incident and reflected beams while that of emitted radiation is parallel to this plane. Wire grid polarizers

transmit radiation when the E-vector is perpendicular to the wire and reflect radiation when the E-vector is parallel to the wire. In this paper we use the convention of 0-degree polarizer angle to mean the wire grid is oriented vertically. Consequently, in the intensity difference images (0 - 90 degrees) shown below, reflected (emitted) polarized regions appear brighter (darker) than unpolarized regions. For smooth, horizontal man made surfaces and observing sensor angles near Brewster's angle, the reflected radiation component (E-perpendicular) is near zero, and the emitted component (E-parallel) dominates.

4.1 Discussion of the LWIR imaging polarimeter data

Figure 8(a) shows four mines, a disk of RTV 3110, and a rubber hockey puck buried flush to the surface in sand. The mines, from front to back, are as follows: TS50, VS50, M14, and PFM-1 (Butterfly), followed by the RTV disk and the puck. It is clear from the image that these materials are easily distinguished from this relatively unpolarized and clutter free background. The polarization appears due to reflection of the warm surroundings. It is noteworthy that the top surfaces of the sandbox frame appear significantly polarized, wood being a material used in the construction of certain land mines. A more subtle, yet equally important feature of this image is the contrast evident between the disturbed and undisturbed regions of the sand. A horizontal (i.e., left to right) strip at the rear of the sandbox, extending to the butterfly mine was purposefully raked with a spring steel rake, and the areas surrounding the mines were disturbed to bury them. These areas are clearly distinguishable in the original images. This effect is attributed to a combination of contrast reduction caused by fine-grained particle coatings on the disturbed soil and the increased moisture content of the surface sand in the disturbed regions.^{20,21} In these measurements the undisturbed sand had received two days of sun since the last rainfall, so the surface was very dry. The disturbed areas brought noticeably moist, but not wet, sand to the surface.

Figure 8(b) shows the sandbox rotated 180° so the objects are in reversed order. In this case, the entire surface was well raked and graded to a uniform depth. The uniformity in the image is evident.

Figure 10 shows the same objects in a more natural background consisting largely of undisturbed aged bark mulch. Other elements in the scene include a rock (about 15 cm diameter), a bush, and some small weeds. The two objects labeled "CARC panel" are metal panels painted with CARC. In this image, the absolute value of the polarization difference is taken so emission and reflection polarization energy both appear bright. The bush, a small central portion of the rock, and the rear CARC panel were deliberately underexposed while the front CARC panel and the left edge of the mulch were overexposed. This was done to achieve maximum sensitivity in the region containing the mines and is a limitation of the 8-bit digitizer in the frame grabber rather than the imaging polarimeter itself. Clearly, the mines are evident in Fig 10, yet there is much less contrast against the bark mulch, which itself yields partially polarized radiation, compared to the sand in Figs 8(a) and 8(b). Although the contrast is low, most of the clutter in the thermal image has been suppressed. By contrast, Fig 11, showing only the 0-degree image component of Fig 10, exhibits the wide range of thermal variations present in the scene. Notice that in polarization the two CARC panels appear very similar on the gray scale in Fig 10 while in the intensity image (Fig 11) the sunlit CARC panel appears quite hot in comparison to the shaded panel near the bush.

Figure 9 shows the sandbox using the same camera and lens but now at a greater range (about 4 m). The mines and soil disturbances are still evident, but the scene also includes a large wood guardrail, a strip of aged bark mulch below it, and a vertical asphalt curbing. For two materials in the scene, wood and asphalt, we have both horizontal and vertical surfaces of the same materials in view at the same time. It is useful to observe how the contrast flips for the same material as the orientation and therefore the plane of incidence changes between horizontal and vertical. The vertical surfaces of the guardrail, sandbox wall, and curbing are bright in this difference image (0 - 90 degrees) because the plane of incidence is now horizontal.

There are several other basic polarization phenomena to be observed in this scene (Fig 9). The railing is painted dark brown and consequently heats up considerably during the day. The left end of the railing is seen to yield polarized emission as the view angle increases off-normal. The mulch exhibits polarized reflection when viewed at this grazing angle. The odd-shaped reflective lines near the sandbox are tar-filled cracks in the asphalt. One of the CARC panels is located on the asphalt just beyond the top left corner of the sandbox. Even though the CARC can be distinguished from the asphalt when we know where to look for it, the two exhibit nearly the same degree of linear polarization as seen here in the gray scale image. This confirms the quantitative DoLP results (Fig 5) obtained earlier in the LWIR spectral polarimeter data for CARC on an asphalt background. The top horizontal edges of the wood sandbox are dark, indicating polarized emission from the wood. Since the wood is in contact with the hot sand and asphalt, it can be expected to be warm; however, when wood is in contact with cool moist soil, this may not be the case. Finally, notice the rings of disturbed soil around the mines in the sandbox. Our

consistency of detecting disturbed soil encourages us to refine our approach by using narrow band filters (multispectral) that are optimized for the characteristic spectral features of various soil types.^{20, 21}

4.2 LWIR imaging polarimeter - importance of spatial scale

These imaging polarimeter data yield an empirical understanding of the spatial scale requirements for a polarization sensor. If the sensor instantaneous field-of-view (IFOV) is too large for the spatial scale of the target features in the scene, the signal at the detector will be a superposition of the polarization contributions within one IFOV (and similarly, for an imaging spectrometer, the spectral features will be a weighted mixture of the target and background features within the IFOV). Consequently, the distinctive polarization attributes of the target can be washed out and lost in the clutter. Even though it does not provide a direct example of a wrong choice of spatial resolution, the degree of polarization exhibited by the bark mulch, that is the clutter, in Fig 10 illustrates the importance of spatial scale. Here the bark mulch and the flush-buried mines are similar in dimensions. As a result the bark mulch is an effective background clutter for mines in this instance where the only attribute being measured is the degree of polarization with a broadband polarimeter. It is correct to infer from this that an investigator who chooses to apply broadband polarimetry alone to scenes of arbitrary spatial scale will fail.

For a more direct illustration of the importance of spatial scale, Figs 12(a) and (b) are polarization difference images, recorded under controlled conditions, of the box of bark mulch shown in the photograph, Fig 13. For Figs 12(a) and (b) the angles between the illuminator, the normal to the bark mulch box and the polarimeter were held constant while the range between the bark mulch and the polarimeter increased from 12 ft in Fig 12(a) to 45 ft in Fig 12(b). The IFOV footprint at the bark mulch is about 1/2 inch in Fig 12(a) and 2 inches in Fig 12(b), compared to a mean dimension of 3/4 inches for the bark mulch chips. As expected the polarized reflectance from individually resolved bark mulch chips are visible in Fig 12(a) but washed out by the coarser spatial resolution in Fig 12(b).

Clearly, in the design of a hyperspectral polarimeter, the spatial resolution is an enabling design parameter, making the spectral and polarization attributes of the target accessible to be used as additional discriminants. At the same time, no amount of effort in tailoring the spatial resolution will help in the event the spatial dimension of the clutter is the same as that of the target. In this situation the broad spectral band imaging polarimeter will not be successful, and the need to invoke hyperspectral polarimetry, thereby adding the spectral attributes for discrimination purposes, is clear.

5. RESULTS

5.1 LWIR imaging polarimeter - performance

The microbolometer FPA demonstrated strong performance when applied in the broad spectral band imaging polarimeter mode with the 1200 line/mm polarizer mounted in the optical train. In addition to the imaging polarimeter that we have assembled around an uncooled FPA being affordable, we can report the following performance on static scenes:

1. With f/1 optics and modest additional effort to hold the temperature of the optics constant when operating in the field, the polarimeter delivered a NEDT $< 0.1^{\circ}\text{C}$.
2. The polarimeter consistently detected disturbances in sand, soil and bark mulch created by burying objects and bringing fresh sample to the surface.
3. The system detected flush-buried antipersonnel mine casings and simulants both night and day, with and without cloud cover.

Even though we are quite pleased with the performance of the uncooled FPA microbolometer in the mode of a broadband imaging polarimeter, we have yet to demonstrate its performance as a multispectral imaging polarimeter. Recognizing that the NEDT of the current uncooled FPA technology is an order of magnitude higher than for cooled FPAs, we know the division of the available polarization signature into say 10 (multispectral) or 100 (hyperspectral) bins implies the best uncooled FPA performance as a multi- or hyper-spectral polarimeter will be marginal. Nevertheless, we intend to determine this experimentally since our experience to date with the uncooled FPA indicates that its actual NEDT performance can be as much as a factor of two better than the nominal specification.

5.2 Polarimeter design process and some useful insights

Our current understanding of the best design process for an imaging multi- or hyper-spectral polarimeter application^{10,15,16,25-28} includes the following points:

1. The spatial resolution will be determined and even constrained by the spatial scale of the target and primary clutter sources in the scenes and scenarios of interest. A polarimeter geared toward mine detection requires a spatial scale consistent with the mine dimensions and respective of the spatial scale of primary clutter sources. Recall the bark mulch images in Figs 12(a) and (b).
2. Given this spatial scale constraint the spectral range, resolution, and choice of detector are a compromise between the need for signal (broadband operation) and the availability of distinguishing "narrow" spectral features to exploit (multi- or hyper-spectral operation).
3. Prerequisite to this process is a very clear realization¹⁰ that at least two, and preferably three, independent polarization measurements must be acquired simultaneously at near perfect registration. Otherwise the polarization signal can be washed out by time-dependent effects such as sensor platform motion, substantial changes in the downwelling radiation incident upon the target, and even small temperature and sensitivity drifts in the detector.
4. Hyperspectral polarimetry offers essential advantages over broadband polarimetry in applications such as demining where small targets occupy one IFOV or less (subpixel) and require the use of spectral discrimination techniques.

6. CONCLUSIONS

6.1 Rationale for combining hyperspectral and polarimetric IR imaging

We have presented several pieces of evidence to motivate the marriage of hyperspectral and polarimetric IR imaging. Basically we recognize that polarimetric and hyperspectral attributes supplement one another, and that all available independent supplemental pieces of information can be used to advantage for discriminating target objects from background clutter. While we expect to exploit the broadband attributes of man made objects, knowledge of the characteristic spectral features for a large number of both man made and natural background materials (for example, quartz sand and the mine casing spectra presented above) enables an additional hyperspectral polarization discrimination capability that comes in the form of the narrow spectral features. The reflected radiation from the sun, sky, clouds and local environment can either enhance or wash out the contrast provided by the target emitted radiation and spectral features, but polarization segregates the emitted and reflected radiation so as to maintain the contrast of the emitted component which is characteristically strong in man made objects. Hyperspectral polarimetry offers one more attribute, angle of polarization (AoP), to be used as a discriminant; AoP is a very sensitive and therefore powerful discriminant, but it also demands the very best in sensor performance to measure.¹⁰

Then, there is the issue of spatial scale. On the one hand, we have confirmed that broadband polarimetry alone cannot eliminate background clutter when the man made objects of interest cover a wide range in size from land mines to trucks. On the other hand, imaging polarimetry relieves the exceedingly heavy computational and signal processing burden of pure hyperspectral sensing while still allowing the sensor to be customized to the application. Neither hyperspectrometry nor polarimetry taken separately is as effective as when these tools are combined.

6.2 Overall status of IR polarimetry effort

More important than any one of the above proof-of-concept demonstrations, each of which represents a successful step in the risk-reduction sense, is to put all these elements together into the same sensor, a Hyperspectral Imaging IR Polarimeter; this is our goal and represents our most significant effort to go beyond state-of-the-art. If we are successful, we will have captured the spectral, spatial and polarimetric attributes in one sensor, all simultaneous in time and perfectly registered spatially.

Going beyond the current proof-of-concept sensor configuration intended for static scenes, we now have an innovative sensor design in hand and are building the prototype hyperspectral imaging IR polarimeter. We have verified in a laboratory

visible-band implementation that our design allows us to extract any angle or degree of polarization in the scene at a frame rate determined by the multiplexer and signal processor. For initial application purposes we are using a cooled FPA in the IR prototype, yet we anticipate the opportunity to test uncooled FPAs for the demining/UXO applications. Having all of the spectral, spatial and polarimetric attributes available to us at the same time provides a powerful discrimination and clutter suppression capability that is quite unique. All indications from analyzing our sequentially collected data are that these attributes taken together will comprise a reliable sensor for the detection of surface and flush-buried land mines, as well as for disturbed soil and shallow-buried mines over diurnal cycles or with microwave enhance heating.

Laboratory measurements or other knowledge of the spectral reflectance and polarimetric properties of the targets in each application are necessary to support hyperspectral polarimetry. Our laboratory data help directly to (a) determine the locations and widths of spectral features needed to infer the value added of hyperspectral over multispectral polarimetry and thereby to (b) determine in potential applications whether broadband, multispectral or hyperspectral polarimetry applies.

In summary, using somewhat sparse, but very reproducible, data from field and laboratory measurements to date, we have presented a case for hyperspectral LWIR polarimetry in applications requiring the detection of plastic, as well as metal, flush-buried and scattered surface mines. This approach applies equally well to partially exposed unexploded ordnance (UXO) and to shallow-buried mines subjected to microwave enhanced heating and diurnal cycles. Our rationale for combining hyperspectral and polarimetric imaging was presented.

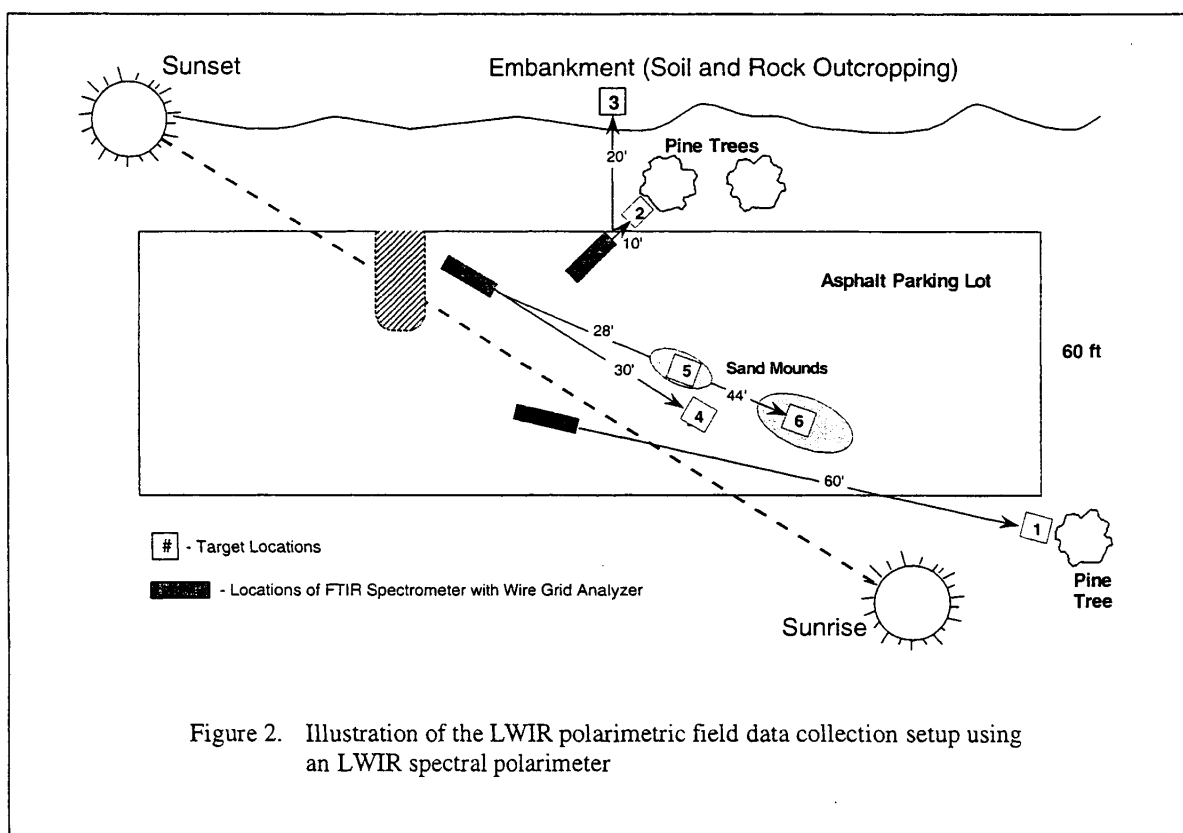
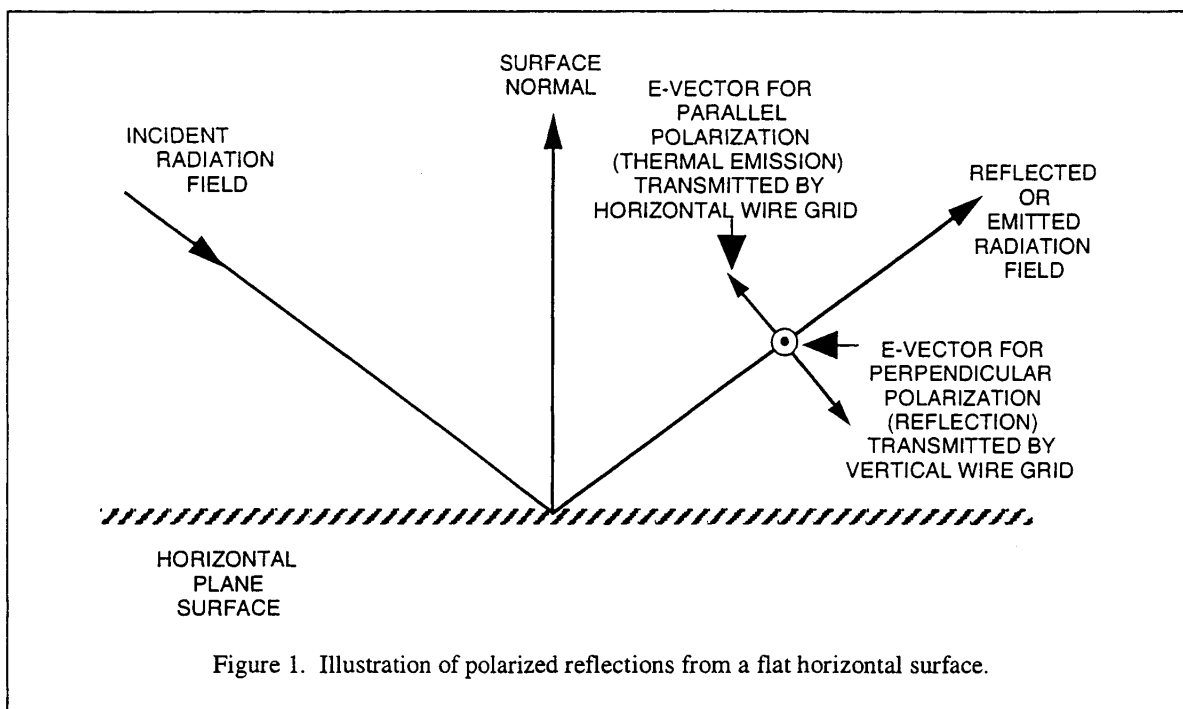
7. ACKNOWLEDGEMENTS

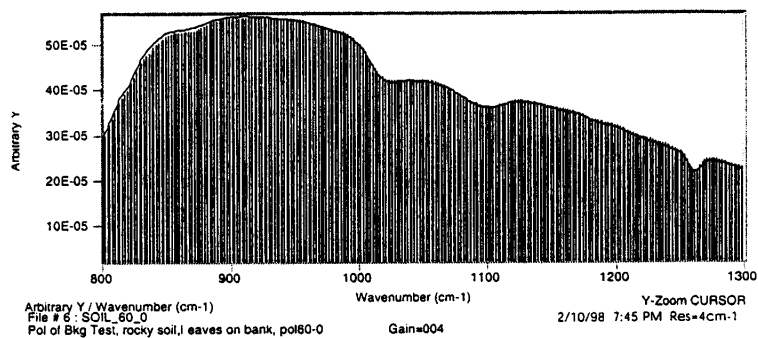
This work was supported by the Northeastern University Humanitarian Demining Multi-Disciplinary Research Initiative (MURI) under contract to the US Army Research Office / Dr. Jim Harvey, Program Director.

8. REFERENCES

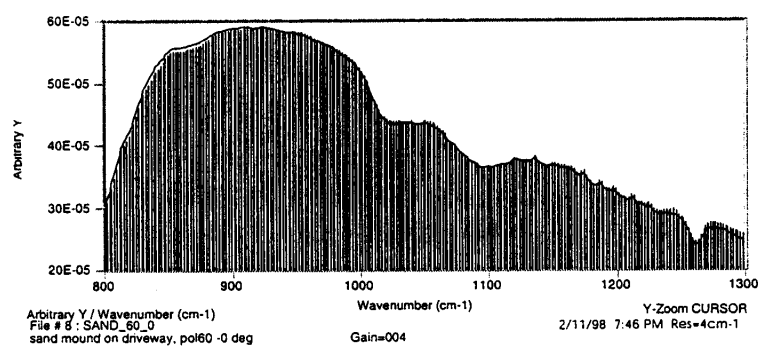
1. M.A. LeCompte, J.A. Conant, and D. Dvornik, "Study of the phenomenology of polarization in aircraft infrared signatures," *Proceedings of the IRIS Specialty Group on Targets Backgrounds and Discrimination*, Monterey, 1988.
2. M.A. LeCompte and J.A. Conant, "POLAR version 3.0 technical manual," Aerodyne Research, Inc., Report No. ARI-RR-1009, 1993.
3. J.A. Conant and M.A. LeCompte, "POLAR version 3.0 users manual," Aerodyne Research, Inc., Report No. ARI-RR-1136, 1995.
4. J.R. Maxwell and S. Weiner, "Polarization radiance (Vol. I, II, III): wavelength dependence of polarized bi-directional reflectance", Report No. 156, ERIM, MI, 1974.
5. R. Tooley, "Man-made target detection using infrared polarization", SPIE, 1166, pp. 52-58, 1989.
6. M.A. LeCompte and J.A. Conant, "Target identification using passive infrared signature polarization," *Proceedings of the Meeting of the IRIS Specialty Group on Targets, Backgrounds and Discrimination*, Naval Postgraduate School, Monterey, 1990.
7. M.A. LeCompte and J.A. Conant, "Modeling and analysis of aircraft infrared signature polarization," *Proceedings of the IRIS Specialty Group on Targets Backgrounds and Discrimination*, Monterey, 1990.
8. M.A. LeCompte, "Aircraft infrared signature polarization model and measurement comparison," *Proceedings of the IRIS Specialty Group on Targets Backgrounds and Discrimination*, Naval Training Center, Orlando, 1992.
9. M.A. LeCompte, and F. Iannarilli, "Infrared signature polarization for passive non-cooperative target recognition," *Proceedings of the IRIS Specialty Group on Targets, Backgrounds and Discrimination*, Naval Training Center, Orlando, 1992.
10. F.J. Iannarilli and M.A. LeCompte, "Predicted performance of counter-air target ID using IR polarimetry *Proceedings of the IRIS Specialty Group on Targets, Backgrounds and Discrimination*, 1994
11. M.R. Hess, G.A. Freund, D.L. McMaster, D.B. Nichols, M.A. LeCompte, F.J. Iannarilli and J.E. Rice, "Measurement and analysis of aircraft infrared signature polarization", *IRIA-IRIS Proceedings, 1994 Meeting of the IRIS Specialty Group on Targets, Backgrounds and Discrimination*, IV, 207-226, 1994..
12. M.R. Hess, M.A. LeCompte, D.B. Nicholls, R.R. Keever, W.B. Shepard, and H.R. Woodman, "Aircraft IR signature polarization measurements at Long Jump '94," *Proceedings of the IRIS Specialty Group on Targets Backgrounds and Discrimination*, Orlando, 1995.
13. M.A. LeCompte, M. Hess, and D. Nichols, "Aircraft IR signature polarization: modeling and measurements," *Proceedings of the Workshop on Infrared and Millimeter Wave Polarimetry*, Redstone Arsenal, AL, U.S. Army MICOM Special Report RD-MG-96-8, pp. 141-155, 1995.

14. J.R. Simard, "Experimental Evaluation of the Apparent Temperature Contrast Created by Buried Mines as Seen by An IR Imager," Defense Research Establishment Suffield, Canada, Report Number SR 607 (November 1994).
15. Li-Jen Cheng and G. Reyes, "AOTF Polarimetric Hyperspectral Imaging for Mine Detection," Jet Propulsion Laboratory, Proceedings SPIE Vol. 2496, 305 (1997).
16. S. Jones, F.J. Iannarilli, H.E. Scott, and K. Annen, "Data Management for Imaging Spectrometers," Air Force Research Laboratory, Space Vehicles Directorate Report AFRL-VS-HA-TR-98-0005. Hanscom AFB (January 1998).
17. P. Li and A. Maad, "Infrared Imaging of Buried Objects by Thermal Step-Function Excitations," *Applied Optics* **34**, 5809 (1995).
18. M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, Oxford, 1970.
19. A.D. Stocker, et al, "Analysis of infrared hyperspectral measurements by the Joint Multispectral Program," *SPIE Conference on Targets and Backgrounds: Characterization and Representation*, Orlando, SPIE 2469, pp.587-602, 1995.
20. J.W. Salisbury and D.M. D'Aria, "Emissivity of terrestrial materials in the 8-14 micron atmospheric window," *Remote Sens. Environ.* **42**, pp. 83-106, 1992.
21. J.R. Johnson, P.G. Lucey, K.A. Horton, and E.M. Winter, "Infrared measurements of pristine and disturbed soils. 1. Spectral contrast differences between field and laboratory data," *Remote Sens. Environ.* **64**, pp 34-46, 1998.
22. J.N. Cederquist, et al, "Multispectral infrared target detection: phenomenology and modeling," *SPIE Proceedings*, Vol. 1954, 1993.
23. T.J. Rogne, F.G. Smith, J.E. Rice, "Passive target detection using polarized components of infrared signatures," Orlando, SPIE, 1317, pp. 242-251, 1990.
24. A. Garton, *Infrared Spectroscopy of Polymer Blends, Composites and Surfaces*, Hanser Publishers, Munich, 1992.
25. M.A. LeCompte, F.J. Iannarilli, D.B. Nichols, and R.R. Keever, "Multispectral IR signature polarimetry for detection of mines and unexploded ordnance (UXO)," *SPIE Conference on Detection Technologies for Mines and Minelike Targets*, Orlando, SPIE, 2496, pp. 180-191, 1995.
26. B.A. Barbour, M.W. Jones, H.B. Barnes and C.P. Lewis, "Passive IR polarization sensors: a new technology for mine detection", *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets III*, Orlando, SPIE, 3392, pp. 96-103, 1998.
27. M. Larive, L. Collot, S. Breugnot, H. Botma, P. Ross, "Laid and flush-buried mines detection using 8-12 micron polarimetric imager", *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets III*, Orlando, SPIE, 3392, pp. 115-120, 1998.
28. C.A. DiMarzio, T. Vo-Dinh, H.E. Scott, "Some approaches to infrared spectroscopy for detection of buried objects", *SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets III*, Orlando, SPIE, 3392, pp. 158-166, 1998.

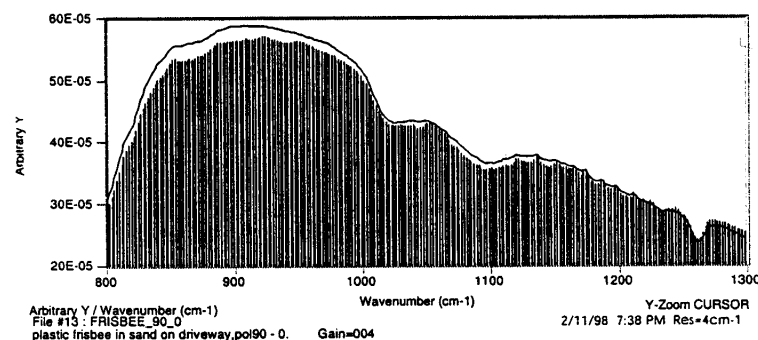




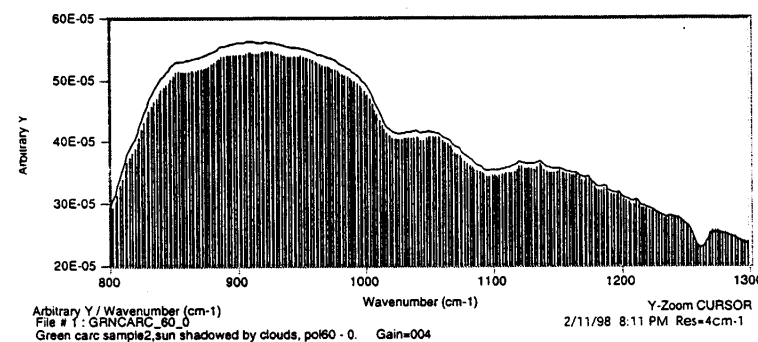
(a) SOIL



(b) SAND

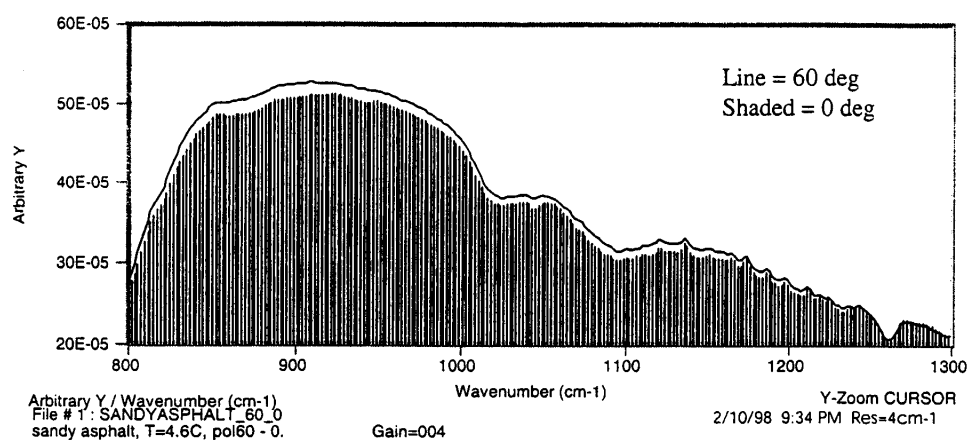


(c) FRISBEE

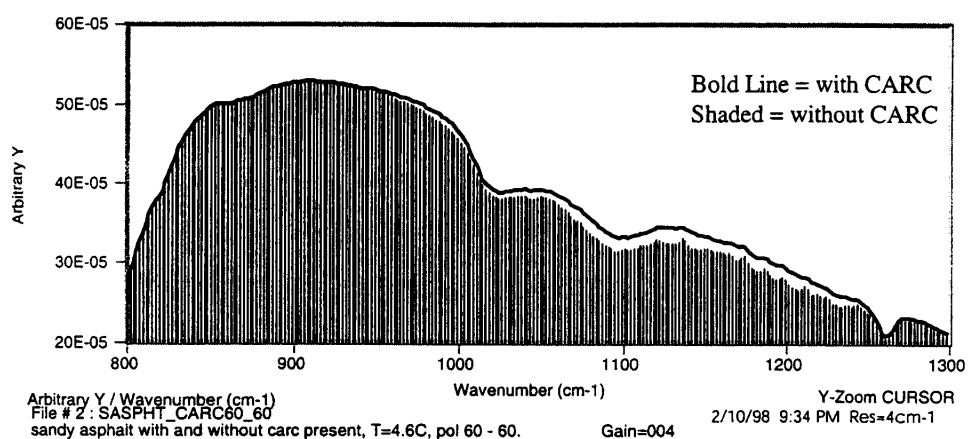


(d) CARC

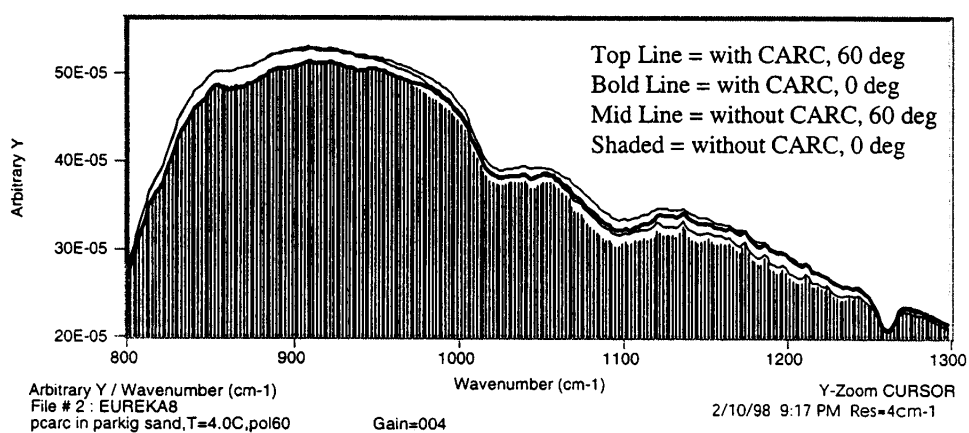
Figure 3. Raw data plots from spectral polarimeter, each at two analyzer angles, for four target materials. The envelope of the shaded area is the spectrum for analyzer angle 0-deg, and the solid line for analyzer angle 60-deg.



(a) Sandy Asphalt
(0, 60 deg)

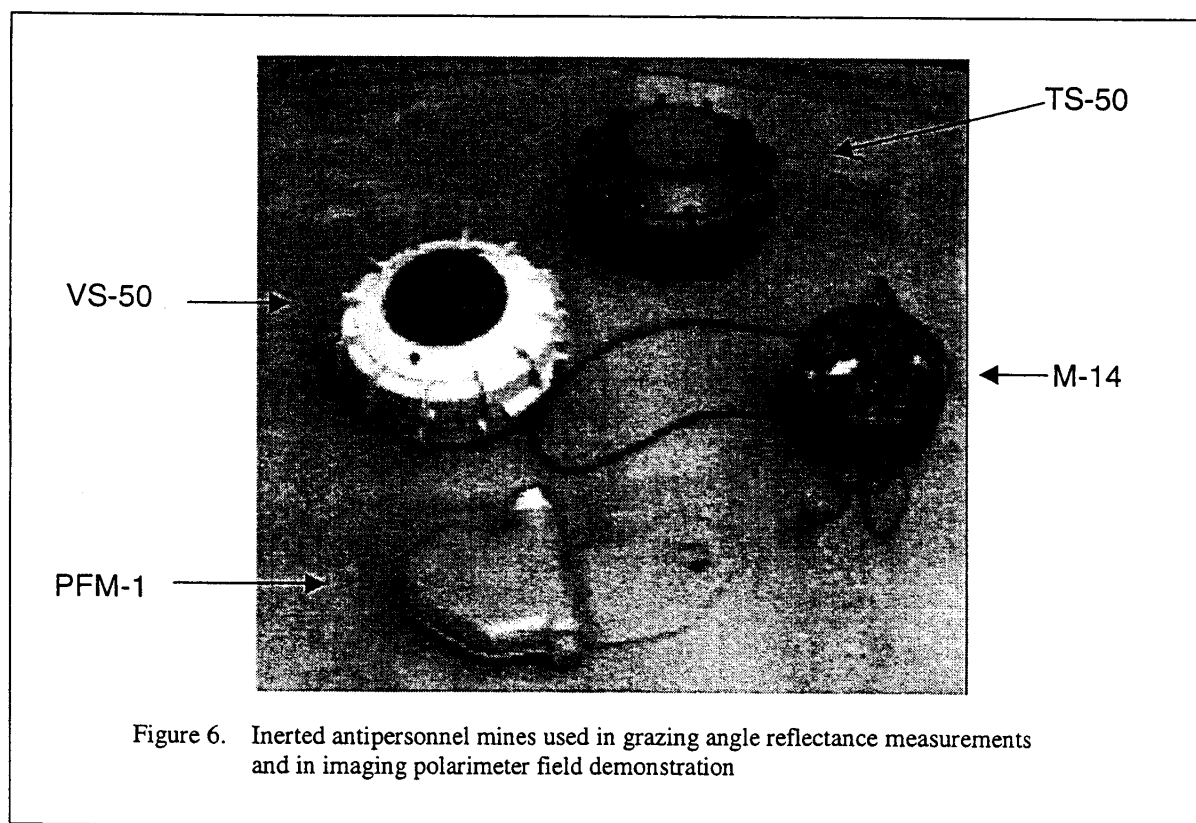
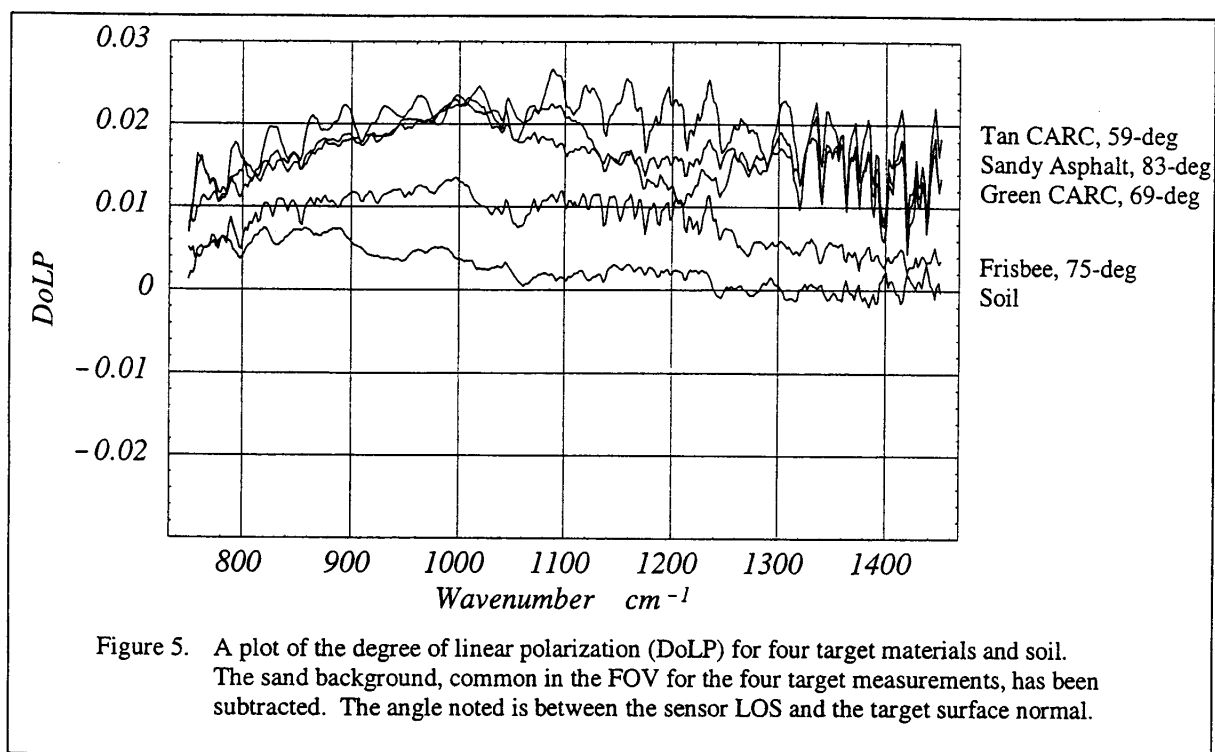


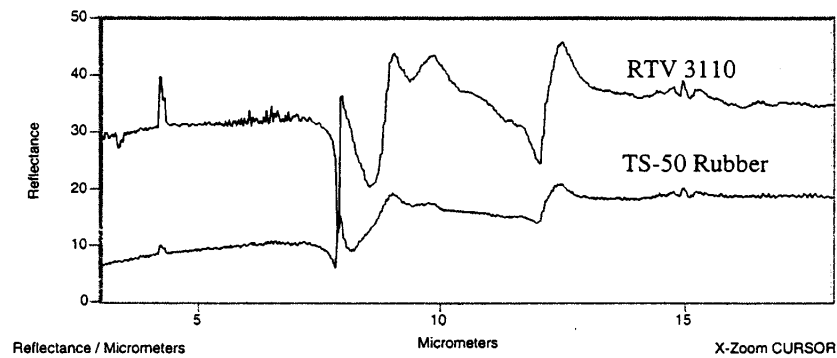
(b) Sandy Asphalt with
CARC (60 deg) and
without CARC
(60 deg)



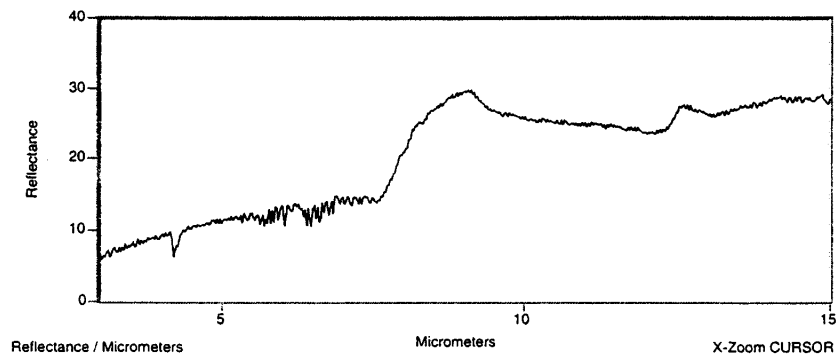
(c) Sandy Asphalt with
CARC (0, 60 deg) and
without CARC
(0, 60 deg)

Figure 4. Spectral polarimeter measurements at the noted analyzer angles for sandy asphalt with and without CARC panel present. The relative intensity for the same analyzer angle is always greater with the CARC panel present in the FOV.

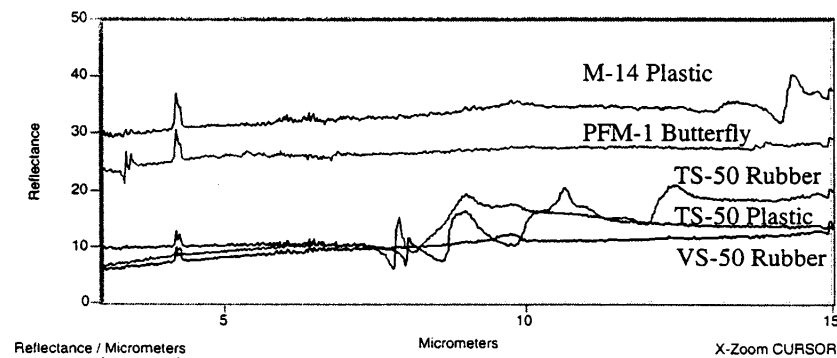




(a) Comparison of grazing angle reflectance of RTV 3110 and TS-50 mine casing (rubber center)



(b) Grazing angle reflectance of Tan CARC



(c) Comparison of grazing angle reflectance of antipersonnel mine casings

Figure 7. Laboratory measurements of grazing angle specular reflectance for several Antipersonnel mine casings and related materials of interest.

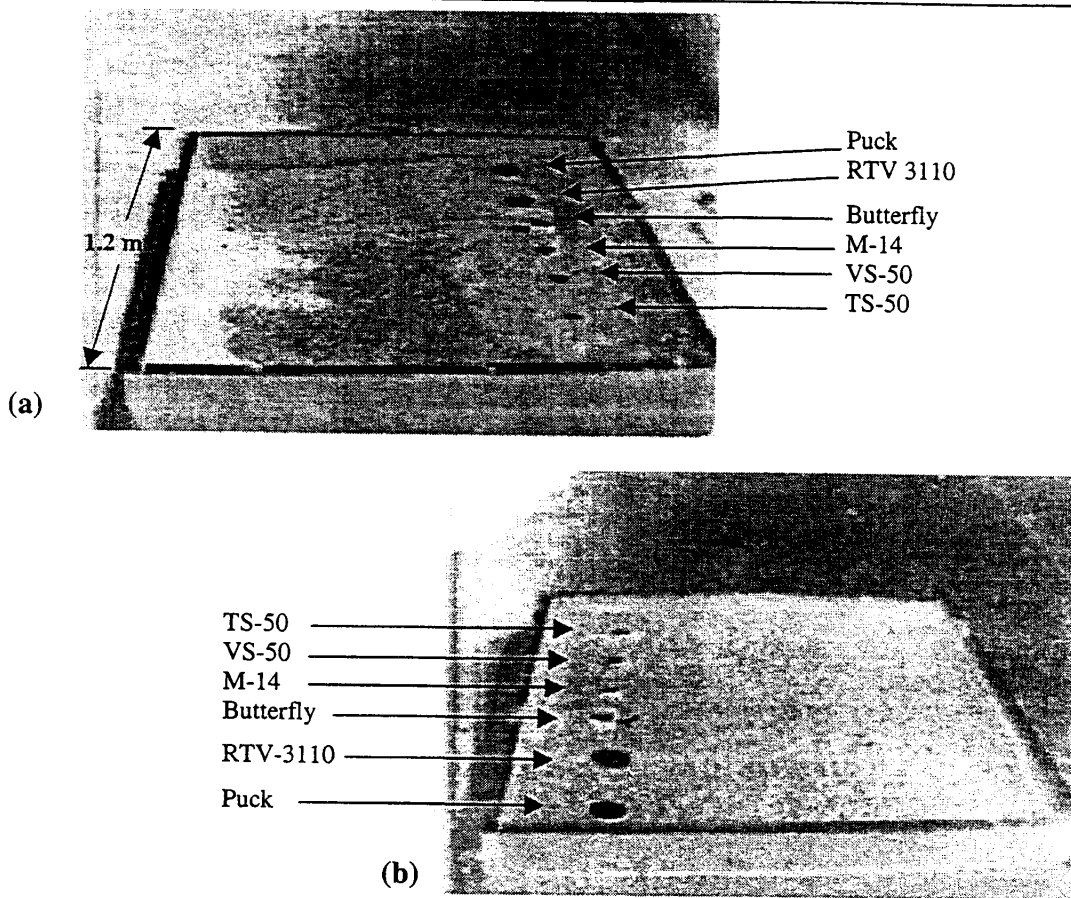


Figure 8. Polarization difference image (0 - 90 degrees) of landmines in sand. (a) Note the raked area at top and disturbed area around lower three mines; (b) Note the sand has been raked and graded to a uniform depth.

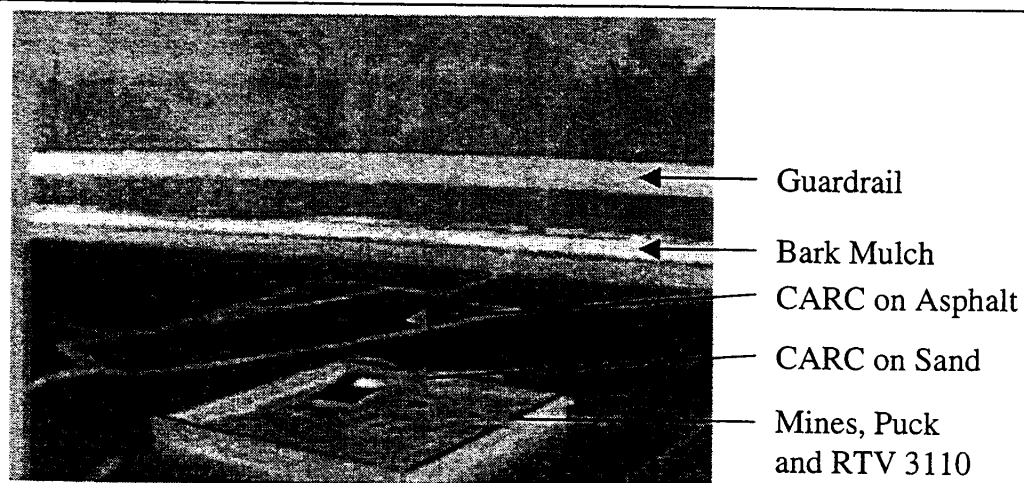


Figure 9. Polarization difference image (0 - 90 degrees) of landmines in sand. Note the polarized emission from the vertical faces of the guardrail and asphalt curbing and the polarized reflection from the horizontal strip of bark mulch beneath the railing.

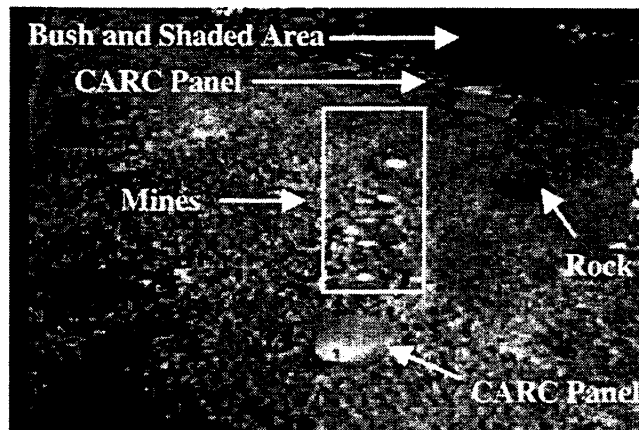


Figure 10. Polarization absolute value difference image, Abs(0-90 degrees), of landmines in bark mulch background with various natural debris and small plants.

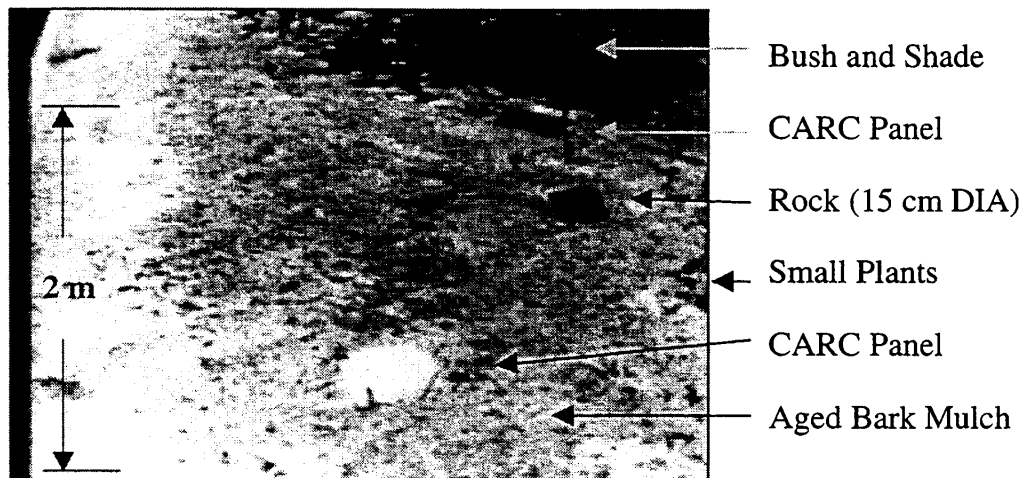


Figure 11. Intensity image (at 0-degree analyzer angle) of landmines in bark mulch background with various natural debris and small plants.

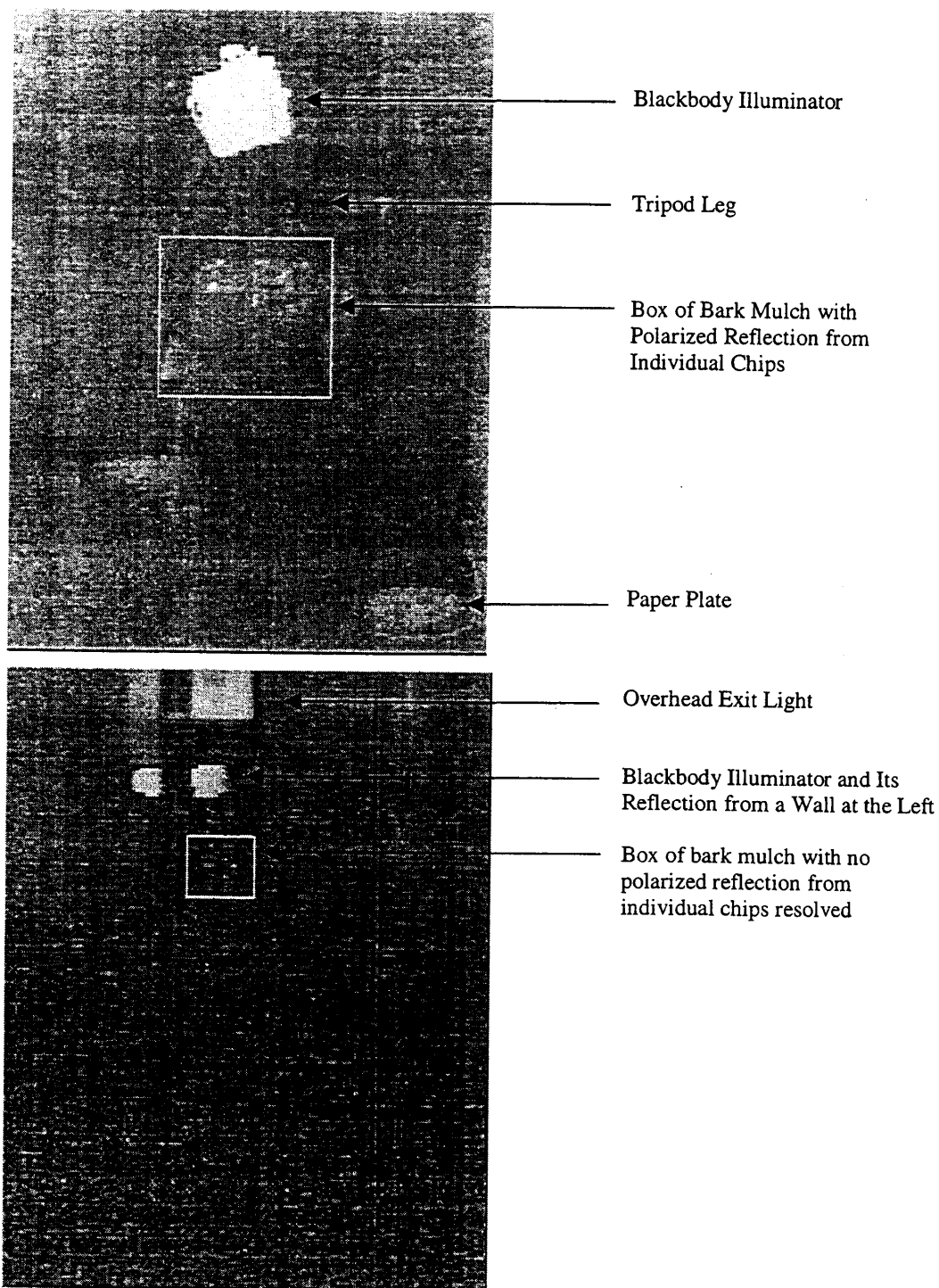


Figure 12. Polarization difference image (0-90 degrees) of bark mulch (a) at a range of 12 ft. Note the polarized reflection from individual bark mulch chips at this range where detector IFOV footprint is about 0.5 inches; (b) at range of 45 ft. Note the bark mulch, directly below the black body illuminator, is not discernible from the background at this range where detector IFOV footprint is about 2 inches.

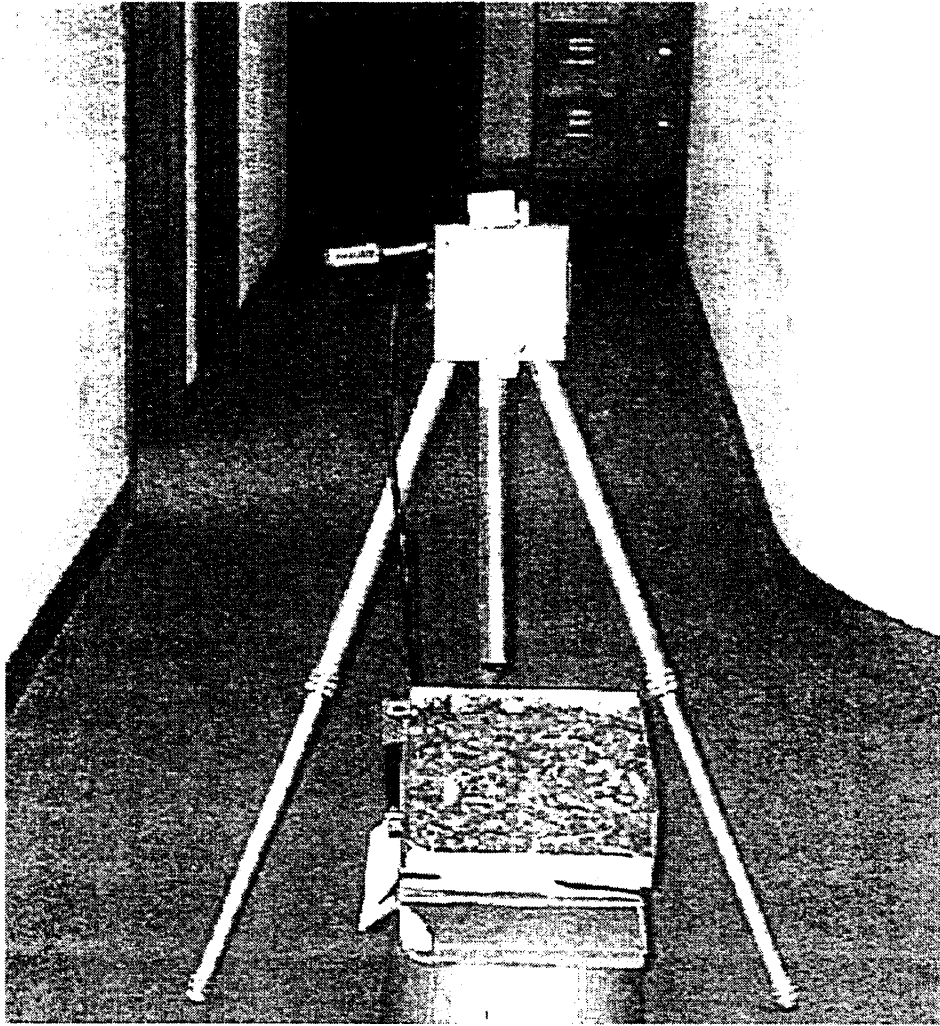


Figure 13. Bark Mulch Spatial Scale

Statistical Clutter Modeling and Parameter Estimation for the Characterization of buried Objects using Frequency Domain Electromagnetic Induction Sensing

Mustafa Özdemir^a, Eric Miller^a, Alan Witten^b

^a Center for Electromagnetics Research, Northeastern University, Boston, MA.

^b School of Geology and Geophysics, The University of Oklahoma, Norman, OK.

ABSTRACT

The problem of low metal content mine characterization from broadband electromagnetic induction (BEMI) data is addressed. A stochastic model describing the spatial distribution of clutter is developed and methods for estimating and removing this unwanted interference are described and tested. After removing the clutter from the signal, a technique is introduced for extracting from BEMI data information describing the location, orientation and structure of the buried object. Examples are provided for spherical and ellipsoidal mines.

1. INTRODUCTION

Broadband electromagnetic induction (BEMI) methods represent a new and very promising sensing technology for the detection, localization, and characterization of buried metallic objects such as landmines and unexploded ordnance.² The data taken over a broad frequency range (tens of hertz to tens of kilohertz) convey information which can be used for not only to detect the presence or absence of an object but also to determine the shape, size, orientation, and material characteristics.

In this paper, we look into problems related to the application of BEMI systems to the characterization and classification of low metal content buried objects. In previous work, we have developed physical models to describe a BEMI system and associated processing methods to extract from BEMI data collected over a grid of points in the neighborhood of the object information regarding the location, orientation and structure of a buried object by estimating dipole moment spectra (DMS).³ We have verified the ability of these methods for purposes of classification using real sensor data for problems where the objects contain significantly metal. However, for problems involving low metal content objects, new processing techniques are needed. In particular, for these cases the signal arising from volumetric inhomogeneities in the electromagnetic properties of the earth (permittivity and conductivity) can be of the same order or magnitude if not larger than the signal arising from the object under investigation. Moreover, this "clutter signal" is known to enter the data *additively* suggesting one method of mitigating the clutter would be to estimate and subtract it from the data.

We propose to model the clutter as a correlated random field which can be described using a polynomial regression model the structure of which is motivated by examination of real clutter data collected with a GEM-3³ sensor. The estimate then subtract processing strategy we propose is designed to reflect the way in which BEMI-type sensors are employed in the field. Currently clutter mitigation amounts to subtracting from data taken in the immediate vicinity of the object target-free secondary data taken on the boundary of this area. Thus, the correlation structure of the clutter is not properly accounted for in the mitigation procedure. Moreover, this approach completely ignores the fact that the sensor is often calibrated in a region close to a suspected target. Thus any information which the calibration data may be able to yield regarding the clutter structure over the object is also absent from the processing.

Here we consider a model-based approach to BEMI clutter mitigation. The data from the calibration region as well as the boundary of the object region are all used to estimate and remove the clutter in the data containing the object signal. After cleaning the data in this manner, we describe a new set of methods for estimating the object characteristics: location, orientation, and DMS. Relative to the statistically optimal algorithm proposed in [3] for this task, the one developed here, though theoretically suboptimal, is faster and still highly accurate.

The remainder of this paper is organized as follows. The BEMI physical model is reviewed in §2. In §3 the clutter model is described while §4 is devoted to the development of our processing methods. Examples demonstrating the performance of this approach are provided in §5. Finally, in §6, we provide conclusions and indicate the future work.

2. PHYSICAL MODEL

We consider an extension of a physical model for EMI proposed in [1] describing the scattering of low frequency electromagnetic radiation by spherical or spheroidal objects of known conductivity and permeability. As seen in Fig. 1 the transmitters and receivers are taken to be square coils (not necessarily co-located) with sides of length $2A$. The target center is located at $r_0 = (x_0, y_0, z_0)$ in the $x - y - z$ coordinate system. We are concerned with processing methods based on multi-frequency data obtained

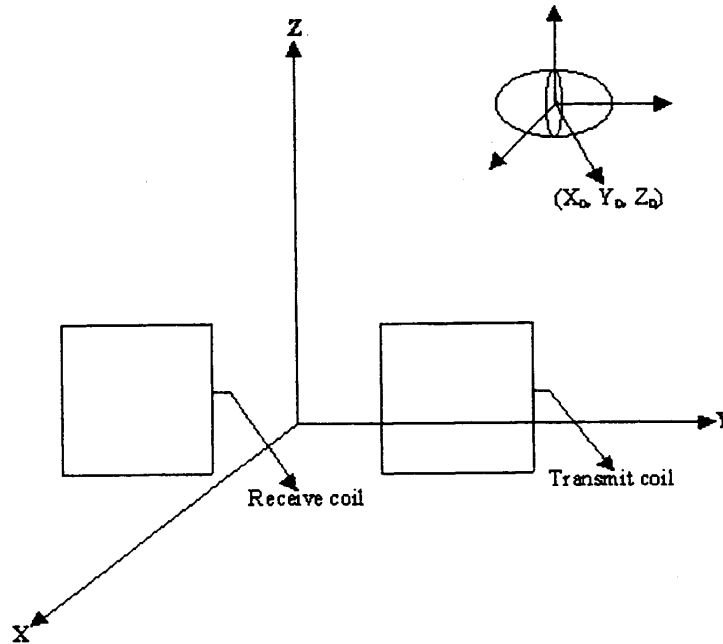


Figure 1. One sensor comprising sensor coils and target object.

from multiple transmitter/receiver locations. Assuming we collect M frequency samples from each of N combinations of transmitters and receivers positions then under the model the k th frequency sample at the n th position is

$$s_{n,k} = \frac{i\omega_k\mu_0}{I} g_n^T R^T \Lambda_k R f_n + w_{n,k} + c_{n,k} \quad (1)$$

where Λ_k is the normalized polarizability tensor for the k th frequency, R is the rotation matrix, g is a 3×1 vector holding the x , y , and z components of the magnetic field produced at r_0 by a current I flowing in the receive coil, g^T indicates the transpose of g , f is the excitation field vector evaluated at the dipole

position and has a similar functional form to that of g , ω is the operating frequency, $i = \sqrt{-1}$, μ_0 is the permeability of free space, $w_{n,k}$ is measurement noise, and finally $c_{n,k}$ is the clutter in the data for the k th frequency sample at the n th position. Details about f and g are provided in Appendix A of [1]. The matrix Λ takes the form:

$$\Lambda(\omega) = \begin{bmatrix} \lambda_1(\omega) & & \\ & \lambda_2(\omega) & \\ & & \lambda_3(\omega) \end{bmatrix}. \quad (2)$$

The three frequency dependent λ 's (here referred to as dipole moment spectra) each are associated with one of the principal axes of the object. For a sphere, all three are identical and closed form expressions can be found for all orders of multipoles.² Finally, R represents a rotation matrix used to transform field quantities between a global frame of reference and the local frame of the object. Here R is parameterized by 3 Euler angles⁷ and explicitly takes the form

$$R = \begin{bmatrix} \cos \phi \cos \psi - \sin \phi \cos \theta \sin \psi & -\cos \phi \sin \psi - \sin \phi \cos \theta \cos \psi & \sin \phi \sin \theta \\ \sin \phi \cos \psi + \cos \phi \cos \theta \cos \psi & -\sin \phi \sin \psi + \cos \phi \cos \theta \cos \psi & -\cos \phi \sin \theta \\ \sin \theta \sin \psi & \sin \theta \cos \psi & \cos \theta \end{bmatrix} \quad (3)$$

From this data set, our goal is to determine the position, the orientation, and the moment spectra of the buried object.

3. CLUTTER MODEL

For low metal content objects a simple additive white Gaussian noise model is not satisfactory. The interaction of the transmitted signal with the background medium, usually negligible for sensing metal objects, become prominent here. These effects are manifest in the form of additive, correlated noise in the signal which we term "clutter". In this work we develop a stochastic model describing the distribution of clutter which provides for the spatial correlation seen in this portion of the sensor signal.

Specifically, we consider the following polynomial regression model in the spatial variables x_i and y_i , the x and y position of the i th sensor to describe this clutter at frequency ω_k :

$$c(x_i, y_i, \omega_k) = \sum_{p,q} \alpha_{p,q,k} x_i^p y_i^q + n_{i,k} \quad (4)$$

where the α 's are unknown, random expansion coefficients, and $n_{i,k}$ represents residual, "white" variations not captured by the regression. Collecting the clutter samples at all locations and all frequencies into a signal vector we write the overall model as

$$c = X\alpha + n. \quad (5)$$

For M frequencies $X = I_M \otimes X'$ where I_N is the $N \times N$ identity matrix, \otimes denotes the Kronecker product and X is the block diagonal matrix obtained from all the X' 's, where the element of X' for (i, j) th position is $x_i^p y_j^q$, α is the vector containing $\alpha_{p,q,k}$, and n is the noise vector.

As described in §1, we use (5) to describe the distribution of clutter over two regions of space: a calibration area and a region containing an object to be characterized. As illustrated in Fig. 2 the clutter mitigation procedure we propose makes use of *all* the calibration data and the data taken on the *boundary* of the object region to estimate the clutter signal on the interior of the object region. One supposition here is that the boundary data do not contain any mine signal. Another is that there is some correlation in the clutter from one region to the next which can and should be exploited in the processing. Thus,

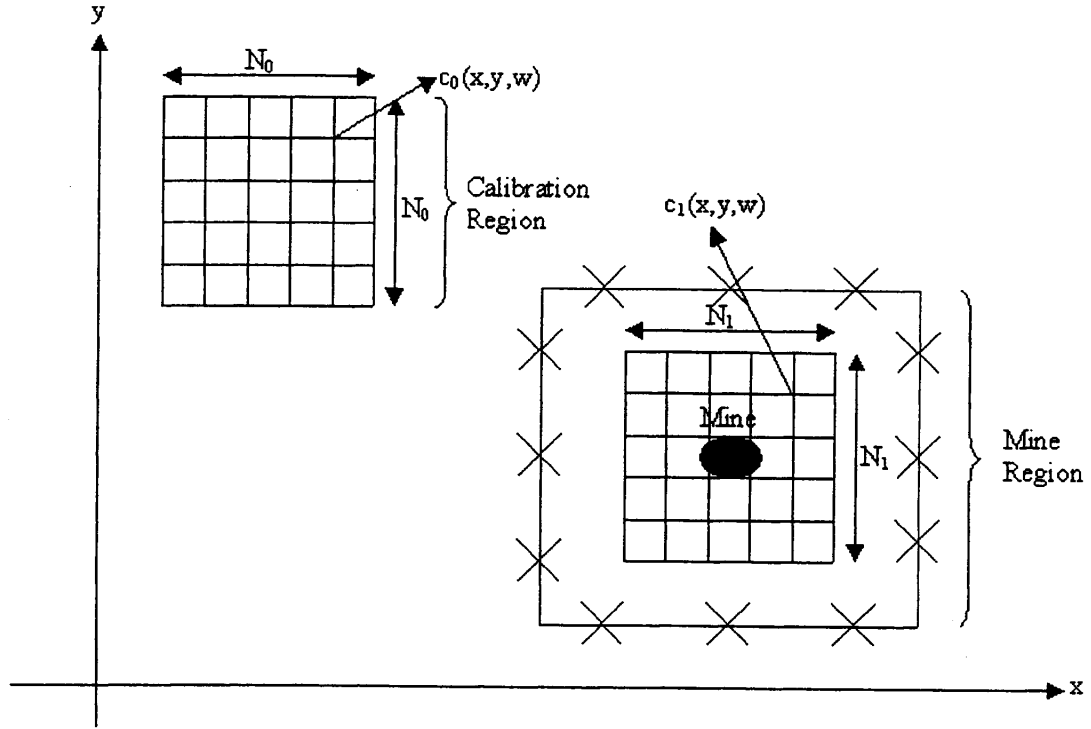


Figure 2. Clutter Model.

we introduce a simple statistical model linking the α vector from the clutter region to that of the object region. Formally, over the calibration region we write the clutter as

$$c_0 = X_0 \alpha_0 + n_0 \quad (6)$$

while over the object area we have

$$c_1 = X_1 \alpha_1 + n_1 = \begin{bmatrix} X_{1,b} \\ X_{1,i} \end{bmatrix} \alpha_1 + \begin{bmatrix} n_{1,b} \\ n_{1,i} \end{bmatrix}. \quad (7)$$

with $X_{1,i}$ built from points interior to the mine region, $X_{1,b}$ from the boundary points (marked by "X" in Fig. 2), $n_{1,i}$ interior noise samples and $n_{1,b}$ boundary noise samples. To complete the model of the clutter we assume that the vector α_1 is $N(0, \sigma_\alpha^2 I)$, and we hypothesize that α_0 and α_1 are related via random walk type model of the form

$$\alpha_0 = \alpha_1 + n_2 \quad (8)$$

Finally, for simplicity we take $n_j \sim N(0, \sigma_j^2 I)$ for $j = 0, 1, 2$.

4. PROCESSING

4.1. Clutter estimation and mitigation

Given the clutter model described in the previous section, our first objective is to find an estimate of α_1 given c_0 (the calibration region data), and $c_{1,b}$ (the clutter data taken on the boundary of the object region), so that we can estimate the clutter data for the whole mine present region. Toward this end, we substitute (8) into (6) to obtain;

$$c_0 = X_0 \alpha_1 + X_0 n_2 + n_0 \quad (9)$$

Combining this with (7), yields the complete clutter model

$$\underbrace{\begin{bmatrix} c_0 \\ c_1 \end{bmatrix}}_c = \underbrace{\begin{bmatrix} X_0 \\ X_1 \end{bmatrix}}_D \alpha_1 + \underbrace{\begin{bmatrix} I & 0 & X_0 \\ 0 & I & 0 \end{bmatrix}}_E \underbrace{\begin{bmatrix} n_0 \\ n_1 \\ n_2 \end{bmatrix}}_n \quad (10)$$

with $En \sim N(0, R)$ and $R = E \text{ blockdiag}(\sigma_0^2 I, \sigma_1^2 I, \sigma_2^2 I) E^T$. Eq. (10) provides a linear model relating all of the clutter data of interest to the expansion coefficients over the region containing the object. Using this model, the linear least squares estimate of α_1 based on the clutter data taken over the calibration region and the boundary of the mine region is³

$$\hat{\alpha}_1 = (D_r^T R_r^{-1} D_r)^{-1} D_r^T R_r^{-1} M_0 c \quad (11)$$

where M_0 is a selection matrix that extracts from c the c_0 and $c_{1,b}$ subvectors, $D_r = M_0 D$, and $R_r = M_0 R M_0^T$. Then, the estimate of the clutter data for the interior of the mine present region is

$$\hat{c} = X_{1,i} \hat{\alpha}_1 = X_{1,i} (D_r^T R_r^{-1} D_r)^{-1} D_r^T R_r^{-1} M_0 c \equiv M_1 c. \quad (12)$$

We mitigate the clutter in the signal as follows: Collecting the data over all frequencies and positions we write the model in (1) as;

$$s = s_0 + c_{1,i} + \omega = s_0 + M_2 c + \omega \quad (13)$$

with s_0 the vectorized form of the first term in (1), $c_{1,i}$ clutter on the *interior* of the object region, and M_2 the matrix which extracts from c the $c_{1,i}$ subvector. The noise vector ω is $N(0, \sigma_\omega^2 I)$. Then, subtracting \hat{c} from the data vector s yields the clutter mitigated data, or cleaned data, \bar{s} ,

$$\bar{s} = s_0 + M_2 c - \hat{c} + \omega = s_0 + (M_2 - M_1) c + \omega \equiv s_0 + M c + \omega$$

Thus, the cleaned data are $N(s_0, R_{\bar{s}})$ with (after some algebra)

$$\begin{aligned} R_{\bar{s}} &= M R_c M^T + \sigma_\omega^2 I \\ R_c &= \sigma_\alpha^2 D^T D + R \end{aligned}$$

4.2. Target parameter estimation

Given \bar{s} , our aim is to estimate the parameters of the detected object: the co-ordinates of the object center, the moment spectra, and the three rotation angles. Here we take a two-step approach to this procedure. First, we use the data to estimate the three location parameters of the object, (x_0, y_0, z_0) and a collection of quantities related to the Euler angles and the DMS. Second, we use these estimates to separately extract orientation and DMS information. The motivation for this approach is primarily computational. As described in greater detail below, each stage requires the solution of a problem involving a single large parameter vector which is *linearly* related to the data and a substantially smaller set of parameters for which the relationship is *non-linear*. By pursuing a two step strategy, we can exploit this structure to obtain an estimation approach requiring two small non-linear search routines rather than one larger one. Moreover, the first such routine for the location parameters is better behaved in terms of local minima than the second search for the Euler angles. Thus, we are able to effectively partition the overall estimation problem.

Our approach to the first subproblem is to start by defining the symmetric matrix M_k

$$M_k = R^T \Lambda_k R = \begin{bmatrix} \mu_{11,k} & \mu_{12,k} & \mu_{13,k} \\ \mu_{12,k} & \mu_{22,k} & \mu_{23,k} \\ \mu_{13,k} & \mu_{23,k} & \mu_{33,k} \end{bmatrix} \quad (14)$$

Substituting (15) into (1), "stacking" the data from all transmitter-receiver pairs for all frequencies, we arrive at the following model for the cleaned data

$$\bar{s} = B(r_0)\mu + \Omega. \quad (15)$$

where, for M frequencies $B = I_M \otimes B_1$ with B_1 a matrix constructed from the f_n and g_n vectors. The vector μ is comprised of the six *unique* elements of each M_k . Finally, the noise vector Ω is zero mean and Gaussian with variance $R_{\bar{s}}$.

Eq. (15) is used in a least squares approach to determine the location of the object, r_0 and μ , as follows:

$$\hat{\mu}_0, \hat{r}_0 = \arg \min_{\mu, r_0} \|\bar{s} - B(r_0)\mu\|_{R_{\bar{s}}}^2 \quad (16)$$

The solutions³ are found:

$$\hat{r}_0 = \arg \min_{r_0} \|\bar{s} - B(r_0)(B^T(r_0)R_{\bar{s}}^{-1}B(r_0))^{-1}B^T(r_0)R_{\bar{s}}^{-1}\bar{s}\|_2^2 \quad (17)$$

$$\hat{\mu} = (B^T(\hat{r}_0)R_{\bar{s}}^{-1}B(\hat{r}_0))^{-1}B^T(\hat{r}_0)R_{\bar{s}}^{-1}\bar{s} \quad (18)$$

The goal of the second processing step is to use $\hat{\mu}$ to estimate λ and α = the vector of three Euler angles. Via (14), we start by using $\hat{\mu}$ to build \hat{M}_k in the obvious manner. According to (14), we should be able to find a *single* rotation matrix which simultaneously diagonalized *all* of the M_k 's to produce the diagonal Λ_k 's. We use this observation to construct the following penalized least squares cost function

$$C(\alpha_k, \lambda) = \sum_k \|R(\alpha)\hat{M}_k R^T(\alpha) - \Lambda'_k\|_F + \text{penalty} \quad (19)$$

where $\|X\|_F$ is the Frobenius norm of the matrix X ,

$$\Lambda'_k = \begin{bmatrix} \lambda'_{11,k} & \lambda'_{12,k} & \lambda'_{13,k} \\ \lambda'_{12,k} & \lambda'_{22,k} & \lambda'_{23,k} \\ \lambda'_{13,k} & \lambda'_{23,k} & \lambda'_{33,k} \end{bmatrix}$$

is the matrix containing the moment spectra and it is not generally diagonal due to the fact that the noise in the data will prevent the exact simultaneous diagonalization of all the M_k . With this in mind, the goal of penalty is to (a) discourage nonzero off diagonal entries in every Λ'_k and (b) to encourage smoothness in the $\lambda_{i,k}$ from ω_k to ω_{k+1} .³

Stacking the unique unknown $\lambda'_{i,j,k}$'s (6 per frequency) into one large vector λ' , we write (19) as;

$$C(\alpha, \lambda) = \|\hat{\mu}'(\alpha) - \lambda'\|_{R_{\bar{s}}}^2 + \beta_1 \|L_{OD}\lambda'\|_2^2 + \beta_2 \|L_D\lambda'\|_2^2. \quad (20)$$

where $\hat{\mu}'$ is the vector of unique elements from $R(\alpha)\hat{M}_k R^T(\alpha)$ over all k . The L_i are used to regularize the problem; specifically, L_{OD} is for off-diagonal elements, and L_D is for diagonal elements. They are built such that

$$\|L_D\lambda'\|_2^2 = \sum_{k=1}^M \sum_{p=1}^3 (\lambda'_{pp}(w_k) - \lambda'_{pp}(w_{k+1}))^2 \quad (21)$$

$$\|L_{OD}\lambda'\|_2^2 = \sum_{k=1}^M (\lambda'_{12}(w_k))^2 + (\lambda'_{13}(w_k))^2 + (\lambda'_{23}(w_k))^2. \quad (22)$$

The regularization parameters β_i in (20) are used to determine the tradeoff in the reconstruction between the two terms in the cost function. The first term enforces fidelity to the data while the second ensures smooth spectra in (21). We note that in general, the on-line determination of β_i is a well-studied, non-trivial issue beyond the scope of this paper.⁴⁻⁶ For simplicity, in the examples in Section 5, we assume that β is known.

To minimize the cost function, we note first that because (21) is quadratic with respect to λ , $\hat{\lambda}$ can be explicitly stated in terms of α and \hat{r}_0 via

$$\hat{\lambda} = \left(R_s^{-1} + \beta_1 L_{OD}^T L_{OD} + \beta_2 L_D^T L_D \right)^{-1} R_s^{-1} \hat{\mu}' \equiv Q(\hat{r}_0, \alpha) \hat{\mu}' \quad (23)$$

so that we can write:

$$\hat{\alpha} = \arg \min_{\alpha} C(\hat{r}_0, \alpha, Q(\hat{r}_0, \alpha) \hat{\mu}') \quad (24)$$

$$\hat{\lambda} = Q(\hat{r}_0, \hat{\alpha}) \hat{\mu}' \quad (25)$$

In our experiments we have found that C exhibits many local minima in terms of the orientation angles. Thus, we have adopted the following strategy: We first impose a coarse grid on the three dimensional space of all permitted orientation angles, then, for each α -value in the grid, the value of the cost function C is found. We use that α values with the smallest overall cost for that cell to initialize a full 3D non-linear least squares scheme to find the final values of $\hat{\alpha}$. Using these values, we construct $\hat{\lambda}$ according to (25).

5. EXAMPLES

In this section, we compare our processing approach to that of a baseline method in which clutter mitigation is performed by subtracting from the interior data at a fixed y the average of the two corresponding horizontal samples taken on the boundary. We demonstrate and analyze the performance of the parameter estimation of our approach under two mine shapes. We simulate data taken on a 9×9 grid of 81 cm^2 pixels by a monostatic transmit/receive system comprised of square coils 5 cm on a side. Ten frequencies logarithmically spaced between 0 and 4.3 kHz are used. One corner of the grid is taken to be $(-0.4, -0.4) \text{ m}$ while the opposite is at $(0.4, 0.4) \text{ m}$.

As a first example, we consider a spherical object located at $(x_0, y_0, z_0) = (0, 0, .10) \text{ m}^*$, and with radius 5 cm . The medium as well as the object are taken to be non-ferrous and the conductivity of the sphere is 10^6 S/m . We assume that the sphere's response can be modeled as a dipole and we use the results of [2] to compute the dipole moment spectrum (DMS). The real and imaginary parts of this spectrum are shown as a solid line in Fig. 3. Because the sphere is rotationally invariant, for this problem there is no need to estimate the rotation angles so that the problem here reduces to determining the location and the DMS. To demonstrate the performance of our approach, we perform 100 Monte Carlo simulations at a signal to clutter plus noise ratio (SCNR) of 10 dB . In our model, the sample mean of the estimated object center is $(-0.0004, -0.0009, 0.1017) \text{ m}$ with a standard deviation of $\pm(0.0003, 0.0010, 0.0021) \text{ m}$. In the baseline method, it is $(0.0237, -0.0059, 0.1144) \text{ m}$ with a standard deviation of $\pm(0.0156, 0.0035, 0.0123) \text{ m}$. In Fig. 3 the dotted lines show the sample mean of the estimated DMS according to both methods. We see from these results that our approach is highly accurate and better when compared with the baseline method both in terms of estimating the position as well as the moment spectrum.

As a second example, we consider an ellipsoid object again located at $(x_0, y_0, z_0) = (0, 0, .10) \text{ m}$ and which has been rotated using $\phi = 1.70$ radians, $\psi = 1.70$ radians and $\theta = 1.70$ radians. In this case, we

*Increasing depth here corresponds to increasing z

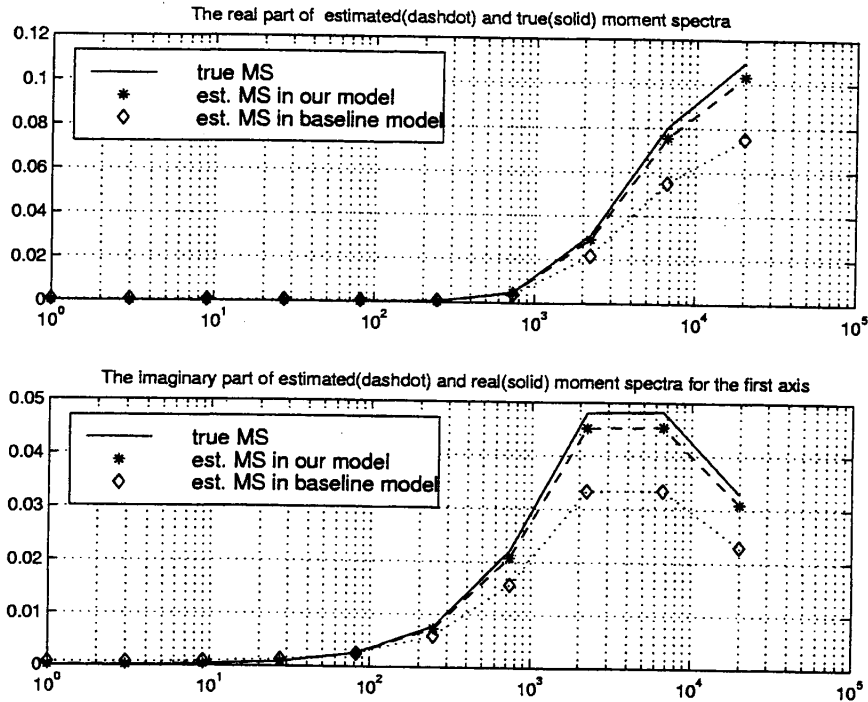


Figure 3. The real and imaginary part of estimated and real moment spectra of sphere mine

presently have no closed form expression for the frequency dependent DMS of such an object. However, under the assumption that the scattering characteristics of an eccentric object will be substantially different among axes, we hypothesizes DMS spectra shown in Fig. 4, Fig. 5, and Fig. 6 as solid lines and examine the performance of our approach under these conditions.

In this case, we estimate the center of object, moment spectra for all three axes and three rotation angles. After performing 100 Monte Carlo simulations at 10 dB SCNR, for our approach the sample mean of the estimated object center is $(-0.0005, 0.0001, 0.0985)m$ with a standard deviation of $\pm(0.0004, 0.0003, 0.0012)m$. For the baseline case the mean is $(0.0578, -0.0011, 0.1107)m$ with a standard deviation of $\pm(0.0562, 0.0025, 0.0129)m$. The sample means of the estimated rotation angles under our processing scheme are $\phi = 1.66751$ radians, $\psi = 1.7163$ radians, and $\theta = 1.9097$ radians with standard deviations of $\pm(0.0041)$, $\pm(0.0014)$, and $\pm(0.0073)$, respectively. Using the simpler clutter mitigation algorithm, the results are means of $\phi = 1.5050$ radians, $\psi = 1.2052$ radians, and $\theta = 2.7402$ radians with standard deviations of $\pm(0.072)$, $\pm(0.0371)$, and $\pm(0.0053)$, respectively. Finally, the real and imaginary parts of the estimated and actual DMS using both clutter reduction methods are shown in Fig. 4–6.

For completeness we have gathered the sample means and standard deviations of our parameter estimates for both objects and both processing methods in the following table. These results along with those seen in Fig. 4–6 clearly point to the utility of taking a rigorous processing approach to clutter modeling and mitigation.

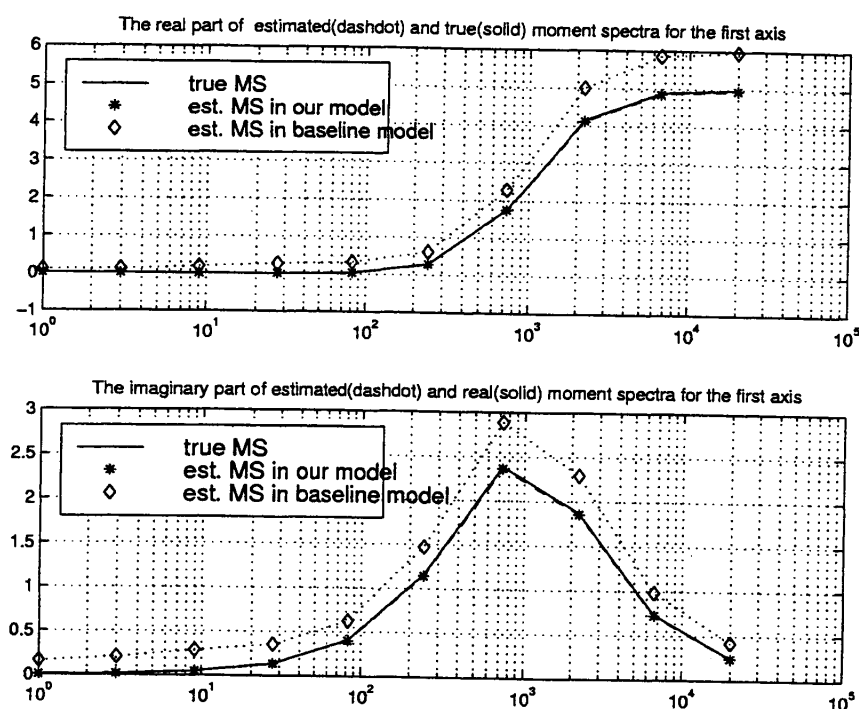


Figure 4. The real and imaginary part of estimated and real moment spectra of spheroid mine for the first axis.

	Sphere Case				Ellipsoid Case			
	Our Model		Baseline Model		Our Model		Baseline Model	
	mean	Standard Deviation	mean	Standard Deviation	mean	Standard Deviation	mean	Standard Deviation
\hat{x}_0	-0.0004	0.0003	0.0237	0.0156	-0.0005	0.0004	0.0578	0.0562
\hat{y}_0	-0.0009	0.0010	-0.0059	0.0035	0.0001	0.0003	-0.0011	0.0025
\hat{z}_0	0.1017	0.0021	0.1144	0.0123	0.0985	0.0012	0.1107	0.0129
$\hat{\phi}$	-	-	-	-	1.6675	0.0044	1.5050	0.0072
$\hat{\psi}$	-	-	-	-	1.7163	0.0014	1.2052	0.0371
$\hat{\theta}$	-	-	-	-	1.9097	0.0073	1.7402	0.0053

6. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented an approach for the estimation of the dipole moment spectra, the coordinates of the object center, and rotation angles from BEMI data after estimating and removing clutter from the signal. Our work has been aimed specifically to the problem of characterizing low metal content objects. Using simulated data, we have demonstrated some clear performance gains using our method relative to simpler clutter mitigation approaches.

In the future, we plan to extend this work in a number of ways. First, we intend on validating and modifying (as needed) these models and processing methods based on the use of real BEMI sensor data collected for example using a GEM 3 sensor. On the theoretical side, our interests are in the development of performance analysis metrics both for the estimation of the clutter parameters as well as the object characteristics. Specifically, the Cramer-Rao lower bound (CRLB) is a useful statistical tool for this effort.

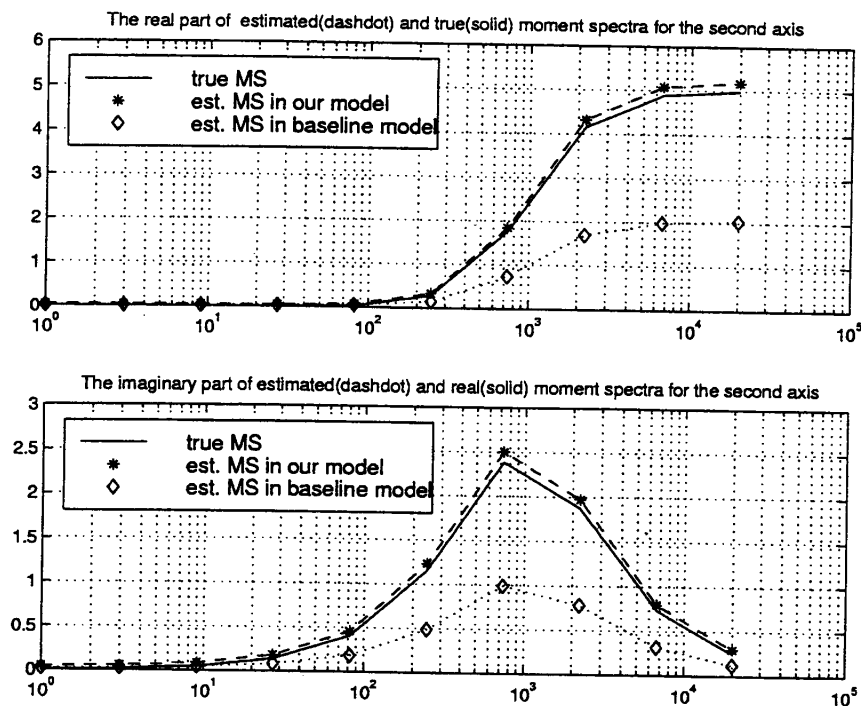


Figure 5. The real and imaginary part of estimated and real moment spectra of spheroid mine for the second axis.

The information provided by the CRLB can be used for example to optimize sensor configuration for a particular characterization task such as maximally distinguishing object X from Y.

REFERENCES

1. Y. Das, J. E. McFee, J. Toews, and G. C. Stuart, "Analysis of an electromagnetic induction detector for real-time localization of buried objects," *IEEE Trans. Geoscience and Remote Sensing* **28**, pp. 278-287, May 1990.
2. I. Won, D. Keiswetter, and E. Novikova, "Electromagnetic induction spectroscopy," *JEEG* **3**, pp. 27-40, March 1998.
3. M. Ozdemir, E. L. Miller, and S. Norton "Localization and characterization of buried objects from multi-frequency, array inductive data," in *SPIE'99 AeroSense Symposium, -Detection Technologies for Mines and Minelike Targets IV, Orlando FL*, April 1998.
4. P. C. Hansen, "The use of l-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.* **14**, pp. 1487-1503, 1993.
5. P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Review* **34**, pp. 561-580, December 1992.
6. M. Belge, M. E. Kilmer, and E. L. Miller, "Simultaneous multiple regularization parameter selection by means of the l-hypersurface with applications to linear inverse problems posed in the wavelet domain," in *Proceedings of SPIE'98-Bayesian inference for inverse problems*, vol. 3459, July 1998.
7. S. Hassani, "Foundations of Mathematical Physics," *Boston: Allyn and Bacon*, pp. 31-35, 1991.

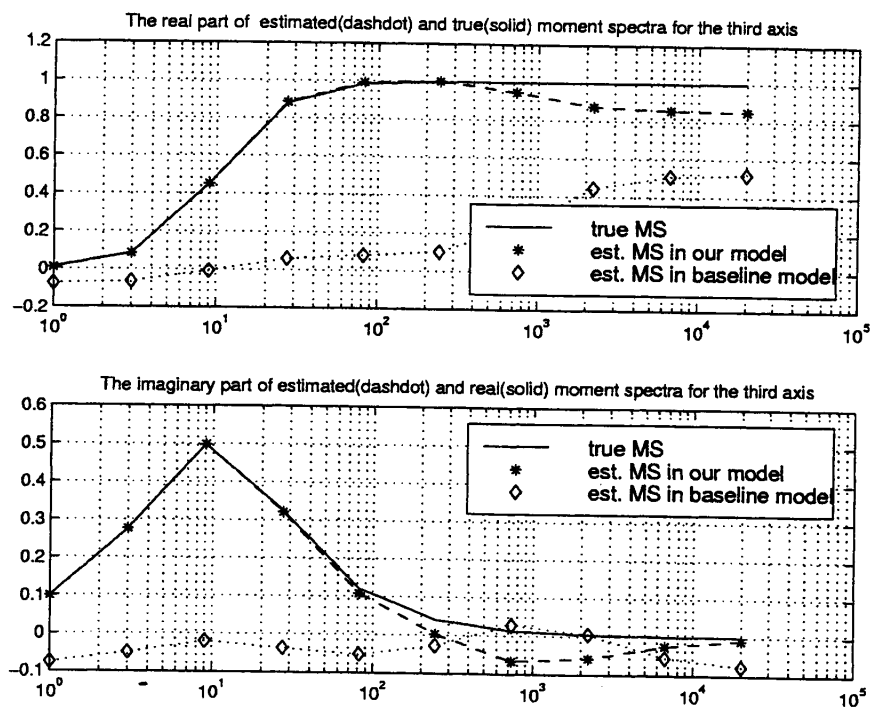


Figure 6. The real and imaginary part of estimated and real moment spectra of spheroid mine for the third axis.

Optimum PML ABC Conductivity Profile in FDFD

Edwin A. Marengo, Carey M. Rappaport and Eric L. Miller

Center for Electromagnetics Research, Northeastern University, Boston, MA 02115

Abstract—We address the problem of optimally choosing the conductivity (σ) profile of a perfectly-matched-layer absorbing boundary condition (PML ABC) with prescribed number of layers so as to minimize reflections for a wide range of incidence angles and for a narrow (CW) or broad frequency-band. A new one-dimensional (1-D), frequency-domain description of 2-D PML performance is developed, validated and used in PML- σ profile optimization. An exhaustive search for PML- σ profiles that minimize reflections over a prescribed wide angle-range is carried out. Our procedure yields PML- σ profiles with better performance than previously reported values, for given number of layers.

Index Terms—Perfectly-matched-layer (PML).

I. INTRODUCTION

The optimal specification of Berenger's perfectly-matched-layer absorbing boundary condition (PML ABC) [1] in finite-difference frequency-domain (FDFD) and finite-difference time-domain (FDTD) EM-field computations is an open question of much interest to the computational electromagnetics community [2]. The challenge translates into profiling the PML conductivity σ so as to minimize discretization and material-contrast reflections (predominant at small incidence angles θ and high frequencies ω) while also allowing strong decay at grazing angles of incidence (say, $\theta > 65^\circ$). In this paper, two-dimensional (2-D) PML performance is modeled using a new 1-D PML formulation which is later used in PML- σ profile optimization. While past efforts [3]–[5], with the exception of Ref.[2], were aimed mostly toward optimizing normal-incidence performance, our focus is on wide-angle and narrow- and/or broad-bandwidth performance.

Our goal is to find conductivity profiles of the form $\sigma_i = \sigma_f \left(\frac{2N-i+1}{2N} \right)^p$, where σ_f and p are parameters that need to be chosen so as to minimize reflections from a PML with N layers and grid-spacing Δ with $i = 1, 2, \dots, 2N$ indicating twice the *half-layer* index in Yee's lattice [6] (with $i = 2N$ at the free space/PML interface). (We associate distinct values of σ to each electric and magnetic field point of the PML, doubling the number of parameters from the original [1].) We consider separately the cases of CW and broad-band operation: In the former, σ -profiles that minimize reflections for a specified incidence angle-range at prescribed, stepped frequencies are computed, which can be used in FDFD; the latter yields a PML that is (globally) optimum over a wide incidence angle-range and a specified broad-bandwidth; hence, it applies to FDTD simulation of broad-band EM pulses. We tackle

both cases numerically by minimizing a cost function, say $F = \int d\omega W(\omega) \int_0^{\pi/2} d\theta \Theta(\theta) \log_{10}[1 + |\Gamma(\theta, \omega)|]$, where $|\Gamma|$ is the PML's reflection coefficient magnitude while $\Theta(\theta)$ and $W(\omega)$ are weighting functions defining, respectively, the relevant ranges of incidence angles θ and frequencies ω . For example, later we will consider hard-weighting functions $\Theta(\theta)$, $W(\omega)$ where $\Theta = 1$ over a prescribed angular range $[0, \theta_{\max}]$ (where, e.g., $\theta_{\max} = 65$ or 75°) and is zero elsewhere, and $W(\omega) = 1$ only within the bandwidth of interest and is zero elsewhere.

The technique developed in the paper is validated with computer-simulated results both in FDFD and in FDTD. We find PML- σ profiles with better performance than previously reported values, for given number of layers. Attention is restricted to the 2-D transverse magnetic (TM) configuration (relative to the PML¹) schematically depicted in Fig. 1. The associated TE results follow by duality.

II. FORMULATION AND VALIDATION

Time-harmonic TM waves inside a PML perpendicular to the x -axis satisfy the equations [7]

$$\begin{aligned} \frac{\partial}{\partial y} \mathcal{H}_z &= j\omega\epsilon_0 \mathcal{E}_x \\ \frac{\partial}{\partial x} \mathcal{H}_z \left(\frac{1}{\epsilon' - j\sigma/\omega\epsilon_0} \right) &= -j\omega\epsilon_0 \mathcal{E}_y \\ \frac{\partial}{\partial x} \mathcal{E}_y \left(\frac{1}{\epsilon' - j\sigma/\omega\epsilon_0} \right) - \frac{\partial}{\partial y} \mathcal{E}_x &= -j\omega\mu_0 \mathcal{H}_z \end{aligned} \quad (1)$$

where ϵ' , considered here to be $\epsilon' = 1$, is the PML dielectric constant. For a y -independent PML configuration such as that in Fig. 1 we obtain, for propagating wave-mode solutions of Eqs.(1),

$$\begin{aligned} \frac{d}{dx} E_y \left(\frac{1}{1 - j\sigma/\omega\epsilon_0} \right) &= -j\omega\mu_0 \cos^2 \theta H_z \\ \frac{d}{dx} H_z \left(\frac{1}{1 - j\sigma/\omega\epsilon_0} \right) &= -j\omega\epsilon_0 E_y \end{aligned} \quad (2)$$

where E_y and H_z are the x -dependent parts of the electric and magnetic field components \mathcal{E}_y and \mathcal{H}_z , respectively, and θ is the propagation angle associated with the propagating wave-mode considered. Next we finite-difference Eqs.(2) while adopting the notation in Fig. 1, obtaining

$$\begin{aligned} E_y^{n+1/2} &= E_y^{n-1/2} + (j\omega + \sigma_{2n}/\epsilon_0)\mu_0\Delta \cos^2 \theta H_z^n \\ H_z^n &= H_z^{n-1} + (j\omega + \sigma_{2n-1}/\epsilon_0)\epsilon_0\Delta E_y^{n-1/2} \end{aligned} \quad (3)$$

¹The configuration in Fig. 1 is TE according to the standard convention wherein the z -axis in Fig. 1 is the reference axis (since 2-D objects associated with Fig. 1 are z -independent). We refer to it as 'TM' to avoid confusion with the mode notation of a parallel plate waveguide configuration to be used later on.

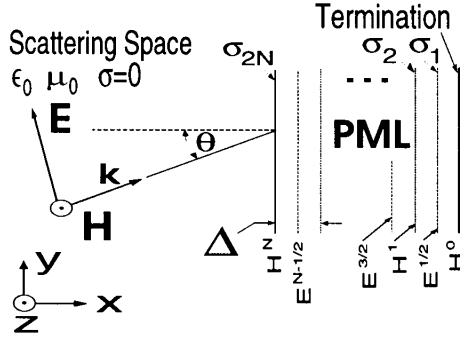


Fig. 1. PML and incident TM wave in 2-D.

for $n = 1, 2, \dots, N$. Any PML analysis based on Eqs.(2,3) can only approximate the numerically-measured value of Γ since Eq.(2) assumes waves with y -dependence $e^{\pm jk \sin \theta y}$ where $k = \omega\sqrt{\mu_0\epsilon_0}$, which are only approximated in the 2-D finite-difference (FD) solution. In particular, Eqs.(3) yield an *approximate* (yet useful) expression for the PML's reflection coefficient Γ , to be discussed next. Numerical FDFD and FDTD experiments are discussed at the end of this section which confirm our model's validity for the purposes of the present study.

By defining the normalized transverse impedance $\eta_n \equiv (E_y^{n-1/2}/H_z^{n-1})/\eta_o$ of one PML layer relative to that of the subsequent layer, where $\eta_o = \sqrt{\mu_0/\epsilon_0}$, one obtains from Eqs.(3)

$$\eta_{n+1} = (j\bar{k} + S_{2n}) \cos^2 \theta + \frac{1}{j\bar{k} + S_{2n-1} + 1/\eta_n} \quad (4)$$

where we have introduced the unitless quantities $\bar{k} \equiv k\Delta$ (the normalized phase constant) with $k = \omega\sqrt{\mu_0\epsilon_0}$ and $S_i \equiv \sigma_i\eta_o\Delta$ (the normalized i^{th} half-layer decay rate). Unless otherwise stated, we will consider perfect magnetic conductor (PMC) loading, $H_z^0 = 0$, so that $\eta_1 = \infty$. In order to obtain an approximate expression for Γ , we compute next the *numerical* transverse free space impedance Z_o defined, following Eq.(4), via

$$Z_o = j\bar{k} \cos^2 \theta + \frac{1}{j\bar{k} + 1/Z_o}, \quad (5)$$

i.e.,

$$Z_o = \frac{2}{-j\bar{k} + \sqrt{-\bar{k}^2 + 4(\cos \theta)^2}} = \cos \theta e^{j \sin^{-1}(\bar{k} \cos \theta/2)} \quad (6)$$

(Z_o is normalized relative to η_o). Termination of a free space FD grid with Z_o (e.g., $\eta_1 = (E_{1/2}/H_0)/\eta_o = Z_o$ in the equations above, with $S_i = 0, i = 1, 2, \dots$) results in a *measured* standing-wave-ratio SWR = 1. Thus, Z_o is the free space impedance in FDFD (again, as described by the model FD equations (3)).

We can now use Eqs.(4,6) to approximate Γ by

$$\Gamma \simeq \frac{\eta_{N+1} - Z_o}{\eta_{N+1} + Z_o}. \quad (7)$$

Z_o behaves like $\cos \theta e^{j\bar{k} \cos \theta/2}$ at low frequencies ($\bar{k} \ll 1$) but its phase differs from the theoretical value for a continuum, $\bar{k} \cos \theta/2$, by about 2% for $\bar{k} = 2\pi/10$ at $\theta = 0$. The additional phase is a consequence of dispersive effects of discretization: It was shown in [8] (see Eq.(2.19), pp. 24) that for a 1-D free space FD grid

$$\omega\sqrt{\mu_0\epsilon_0}\Delta/2 = \sin(\kappa\Delta/2) \quad (8)$$

where $\kappa (\neq k = \omega\sqrt{\mu_0\epsilon_0})$ is the wavenumber in the FD model (i.e., Eq.(8) is the discretization-corrected dispersion relation). Eq.(8) applies to our framework with the substitution $\bar{k} \rightarrow \bar{k} \cos \theta$ (and $\bar{\kappa} \rightarrow \bar{\kappa} \cos \theta$) and thus the discretization-corrected phase-shift between adjacent E- and H-field points in one-way wave-propagation in free space is $\sin^{-1}(\bar{k} \cos \theta/2)$, which is the result in (6).

To examine the validity of the model above we used both FDFD and Fourier-transformed FDTD data. Fig. 2 shows plots of $|\Gamma|$ versus θ and parameterized by various values of \bar{k} for an 8-layer PML proposed in [2]: $\sigma_f = 0.018/\Delta$ and $p = 3.7$. The figure shows data points for $|\Gamma|$ as measured in (2-D) FDFD (using the measured SWR) for a parallel plate waveguide excited by single TM modes and terminated by an x -normal PML with PEC-wall termination. The TM mode corresponding to $\theta = 0$ is the TEM mode. For given TM mode, θ and ω , the transverse waveguide dimensions are given by well known formulas (see [9], pp. 287-290) and can be used to compute other operational parameters [9]. Also shown are values of $|\Gamma|$ corresponding to Fourier-transformed FDTD data. In contrast to the FDFD results – all of which involve a parallel plate waveguide grid – our FDTD simulations correspond to transient plane-wave propagation in a (2-D) PML-terminated free space grid (time-gating was used to isolate the PML reflections). In particular, our FDTD simulations were used to measure the PML response to an incident, wide-band Gaussian pulse of the form $e^{(x-ct)^2/W^2}$ with $c = 1/\sqrt{\mu_0\epsilon_0}$, $W = 6\Delta$ and Courant number $r = c\Delta t/\Delta = 0.08$ (see also [2]). In addition, Fig. 2 provides data points for the reflection coefficient as measured in the time-domain (FDTD) by the ratio of the peak amplitudes of the reflected and incident wave-pulses. The results based on Eqs.(4,6,7) (solid lines labeled $\bar{k} = 2\pi/10$ and $\bar{k} = 2\pi/100$) are seen to follow closely those of FDFD and Fourier-transformed FDTD simulations (the FDFD $|\Gamma|$ -values for $\bar{k} = 2\pi/25$ also agree with Eqs.(4,6,7) (results not shown for clarity)). Those associated with the continuum analog of formula (7) (curves marked with a triangle) differ visibly from the FDFD and Fourier-transformed FDTD results for non-grazing incidence and high \bar{k} (they behave well only for low \bar{k} and high θ (values for $\bar{k} = 2\pi/100$ marked with a small +), as expected (inter-grid phase-shift is then negligible)). Other profiles were studied in a similar fashion and analogous results were obtained where model predictions based on Eqs.(4,6,7) were consistent with FDTD and FDFD reflection coefficient measurements. We may now proceed with confidence to use Eqs.(4,6,7) as design tools in the search for optimum σ coefficients.

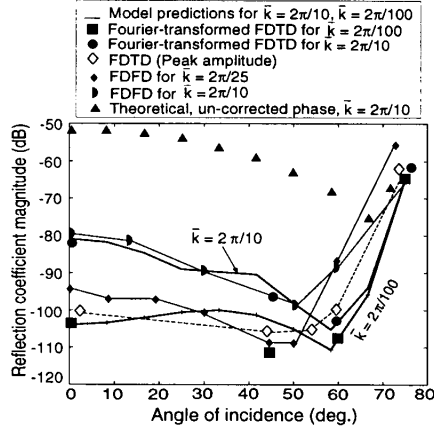


Fig. 2. Reflection coefficient results for PML- σ profile in [2].

III. RESULTS AND DISCUSSION

Using the PML-model in Sec. II, an automated search for optimum values of σ_f and p for an N -layer PML profiled via $\sigma_i = \sigma_f \left(\frac{2N-i+1}{2N} \right)^p$, $i = 1, 2, \dots, 2N$ ($i = 2N$ for the half-layer at the free space/PML interface) was carried out both for fixed \bar{k} (CW case) and for $2\pi/100 \leq \bar{k} \leq 2\pi/10$ (broadband case). We considered the cost function $F = \int_{\text{BW}} d\omega \int_0^{\theta_{\max}} d\theta \log_{10} [1 + |\Gamma(\theta, \omega)|]$ defined in Sec. I, where BW is the relevant bandwidth, as well as the L^2 norm of $|\Gamma(\theta, \omega)|$, $R = \int_{\text{BW}} d\omega \int_0^{\theta_{\max}} d\theta |\Gamma(\theta, \omega)|^2$, i.e., the energy reflection coefficient for bandwidth BW averaged over the relevant θ -range, where, e.g., $\theta_{\max} = 65$ or 75° . The later characterization applies to both frequency and time domains. It turns out that for low $|\Gamma|$ (e.g., in the vicinity of the optimized solution) $F \approx \int_{\text{BW}} d\omega \int_0^{\theta_{\max}} d\theta |\Gamma(\theta, \omega)|$ (the L^1 or area norm); it follows that (σ_f, p) -searches based on F and R should yield similar results.

Table I summarizes the results of the search for an 8-layer PML, obtained using both F and R and for $\theta_{\max} = 65$ and 75° . The (σ_f, p) -values in Table I correspond to absolute minima of F and R in $10^{-3} \leq \sigma_f \leq 10^3$ and $0 \leq p \leq 7$ (values obtained via exhaustive numerical search). The search reveals the need for σ -profiles that are low enough to minimize discretization and material-contrast reflections at close-to-normal incidence while being also strong enough to absorb waves propagating at close-to-grazing angles. For example, we see that the best σ -profiles for $\theta_{\max} = 65^\circ$ are consistently lower than for $\theta_{\max} = 75^\circ$. This results in reduced reflections in the small θ region in the $\theta_{\max} = 65^\circ$ case. Besides, the $\theta_{\max} = 75^\circ$ case demands higher σ to enhance wave attenuation in the large θ region $65^\circ \leq \theta \leq 75^\circ$. The results of broad-band optimization are in between those of low- and high-frequencies, as expected. Remarkably, the broad-band case optimization with $\theta_{\max} = 75^\circ$ yielded a σ -profile very similar to that reported in [2] and obtained from wide-band and wide propagation-angle FDTD simulations (the same profile used in Sec. II, i.e., $\sigma_f = 0.018/\Delta$ and $p = 3.7$). We also note that there is little variation among the optimum (σ_f, p) -values corresponding to various values (ranges) of \bar{k} due to the broadband

nature of the PML. Furthermore, we have found by numerical experiments that $|\Gamma|$ varies little with \bar{k} for $\bar{k} \lesssim 2\pi/50$; hence, our results for $\bar{k} = 2\pi/100$ represent well the PML's low frequency behavior.

TABLE I
SUMMARY OF OPTIMIZATION RESULTS

Cost Function	\bar{k}	θ_{\max}	$\sigma_f \Delta$	p
F	$2\pi/10$	75°	0.019	3.49
R	$2\pi/10$	75°	0.018	3.51
F	$2\pi/100$	75°	0.026	3.71
R	$2\pi/100$	75°	0.026	3.72
F	$2\pi/10 - 2\pi/100$	75°	0.022	3.65
R	$2\pi/10 - 2\pi/100$	75°	0.021	3.65
F	$2\pi/10$	65°	0.014	3.47
F	$2\pi/100$	65°	0.020	3.82
F	$2\pi/10 - 2\pi/100$	65°	0.016	3.73

Fig. 3 shows plots of $|\Gamma|$ vs θ and parameterized by various values of \bar{k} for an 8-layer PML with $\sigma_f = 0.019/\Delta$ and $p = 3.49$ (optimum coefficients for $\bar{k} = 2\pi/10$ and $\theta_{\max} = 75^\circ$). Fig. 4 shows analogous results for $\sigma_f = 0.026/\Delta$ and $p = 3.71$ (optimum coefficients for $\bar{k} = 2\pi/100$ and $\theta_{\max} = 75^\circ$). The values of $|\Gamma|$ in the figures were calculated from the measured SWR in FDTD simulation of a parallel plate waveguide excited by single TM modes and terminated by an x -normal PML with PEC-wall termination (the same configuration outlined in connection with Fig. 2). The broadband time-domain performance of $\sigma_f = 0.022/\Delta$ and $p = 3.65$ (results not shown) is similar to that reported in [2] for $\sigma_f = 0.018/\Delta$ and $p = 3.7$. Figs. 3 and 4 show that, for fixed normalized decay rate profile $S_i = \sigma_i \eta_0 \Delta$ and as function of \bar{k} alone, PML performance tends to degrade with increasing \bar{k} as a consequence of discretization (cf. Eq.(4) and associated discussion). The PML performance at grazing incidence is, on the other hand, less sensitive to discretization (since $\bar{k} \cos \theta \rightarrow 0$ as $\theta \rightarrow \pi/2$); the latter effect is also apparent in Figs. 3 and 4. We see that large-angle PML performance improves noticeably for low-frequency optimal profiles (compare results in Fig. 4 to those in Fig. 3) due to reduced optimization-constraints deriving from discretization effects at small θ . When compared to its high-frequency counterpart, the low-frequency-optimized PML design has, thus, higher conductivity, which improves large θ decay rate. For the same reason, however, the $|\Gamma|$ -values for $\bar{k} = 2\pi/25$ at $\theta < 50$ are visibly worse in Fig. 4 than in Fig. 3 (the low-frequency design is too aggressive). The results shown in Fig. 3 are comparable but slightly better than those for $\sigma_f = 0.018/\Delta$ and $p = 3.7$ (see Fig. 2) at the same \bar{k} . Finally, the best PML performance in Fig. 4 (at $\bar{k} = 2\pi/100$) is somewhat better than the best PML performance in Fig. 3 (at $\bar{k} = 2\pi/25$).

Fig. 5 summarizes optimization results for other values of N (in the 0-to- 75° angular range). Similar (σ_f, p) -behavior with N was reported in [5] for normal-incidence PML performance. For 16 layers, we have found that, for the optimum σ -profiles shown, $|\Gamma|$ ranges from -150 to -180 dB for $\bar{k} = 2\pi/100$ and from -130 to -160 dB for $\bar{k} = 2\pi/10$

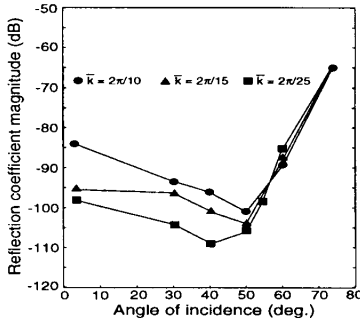


Fig. 3. $|\Gamma|$ vs θ for $\sigma_f \Delta = 0.019$ and $p = 3.49$.

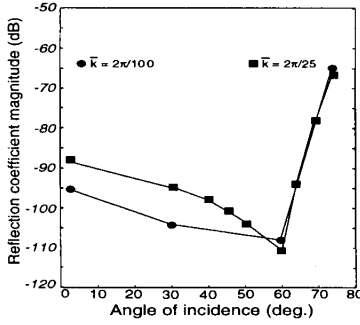


Fig. 4. $|\Gamma|$ vs θ for $\sigma_f \Delta = 0.026$ and $p = 3.71$.

and the range $\bar{k} = 2\pi/10$ - $2\pi/100$. The figures above correspond to the 0-to-75° angular range.

Finally, the same procedure can be applied to σ -profiles of other forms. For example, next we used the PML-model and performance measures above to search for PML σ -profiles of the form $\sigma_i = \sigma_f^a \left(\frac{2N-i+1}{2N}\right)^{p^a} + \sigma_f^b \left(\frac{2N-i+1}{2N}\right)^{p^b}$. An extensive numerical search based on the cost function F for $N = 8$, $\bar{k} = 2\pi/100$ and $\theta_{\max} = 75^\circ$ with integer p^a and p^b yielded $p^a = 2$, $\sigma_f^a \Delta = 0.0018$, $p^b = 4$ and $\sigma_f^b \Delta = 0.0265$. Plots of $|\Gamma|$ vs θ corresponding to this σ -profile for $\bar{k} = 2\pi/25$ and $\bar{k} = 2\pi/100$ are shown in Fig. 6. The performance of this PML is a little better than that obtained with two search-parameters (i.e., σ_f, p) (compare results in Fig. 6 with those of Fig. 4, both of which correspond to optimization at $\bar{k} = 2\pi/100$). For $\bar{k} = 2\pi/100$, we note improvement of about 10 dB at $\theta = 60^\circ$ and of 2 dB at $\theta = 75^\circ$ with respect to the profile in Fig. 4. On the other hand, for $\bar{k} = 2\pi/25$, $|\Gamma|$ -values for $\theta \leq 30^\circ$ in Fig. 6 are 1 dB lower than in Fig. 4 while for $\theta > 30^\circ$, $|\Gamma|$ -values in Fig. 6 are 2 or more dB below those in Fig. 4. Thus, it appears that there is a limit to PML performance near our optimized solutions (for the number of PML layers considered, $N = 8$). It is expected that for larger N , the increase in number of σ -profile parameters will have greater impact. This subject will be addressed elsewhere.

Despite its simplicity (relative to PML descriptions including 2- and 3-D discrete dispersion [8]), our 1-D based PML performance predictor (i.e., Eqs.(4,6,7)) has been found to work well in connection with 2-D FDFD and FDTD simulations, yielding a powerful tool for PML optimization. The sought-after PML- σ profiles were found

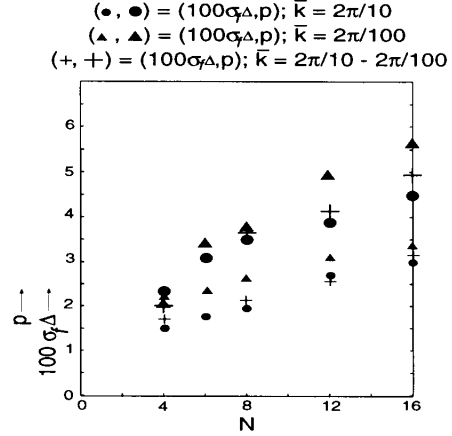


Fig. 5. $\sigma_f \Delta$ and p versus N for $\bar{k} = 2\pi/10, 2\pi/100$ and the range $2\pi/10$ - $2\pi/100$.

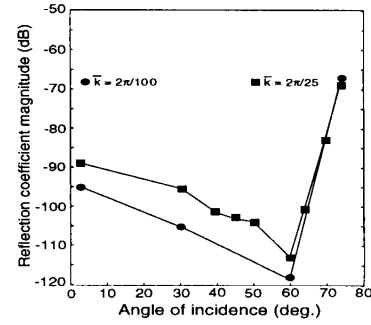


Fig. 6. $|\Gamma|$ vs θ for $\sigma_f^a \Delta = 0.0018$, $p^a = 2$, $\sigma_f^b \Delta = 0.0265$ and $p^b = 4$.

to exhibit better CW and/or broad-band and wide-angle performance than the best reported values to date [2]. The optimal σ -profiles reported in the paper were found to be consistent with optimal σ -profile recipes proposed in [3] and [5] for the normal-incidence case. As expected, our PML designs were, however, more aggressive (larger σ) in order to accommodate for large-angle performance which was not addressed before.

REFERENCES

- [1] J.P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves", *J. Comp. Phys.*, vol. 114, pp. 185-200, 1994.
- [2] S.C. Winton and C.M. Rappaport, "Profiling the perfectly matched layer to improve large angle performance", *IEEE Trans. Antenn. Propagat.*, In Review.
- [3] S. Gedney, "An anisotropic perfectly matched layer-absorbing medium for the truncation of FDTD lattices", *IEEE Trans. Antenn. Propagat.*, vol. 114, pp. 1630-1639, 1996.
- [4] D. Katz, E. Thiele and A. Taflov, "Validation and extension to three dimensions of the Berenger PML absorbing boundary", *IEEE Guided Wave Lett.*, vol. 4, pp. 268-270, 1994.
- [5] Z. Wu and J. Fang, "Numerical implementation and performance of perfectly matched layer boundary condition for waveguide structures", *IEEE Trans. Micro. Theo. Tech.*, vol. 43, pp. 2676-2683, 1995.
- [6] K.S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media", *IEEE Trans. Antenn. Propagat.*, vol. AP-14, pp. 302-307, 1966.
- [7] C.M. Rappaport, "Interpreting and improving the PML absorbing boundary condition using anisotropic lossy mapping of space", *IEEE Trans. Magn.*, vol. 32, pp. 968-974, 1996.
- [8] C. Christopoulos, *The transmission-line modeling method*, IEEE Press, USA, 1995.
- [9] D.H. Staelin, A.W. Morgenthaler and J.A. Kong, *Electromagnetic waves*, Prentice-Hall, USA, 1994.

IMPLEMENTATION OF A TWO DIMENSIONAL PLANE WAVE FDTD USING ONE DIMENSIONAL FDTD ON THE LATTICE EDGES

Scott C. Winton and Carey M. Rappaport
Center for Electromagnetics Research
Northeastern University
Boston, MA 02115

Abstract Development and testing of angle independent absorbing boundary conditions (ABCs) can be improved by simulating waves incident on the ABC at a single angle. By using one-dimensional Finite Difference Time Domain (FDTD) as the lattice side edge condition, the creation and numerical propagation of a two dimensional plane wave with arbitrary incident angle is possible. The application and extent of usefulness of the method are examined and extensions to increase the range of usefulness are introduced.

I. INTRODUCTION

Until recently, the use of the FDTD method for numerical solutions of electromagnetic scattering problems was severely hampered by the poor performance of absorbing boundary conditions used to prevent reflections of EM waves at the lattice edges. With the advent of the Berenger Perfectly Matched Layer[1], this problem has been significantly reduced and the computational efficiency of FDTD problems has been significantly improved.

Understandably, this has created considerable interest in improving and optimizing the PML and angle-independent ABCs as a whole. But this effort has been impaired by the fact that commonly used excitations, such as point or line sources, generate waves incident on the ABC at all angles. This complicates the analysis of the performance of the ABC and impairs design optimization. The problem is ameliorated by introducing plane wave sources incident on the ABC at a single angle. With such an excitation, the performance of the ABC is clearly defined. However, because of the difficulty in dealing with propagation on the lattice edge, the creating and propagating such a wave is difficult.

For a two-dimensional FDTD simulation using the standard Yee cell formulation[2], the update of a given spatial grid point requires data from the four adjacent grid points. Clearly this creates a problem at the lattice edges. Typically, in scattering simulations, a Mur total/scattered field region separation avoids the need to calculate incident waves on the edges[3]. However, to test ABCs, the incident wave must be a uniform plane wave without deformations along the edge. ABCs cannot be used at these edges since they fail for plane waves propagating at steep grazing angles. Furthermore, the values on the edges must not be specified analytically because the numerical values of the fields inside the discretized ABC under test are not known.

II. INTEGRATION OF ONE-DIMENSIONAL EDGE FDTD WITH A TWO-DIMENSIONAL GRID

For the following discussion, consider a two-dimensional grid on which a boundary value FDTD simulation will be run. The “front” of the grid is the source boundary value while the ABC to be tested is positioned at the “back”. What is desired is to create a plane wave with a phase front at an angle θ with respect to the front wall. For this discussion a transverse electric (TE) wave is considered. The x direction is front to back and the y direction is left to right as shown in Figure 1.

Instead of ABCs at the “left” and “right” edges, 1-D FDTD is used. The update of each spatial point on these edges will require data from only the grid points preceding and succeeding it. At each time step, information from this 1-D FDTD is passed to the larger 2-D FDTD to update spatial points adjacent to the left and right edges. If the wave is normally incident on the ABC, the formulation of the plane wave is quite simple. In this case, H_x is zero and both E_z and H_y are uniform left to right. Therefore, no information is obtained from the transverse difference and the calculations along each grid line running front to back reduces to as a 1-D FDTD algorithm. To illustrate this, consider the time harmonic Maxwell’s curl equations for TE waves in lossless media:

$$\frac{\partial E_z}{\partial y} = -j\omega\mu H_x \quad (1a)$$

$$\frac{\partial E_z}{\partial x} = j\omega\mu H_y \quad (1b)$$

$$\frac{\partial H_y}{\partial x} + \frac{\partial H_x}{\partial y} = j\omega\epsilon E_z \quad (1c)$$

These have the familiar (forward propagating) solutions:

$$E_z = E_o e^{-jk(x \cos \theta + y \sin \theta)} \quad (2a)$$

$$H_y = -\cos \theta \frac{E_o}{\eta} e^{-jk(x \cos \theta + y \sin \theta)} \quad (2b)$$

$$H_x = -\sin \theta \frac{E_o}{\eta} e^{-jk(x \cos \theta + y \sin \theta)} \quad (2c)$$

It is clear that for normal incidence, ($\theta = 0$), this becomes a TEM wave with no y dependence. Equations (1a) and (2c) become unnecessary and the 1-D FDTD and the 2-D FDTD calculations are identical.

The situation is more complicated when θ is nonzero. Now H_x is nonzero and the 1-D wave is no longer identical to the 2-D wave. In order for the 1-D FDTD simulation to supply the correct data to the 2-D grid, the 1-D wave must propagate with a velocity that keeps pace with the 2-D wave. This velocity is simply the phase velocity of the 2-D wave in the x direction, *i.e.* $v_{1D} = v_o / \cos \theta$, where v_o is the velocity of the 2-D wave in the direction θ . This is analogous to taking a slice along the right (or left) edge of the grid, of an infinite 2-D plane wave. Clearly this “slice” must travel along the edge with greater velocity than the the wave traveling an angle θ . Since the velocity, v_{1D} of this wave is given as $v_{1D} = \frac{\omega}{k_x}$, we can write Eq. (2) for the 1-D wave at $y = 0$ as:

$$E_{z_{1D}} = E_o e^{-jkx \cos \theta} \quad (3)$$

$$H_{y_{1D}} = -\cos \theta \frac{E_o}{\eta} e^{-jkx \cos \theta}$$

The solution in Eq.(3) does not satisfy Maxwell’s curl equations. This problem can be addressed by modifying Ampere’s Law. By taking the partial derivative of Eq.(2c) at $y = 0$ one obtains:

$$\frac{\partial H_x}{\partial y} = jk \sin^2 \theta \frac{E_o}{\eta} e^{-jkx \cos \theta},$$

which, using Eq.(2b) becomes

$$\frac{\partial H_x}{\partial y} = \frac{(1 - \cos^2 \theta) \frac{\partial H_y}{\partial x}}{\cos^2 \theta}.$$

Thus,

$$\frac{\partial H_y}{\partial x} + \frac{\partial H_x}{\partial y} = \frac{1}{\cos^2 \theta} \frac{\partial H_y}{\partial x}$$

Now Eq.(1), for the 1-D wave will become:

$$\begin{aligned} \frac{\partial E_{z1D}}{\partial x} &= j\omega\mu H_{y1D} \\ \frac{1}{\cos^2 \theta} \frac{\partial H_{y1D}}{\partial x} &= j\omega\epsilon E_{z1D} \end{aligned} \quad (4)$$

Having described the changes to Maxwell's equations needed to integrate the 1-D FDTD into the 2-D grid, it is worthwhile examining the discretization of the modified 1D curl equations.

III. DISCRETIZATION OF MODIFIED EQUATIONS AND STABILITY CONSIDERATIONS

Through a straight forward discretization process[4], the discretized 1-D modified Maxwell's curl equations for lossless media become;

$$\begin{aligned} E_i^{n+\frac{1}{2}} &= E_i^{n-\frac{1}{2}} + \frac{R}{\cos^2 \theta} \eta (H_{i+\frac{1}{2}}^n - H_{i-\frac{1}{2}}^n) \\ H_{i+\frac{1}{2}}^{n+1} &= H_{i+\frac{1}{2}}^n + \frac{R}{\eta} (E_{i+1}^{n+\frac{1}{2}} - E_i^{n+\frac{1}{2}}) \end{aligned} \quad (5)$$

where n is the time index, i is the space index and $R = v_o \Delta t / \Delta x$ is the Courant Number, and the vector component designations has been suppressed. Note that the velocity of the modified 1-D wave is $v_{1D} = \frac{v_o}{\cos \theta}$ as is apparent in the discretized wave equation based on Eq.(5).

Of particular importance is the Courant Number, R . Stability analysis indicates that for a FDTD simulation to be stable $R \leq 1$. For Equation (5), a new Courant number $R_{1D} = \frac{R}{\cos \theta}$ must be used instead. Since $R_{1D} \leq 1$, the usefulness of this method is limited to smaller angles. For example, let R_g be the Courant Number of the 2-D FDTD simulation. If R_g is chosen to be 0.5, then the largest angle that may be used is 60° . Clearly larger angles may be used if R_g is chosen to be smaller. However, since R_g is also a measure of how fast the wave moves through the grid, choosing it too small increases computational expense for the entire 2-D grid.

IV. LARGE ANGLE SOLUTIONS

For applicability with large propagation angles without decreasing R_g , changes must be made to the method. This is done by adjusting the Courant Number of the 1-D FDTD edge simulation. A smaller Courant number $R'_{1D} = v_o \Delta t' / \Delta x \cos \theta$ may be chosen such that $R'_{1D} \leq R_{1D}$. Since Δx remains the same, two simulations using R'_{1D} and R_{1D} would be spatially similar at the same physical time t , whenever $n' \Delta t' = n \Delta t$ for some different number of new time steps. The the stability condition is now $R'_{1D} \leq 1$ and thus θ may be increased.

Some care must be exercised in order to insure that the correct data is being passed to the 2-D grid. Let $m = \Delta t / \Delta t'$. If m is an integer ($= n' / n$), then the m^{th} iteration of the 1-D FDTD is used to update the 2-D interior of the grid. If m is not an integer, then the correct value of E_z on the

edge of the lattice must be interpolated with respect to time from two or more iterations of the 1-D FDTD and **then** supplied to the 2-D grid.

The interpolation process is quite straight forward, with emphasis given to insuring the correct timing. Assume the 1-D FDTD simulation has a time sample interval of $\Delta t'$ and the 2-D simulation has a time sample interval of Δt . To meet the above criteria, $\Delta t' \leq \Delta t$. In order to insure that interpolation, and not extrapolation, is being performed, the 1-D simulation must be performed until $n'\Delta t' \geq n\Delta t$. The number of previous time values that must be stored in order to perform the interpolation is equal to the order of interpolation desired. Increasing the order will increase accuracy, but since previous values must be stored for **every** point on the 1-D grid, the order should be kept as low as possible.

Once the values needed for interpolation have been identified and calculated, any standard interpolation algorithm, such as Lagrange Interpolation, can be used. After the interpolated edge E_z values have been calculated, they can then be supplied to the 2-D grid. The edge H_y values need not be interpolated.

V. FDTD SIMULATION RESULTS

Several experiments were performed using the methods described above using a variety of parameters. Excellent results were obtained for angles ranging from 0 to 85 degrees. The general method of each experiment was the same. A Gaussian pulse plane wave was created along the initial boundary at $x = 0$ with time variation corresponding to various propagation angles and with various values of R_g . In each of these experiments, the plane wave encounters a PML ABC at the back of the grid. The ABC in question is from [5] with 8 PML layers and conductivity profile $\sigma_i = \sigma_f(i/8)^{3.7}$. Figure 2a shows a 50×50 view sampled from a 200×200 grid. The propagation angle is 45° degrees and the Courant Number R_g is 0.5. Note that the wave propagates without edge distortion. Figure 2b is the same wave 200 time steps later. The wave has encountered the ABC and no reflection is visible, even in the lattice corner, where the 1-D FDTD accurately extends the 2-D ABC interaction calculation to the edge. In the scattered field view 2c, which is at the same time step as 2b, the magnification has been increased by 5 orders of magnitude and the incident field has been removed. The features to note are that the scattered wave satisfies Snell's law and that the scattered wave is uniform along the 45° angle, *i.e.*, the introduction of the 1-D FDTD on the right edge of the grid has not introduced any additional reflection artifacts. Figure 3a is once again a 50×50 view of a 200×200 grid. Here the Gaussian pulse plane wave is incident on the ABC at 70° . Once again R_g is 0.5 but now $R_g' = 0.25/\cos\theta$. The 1-D wave is traveling at one-half the velocity needed to keep pace with the 2-D wave, so only every second time sample is passed to the 2-D grid. Figure 3b is the same wave 150 time steps later. As with 45° wave, the interaction with the ABC has produced no visible reflection. Once again it may be noticed that the scattered wave, Figure 3c, obeys Snell's law and is uniform along the Snell angle. Clearly the visible reflection is due solely to the plane wave interacting with the ABC, which is the desired information.

VI. CONCLUSIONS

A method for testing angle-independent ABCs has been described. By using a one-dimensional FDTD simulation on the left and right edges of a two-dimensional grid, a plane wave incident on a ABC at the back edge of the grid at a single angle can be created and propagated. This method will greatly simplify the analysis of angle-independent ABC performance. The method has been tested using a Gaussian pulse plane wave with a variety of parameters and has been shown to give excellent results. Finally, since all of the desired information is found in the 1-D simulations, it can be concluded that the analysis of angle-independent ABCs may be carried out using only 1-D simulations.

REFERENCES

- [1] Berenger, J., "A Perfectly Matched Layer for the Absorption of Electromagnetic Waves", *Journal of Computational Physics*, Vol. 114, No. 1, pp. 185–200, October 1994.
- [2] Yee K.S., "Numerical Solution of Initial Boundary-Value Problems Involving Maxwell's Equations in Isotropic Media", *IEEE Trans. Ant. Prop.*, Vol. AP-14, No. 1, pp. 302–307, May 1966.
- [3] Mur, G., "Absorbing Boundary Conditions for the Finite-Difference Approximation of the Time-Domain Electromagnetic-Field Equations", *IEEE Transactions on Electromagnetic Compatibility*, Vol. EMC-23, No. 4, pp. 377–382, November 1981.
- [4] Sadiku, M.N.O., "Numerical Techniques in Electromagnetics", pp. 139–144, CRC Press, Boca Raton, 1992.
- [5] Rappaport, C., "Interpreting and Improving the PML Absorbing Boundary Condition Using Anisotropic Lossy Mapping of Space", *IEEE Trans. on Magnetics*, pp. 968–974, May 1996.

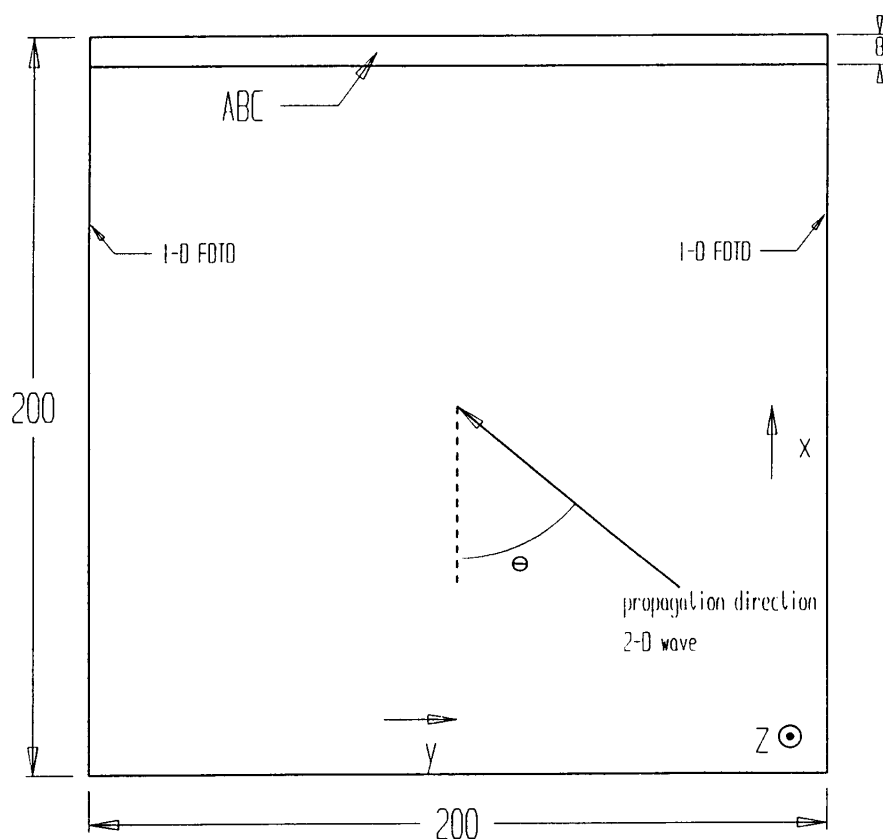


Figure 1
Geometry of grid.

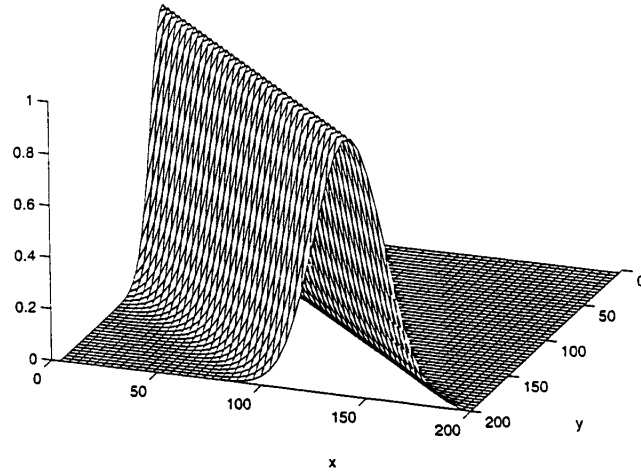


Figure 2a Gaussian pulse plane wave generated at $x = 0$ incident on the ABC at 45° . After 500 time steps, the 1-D FDTD simulation at $y = 200$ aligns perfectly with the 2-D FDTD simulation throughout the grid. ABC exists for $192 \leq x \leq 199$.

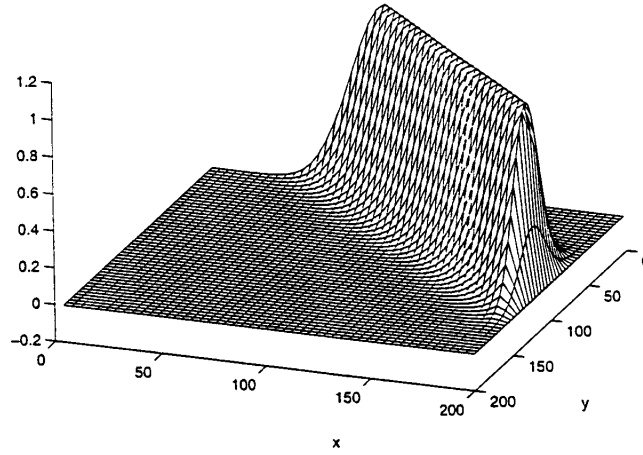


Figure 2b Gaussian pulse of Figure 2a, 200 time steps later: total field. The pulse has encountered the ABC, and is almost completely absorbed.

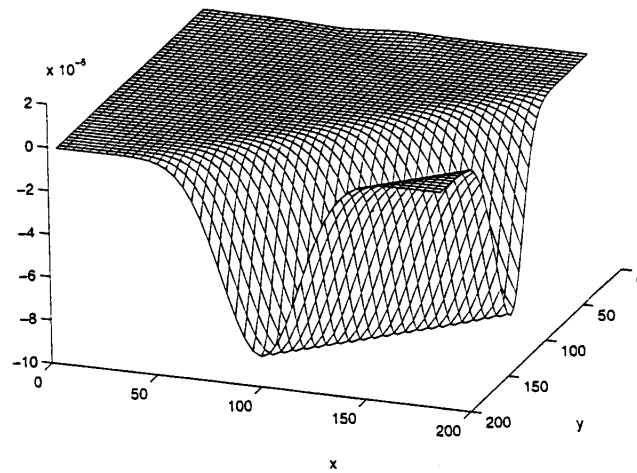


Figure 2c Scattered field at the same time of Figure 2b showing the residual reflection of the ABC.

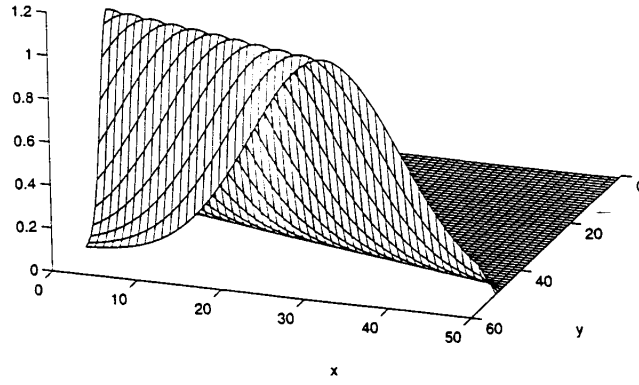


Figure 3a Gaussian pulse plane wave generated at $x = 0$ incident on the ABC at 70° . After 375 time steps, the 1-D FDTD simulation at $y = 200$ aligns perfectly with the 2-D FDTD simulation throughout the grid. ABC exists for $192 \leq x \leq 199$.

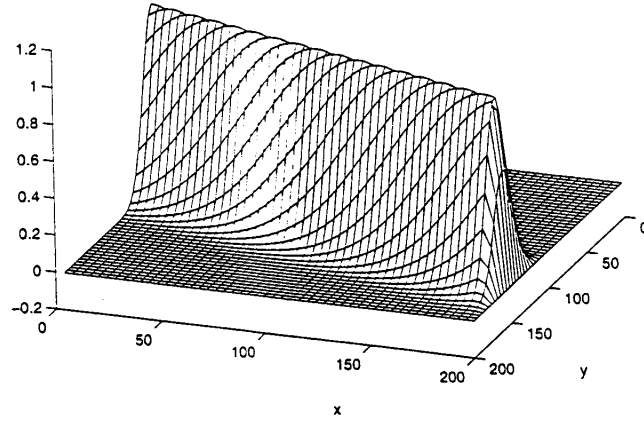


Figure 3b Gaussian pulse of Figure 3a 150 time steps later: total field. The pulse has encountered the ABC and is almost completely absorbed.

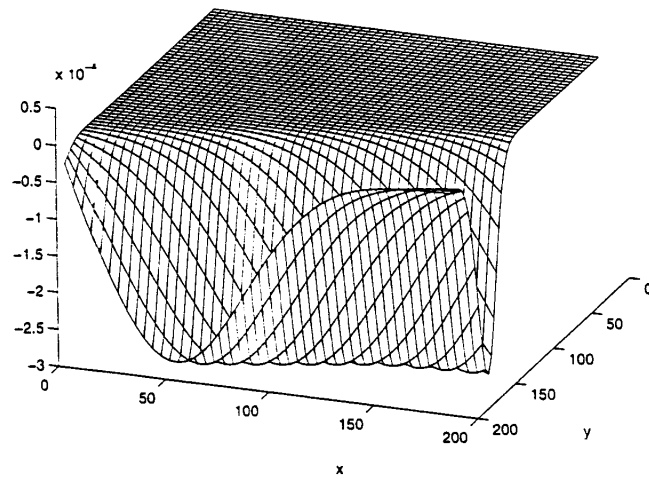


Figure 3c Scattered field at the same time of Figure 3b showing the residual reflection of the ABC.

Modeling Dispersive Soil for FDTD Computation By Fitting Conductivity Parameters

Carey M. Rappaport and Scott C. Winton
Center for Electromagnetics Research
Northeastern University
Boston, MA 02115

Abstract The electrical parameters of soils are strongly dependent on their type, physical characteristics, and electromagnetic excitation frequency. When numerically modeling soil for subsurface sensing simulation, it is particularly important to account for this dispersion. When computations are done in the time domain, this dispersion becomes problematic. Using a difference equation relation—and its corresponding Z -transform—to model dispersion between electric field and current leads to an approximation of complex conductivity in the form of a ratio of polynomials in Z^{-1} (where $Z = e^{j\omega t}$).

It is shown that if the real dielectric constant is held constant at an average value and conductivity only is matched with a single (2,2) Padé approximant $\sigma(f) = (b_0 + b_1 Z^{-1} + b_2 Z^{-2}) / (1 + a_1 Z^{-1} + a_2 Z^{-2})$, then the resulting propagation number $\beta(f)$ and decay rate $\alpha(f)$ will both closely match those corresponding to real soil measurements. In particular, a simple relation governing the frequency behavior of conductivity as a function of soil moisture and density is presented, allowing for the efficient numerical prediction of wave propagation in soils of varying environmental characteristics. Computed FDTD results for scattering in soil—with the computational lattice terminated with a “soil-tuned” PML absorbing boundary condition—clearly show the significance of media dispersion

This variant is nontrivial, since unlike with the frequency domain constitutive relation, electric flux and field are related by convolution in the time domain: $\mathbf{D} = \epsilon * \mathbf{E}$. The electric current is also a more complicated function of electric field in the time domain, since conductivity cannot merely be included as part of a complex permittivity.

The standard approaches to modeling dispersion in the FDTD method involve either recursively computing the convolution (as cleverly developed by Luebbers, *et. al.* [1]); or by approximating the frequency domain dispersive complex dielectric constant with a series of simple rational functions (Debye or Lorentz models) of $j\omega$ [2,3], and then by multiplying the constitutive relation by the denominator and inverse Fourier transforming the result into the time domain. While these methods are effective, they suffer from the limitations of numerical computation. Namely, that for good dispersive media modeling, higher-order, multiple-pole dielectric constant functions are necessary; but as the order of the function increases, so does the required storage of previous time field values for the entire grid, along with the sensitivity and numerical instability of the algorithm. In particular, to suit the conceptual elegance and simplicity of the FDTD method, it is important to keep the media model to at worst second-order. This presents a problem for the conventional complex dielectric constant models, which must accurately approximate both real and imaginary frequency dependencies simultaneously, using at most two poles.

I. INTRODUCTION

Recent interest in ground penetrating radar for locating and identifying buried waste, land mines, and excavation obstacles has motivated the development of advanced computational tools to simulate wave propagation in soil. In particular, the need exists to analyze ultra-wideband signals which might balance the trade-off between penetration depth and target resolution. Soil is a difficult medium to model since it is inhomogeneous, lossy, dispersive, and has an irregular surface boundary. For flexibility in predicting radar scattering from both metal and plastic targets buried in soil with rock inclusions and topped with vegetation, the Finite Difference Time Domain offers significant advantages over other standard computational techniques.

To include the effects of frequency-dependent conductivity and dielectric constant, a dispersive variant of the FDTD algorithm must be employed.

For certain types of media, however, it is possible to separate the modeling of real dielectric constant and conductivity. In both biological tissue and soil, for instance, the lossy dispersive wave propagation is governed almost entirely by the frequency-dependent conductivity. For these media, the real dielectric constant, though frequency-dependent, does not significantly affect either the real propagation constant β , nor the decay rate α [4]. As such, their electrical characteristics can be well-modeled with a constant relative permittivity ϵ' , and a second-order-in-frequency conductivity σ . Further, by modeling σ in terms of powers of the Z -transform variable Z^{-1} (which readily transform to time delays), the conversion of the generalized dispersive Ohm's Law $\mathbf{J}(Z) = \sigma(Z)\mathbf{E}(Z)$ to the time domain is particularly straightforward [5]. The problem addressed with this report is the specific selection of modeling parameters for a typical, well-studied soil, that simply and efficiently accounts for variations in density and moisture content.

MODELING PUERTO RICAN CLAY LOAM USING SECOND-ORDER CONDUCTIVITY

Arguably the most widely-cited soil measurement study is that of Hipp [6], which provides conductivity and real permittivity of San Antonio and Puerto Rican clay loam as a function of moisture (as a percent of dry weight) m , and density (g/cc) d , for the frequency range 30 to 3840 MHz. The method for developing the general second-order conductivity soil model is based on this experimental data set. While other soils will have different electrical characteristics, it is expected that general trends will be similar to that of Puerto Rican clay loam. Also, because of the wide variety of soils, and the difficulty in obtaining carefully generated soil measurements, every attempt was made to keep the model parameters as simple as possible, with mostly linear dependence on their physical characteristics.

First, to maintain easy conversion to time domain it is essential that the conductivity for all moisture and density cases have the form:

$$\sigma(Z) = \frac{J(Z)}{E(Z)} = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2}}{1 + a_1 Z^{-1} + a_2 Z^{-2}} \quad (1)$$

The b_i and a_i coefficients will each be independent functions of m and d . The actual measured values of conductivity correspond to the real part of $\sigma(Z)$, with $Z = e^{j2\pi f \Delta t}$, for FDTD time step Δt . The imaginary part, divided by $j2\pi f \epsilon_0$ adds to the frequency-independent dielectric constant ϵ' . In the time domain, Eqn. (1) becomes:

$$\mathbf{J}^n + a_1 \mathbf{J}^{n-1} + a_2 \mathbf{J}^{n-2} = b_0 \mathbf{E}^n + b_1 \mathbf{E}^{n-1} + b_2 \mathbf{E}^{n-2} \quad (2)$$

and Ampere's Law, as usual, is given by:

$$\nabla \times \mathbf{H}^{n+1/2} = \epsilon_0 \epsilon' \frac{\mathbf{E}^{n+1} - \mathbf{E}^n}{\Delta t} + \frac{\mathbf{J}^{n+1} + \mathbf{J}^n}{2} \quad (3)$$

In a previous publication [7], the best coefficients for Puerto Rican clay loam were determined for each separate sample of moisture and density. Although useful from a numerical view, these coefficients are not very helpful for the practical problem of determining wave propagation for an intermediate soil condition. To address this difficulty, a more unified approach is developed.

For the various moisture and density cases measured in [6], the best modeling coefficient values of Eqn. (1) vary considerably. However, the denominator coefficients: a_1 and a_2 , only differ at most by about 15%. Choosing fixed values $a_1 = -1.6$ and $a_2 = .64$ and allowing variation of the b_i coefficients gives up a little accuracy but provides a more simple model. Instead of finding the values of b_i which minimize a non-linear cost function for each moisture/density case, the current modeling

method simply solves for b_1 , b_2 , and b_3 by setting the real parts of $\sigma(Z)$ in Eqn. (1) to the measured values, at three particular frequencies: 120, 960, and 3840 MHz. While this method arbitrarily emphasizes the fit at these frequencies, it avoids justifying what type of cost function to use. A least-squares cost function for normalized error in σ and ϵ' , for example, is not as precise as for normalized error in the real and imaginary parts of the wave number $k = \beta - j\alpha$; and neither appropriately weighs the error on decaying wave amplitude across the frequency range.

The b_i coefficients for constant density and moisture levels $m = 2.5, 5, 10$, and 20 are first determined, and then fit to simple functions of m : $b_i \approx b_{i0} + b_{i1} \log m$. It was found that while a very good fit is possible for each b_i , the numerator of Eqn. (1) is small for low frequencies (where $Z \approx 1$), so that small errors in the b_i approximations lead to large errors in $\sigma(f)$. To avoid this cancellation problem, approximations are determined for b_1 , b_2 , and the sum $bs = b_0 + b_1 + b_2$, with the sum having the form $bs \approx bs_0 + bs_1 m + bs_2 m^2$. The resulting coefficients for $d = 1.2, 1.4$, and 1.6 are given in Table 1. The average dielectric constant chosen for these models is simply the measured value at 960 MHz. A quadratic least-squares fit to these data for each density, $\epsilon(m) = \epsilon_0 + \epsilon_1 m + \epsilon_2 m^2$ is given in Table 2.

Table 1: Conductivity Numerator Coefficients

Coeff.	Density (g/cc)		
	1.2	1.4	1.6
b_{10}	0.0484917	0.0540739	0.0160977
b_{11}	-0.136191	-0.162553	-0.188476
b_{20}	-0.0125623	-0.0220495	-0.0004839
b_{21}	0.0574154	0.0722359	0.0808164
bs_0	1.14562E-4	7.06024E-5	1.1323E-4
bs_1	-3.9964E-6	-4.218E-6	-1.2685E-5
bs_2	1.01241E-6	1.92513E-6	3.54086E-6

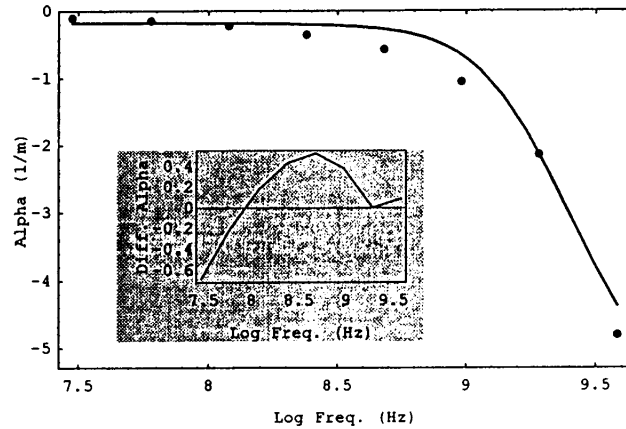
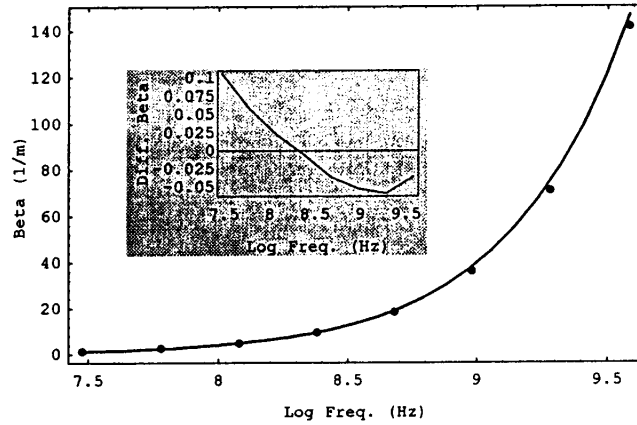
Table 2: Dielectric Constant Coefficients

Coeff.	Density (g/cc)		
	1.2	1.4	1.6
ϵ_0	2.8	2.817	3.95
ϵ_1	0.14	0.171	0.0632
ϵ_2	0.012	0.0181	0.0318

Figure 1 shows the accuracy of the approximation for $d = 1.4$, and the extreme moisture cases, $m = 2.5$ and 20% . Plotted in this figure are the magnitude of the propagation constant β and decay rate

Out[141]=

$$\sigma(z) = \frac{0.0508043 + \frac{0.0441395}{z} - \frac{0.0948717}{z^2}}{1 + \frac{0.64}{z} + \frac{1.6}{z^2}}, \quad \text{Av. Diel. Const.} = 3.2$$



Out[252]=

$$\sigma(z) = \frac{0.239297 + \frac{0.19435}{z} - \frac{0.432891}{z^2}}{1 + \frac{0.64}{z} + \frac{1.6}{z^2}}, \quad \text{Av. Diel. Const.} = 13.5$$

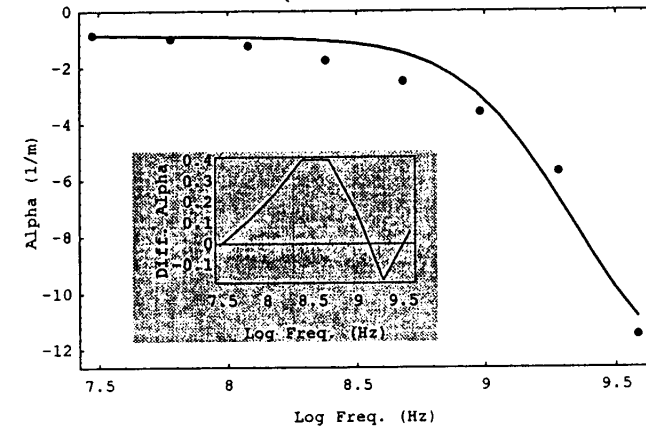
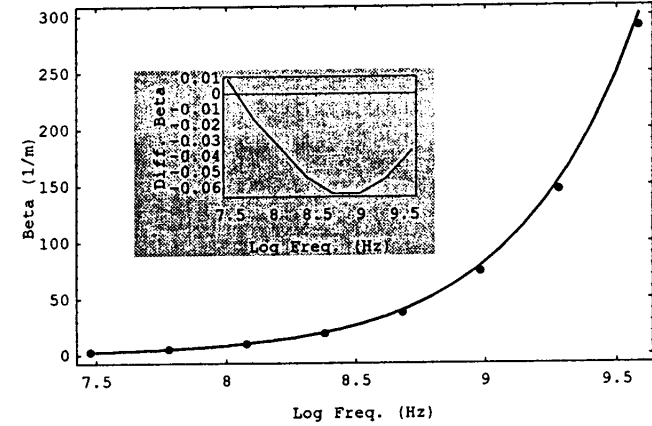


Figure 1 Complex conductivity, frequency-independent dielectric constant, wave propagation number, decay rate, and normalized error, as a function of log frequency, for conductivity model (—), and measured (· · ·) data for a) $m = 2.5\%$, and b) $m = 20\%$.

α of the model and the measured values of Puerto Rican clay loam as a function of frequency for the entire measured frequency range 30 to 3840 MHz. Also shown in the inserts are the fractional errors $\Delta\beta/\beta$ and $\Delta\alpha/\alpha$. The agreement is surprisingly good for such a simple model across two decades of frequency. It should be noted that the fit is even better for the intermediate moisture values 5 and 10%, and similar for the other density cases.

SOIL-TUNED PML ABC

With all FDTD scattering problems, it is necessary to minimize reflections from the lattice boundaries with absorbing boundary conditions (ABC). A novel ABC, the "soil-tuned" Perfectly Matched Layer (PML), has been developed. The soil-tuned PML is a modified version of the Berenger ABC [8,9] which specifically absorbs waves incident from dispersive media. Since the efficiency of transmission of waves into the PML is dependent on the closeness of match of the transverse wave impedance on both sides of the layer, it is essential to select the electrical parameters of the PML appropriately. For soil parameters ϵ'_{soil} , σ_{soil} , μ'_{soil} , the desired impedance match condition is:

$$\begin{aligned}\eta_{\text{soil}} &= \sqrt{\frac{\mu'_{\text{soil}}\mu_0}{(\epsilon'_{\text{soil}} - j\sigma_{\text{soil}}/\omega\epsilon_0)\epsilon_0}} \\ &= \sqrt{\frac{\mu'_{\text{soil}}\mu_0(1 - j\sigma_P/\omega\epsilon_0)}{(\epsilon'_{\text{soil}} - j\sigma_{\text{soil}}/\omega\epsilon_0)\epsilon_0(1 - j\sigma_P/\omega\epsilon_0)}} \quad (4) \\ &= \eta_{\text{PML}}\end{aligned}$$

where σ_P is the usual increasing conductivity profile of the PML layer. While the relations of Eqn. (4) correspond to the transverse impedance only for normal incidence on the PML layer, the split-field or auxiliary equation PML formulation ensure impedance match for all incidence angles, provided a match occurs for normal incidence.

Eqn. (4) therefore specifies the effective dielectric constant and conductivity in the PML layer,

$$\begin{aligned}\epsilon'_{\text{PML}} &= \epsilon'_{\text{soil}} \\ \mu'_{\text{PML}} &= \mu'_{\text{soil}} \\ \sigma_{\text{PML}} &= \sigma_{\text{soil}} + \epsilon'_{\text{soil}}\sigma_P \\ \sigma_{\text{PML}}^m &= \sigma_P\mu_0/\epsilon_0\end{aligned} \quad (5)$$

where the double conductivity term $-\sigma_{\text{soil}}\sigma_P/(\omega\epsilon_0)^2$ is neglected as negligible compared to ϵ'_{soil} for all but the largest PML conductivity layers. If σ_{soil} were frequency independent, Eqn. (5) would provide constant constitutive parameters for the PML equations which could be used directly in the time domain. Since σ_{soil} is dispersive, however, the PML equations must make use of the auxiliary difference

Eqn. (2), with b_i coefficients in Eqn. (1) adjusted to account for the new conductivity values of Eqn. (5c). The soil-tuned PML can be thought of a modification of the dispersive media calculation with split fields and magnetic loss.

NUMERICAL TEST CASE

Using the formula derived in the above section to specify the electrical characteristics of Puerto Rican clay loam with density 1. g/cc, and 10% moisture, a 2-dimensional FDTD calculation simulating plane wave scattering from a buried one wavelength diameter circular metal cylinder was performed. The geometry of the scattering lattice is shown in Figure 2. The lattice is oversized in width, 500 grid points, to prevent reflections from the sides, and is terminated with an 8-cell soil-tuned PML, to prevent reflections from the back lattice boundary. A 0.96 GHz modulated, gaussian envelope plane wave is initiated along the front grid boundary. One-dimensional FDTD calculations on the left and right edges ensure that the plane wave propagates from front to back without distortion. The time and space steps used are $\Delta t = 20$ ps and $\Delta x = 4.6$ mm, the nominal phase velocity is $v = c/\sqrt{\epsilon'} = .383c$, and the Courant number is held at 0.5.

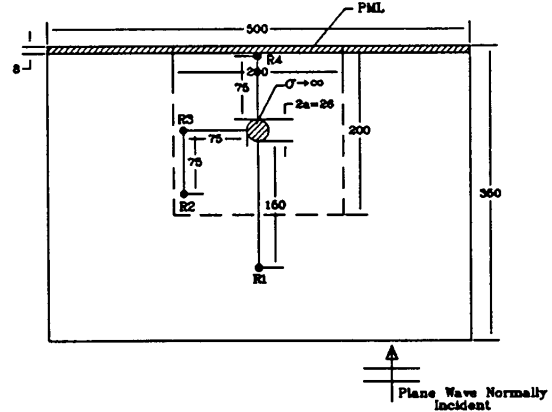


Figure 2 Geometry of the scattering problem

The four surface plots of Figure 3 show the electric field distribution across the central 200 by 200 grid point section of the computational grid (indicated in Figure 2) at various times. The upper left and right plots show the total and scattered field as the incident wave begins scattering from the circular cylinder. The lower two plots show scattered field 100 and 300 time steps later. The scattering is symmetric, as expected. For the 1100 Δt plot, the residual reflections from the PML ABC, at $x = 100$, are visible. The amplitude of these reflections are of the order of 10^{-3} , which is about 3% of the field amplitude incident on the back boundary. Although

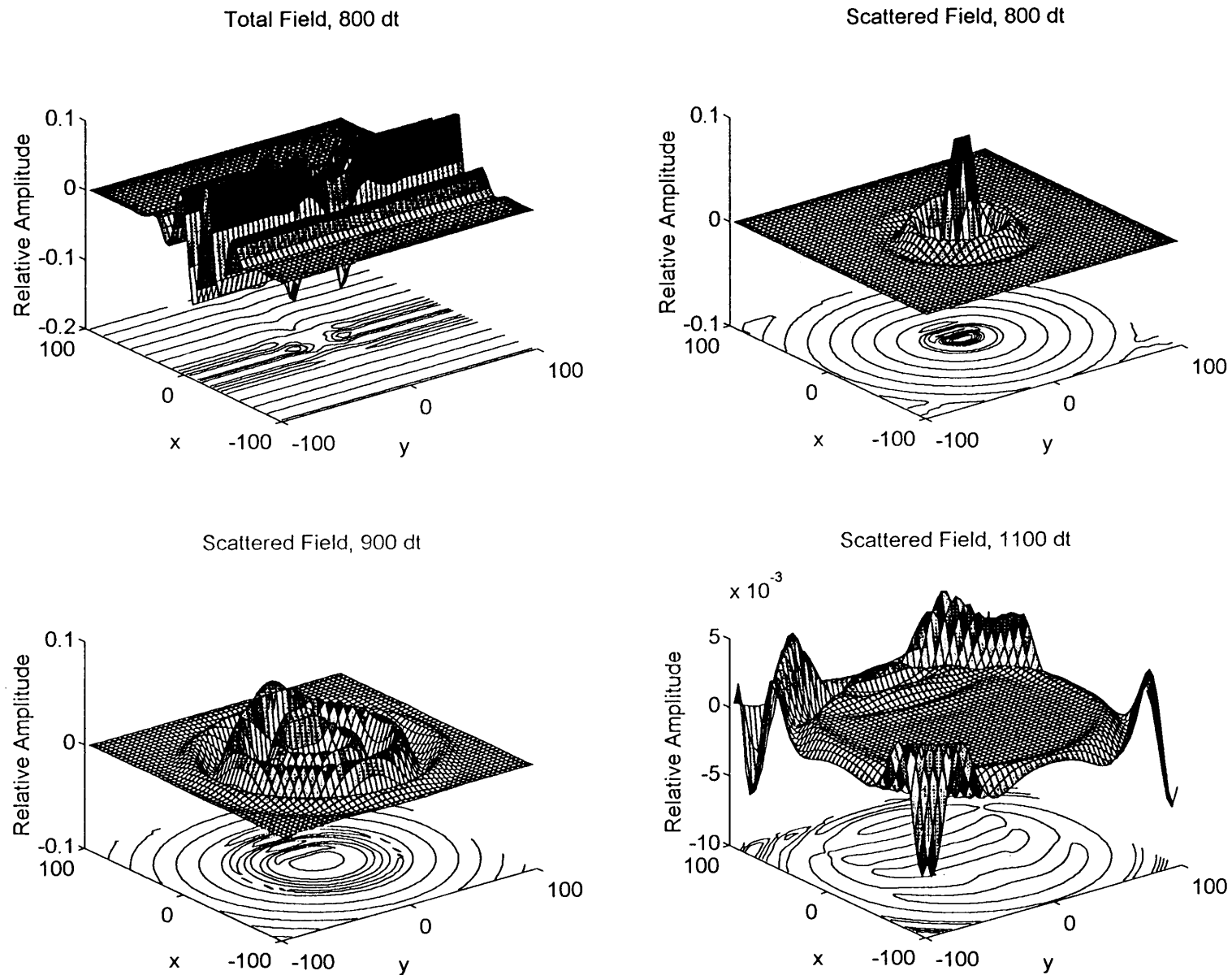


Figure 3 FDTD field plots simulating scattering by a perfectly conducting circular cylinder, surrounded by dispersive clay loam, $d = 1.4$, $m = 10\%$, with plane wave incidence.

still rather significant, the soil-tuned PML generates about one-half the reflections of generated by a conventional PML used to terminate this dispersive medium.

Figure 4 compares frequency independent and dispersive propagation, by showing the received signals at a single point ($R1$ in Figure 2) 150 grid points, directly in front of the circular scatterer. For the frequency independent case (left plots), the conductivity is kept constant at the measured value at 960 MHz, 0.032 S/m. The upper plots give the total field for the two cases, indicating only minor differences in the modulated plane wave propagation. It is interesting to note the much more significant differences in the lower, scattered field plots. In particular, the wave amplitude of the dispersive medium is twice that of the uniform conductivity case, the propagation speed is slightly different, and the higher frequencies have been attenuated—with only nine discernible maxima compared to ten in the uniform case.

CONCLUSIONS

A model of dispersive soil—with simple functional dependence on moisture and density—that can easily be adapted into the FDTD method has been developed. Based on a (2,2) Padé approximant in transform variable Z^{-1} , the model requires storing at most four additional arrays per time calculation. For simplicity, the model maintains constant denominator coefficients, with numerator coefficients being limited to a worst second order in moisture.

Also presented is a soil-tuned PML absorbing boundary condition, which is a modified variant of the PML used to terminate dispersive media lattices. In this new formulation, the PML constitutive parameters are adjusted to ensure transverse impedance matching with the dispersive medium. Its performance for soil is twice as good as the conventional PML.

Clearly, dispersion is important, and may have a greater effect in two or three dimensional scattering applications. Neglecting to model the frequency dependence of soil in wave propagation simulation can lead to significant errors. The current model simply and effectively approximates the dispersion for Puerto Rican clay loam through the entire 30 to 3840 MHz band.

ACKNOWLEDGEMENTS

The authors are grateful to Ann W. Morgenthaler for helpful discussions.

REFERENCES

- [1] Luebbers, R., Hunsberger, F., Kunz, K., Standler, R., and Schneider, M., "A Frequency-Dependent Finite Difference Time Domain Formulation for Dispersive Materials", *IEEE Transactions on Electromagnetic Compatibility*, pp. 222-227, Vol. 32, No. 3, March 1990.
- [2] Kashiwa, T. and Fukai, I., "A Treatment by the FD-TD Method for the Dispersive Characteristics Associated with Electronic Polarization", *Microwave and Guided Wave Letters*, Vol. 16, No. 6, pp. 203-205, June 1990.
- [3] Gandhi, O., "A Frequency-Dependent Finite Difference Time Domain Formulation for General Dispersive Media", *IEEE Transactions on Microwave Theory and Techniques*, Vol. 41, No. 4, pp. 658-665, April 1993.
- [4] Rappaport, C., and Weedon, W., "Efficient Modeling of Electromagnetic Characteristics of Soil for FDTD Ground Penetrating Radar Simulation", *1996 IEEE Antenna and Propagation Society/URSI Symposium Digest*, pp. 620-623, June 1996.
- [5] Weedon, W., and Rappaport, C., "A General Method for FDTD Modeling of Wave Propagation in Arbitrary Frequency-Dispersive Media", *IEEE Transactions on Antennas and Propagation*, accepted for publication March 1997.
- [6] Hipp, J., "Soil Electromagnetic Parameters as Functions of Frequency, Soil Density, and Soil Moisture", *Proceedings of the IEEE*, Vol. 62, No. 1, pp. 98-103, January 1974.
- [7] Weedon W., Rappaport, C., and Silevitch, D., "Modeling and Stability Considerations for FDTD Analysis of Wave Propagation in Soils", *1996 SPIE Aerosense Symposium*, pp. 245-252, April 1996.
- [8] Berenger, J., "A Perfectly Matched Layer for the Absorption of Electromagnetic Waves", *Journal of Computational Physics*, Vol. 114, No. 1, pp. 185-200, October 1994.
- [9] Rappaport, C., "Interpreting and Improving the PML Absorbing Boundary Condition Using Anisotropic Lossy Mapping of Space", *IEEE Transactions on Magnetics*, Vol. 32, No. 3, May 1996, pp. 968-974.

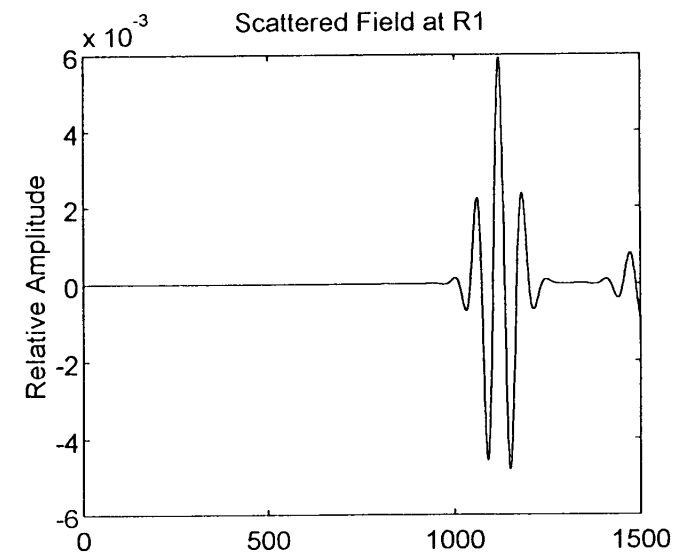
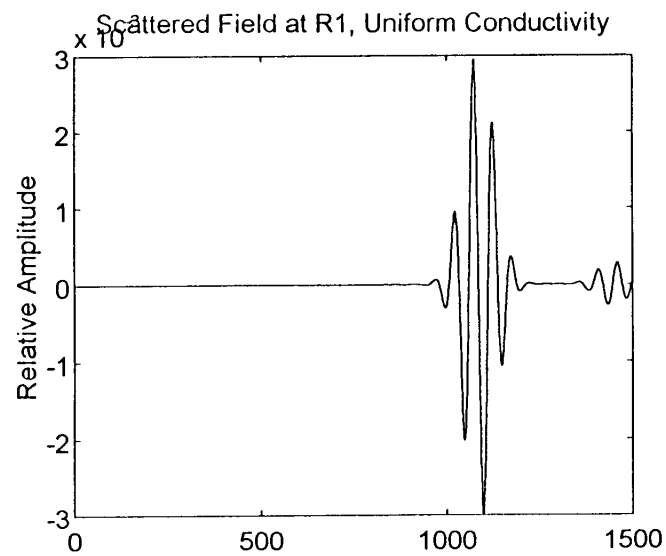
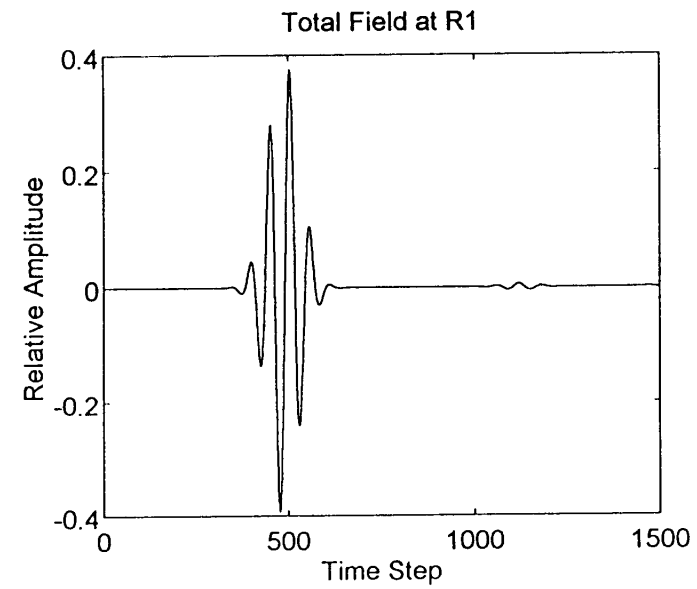
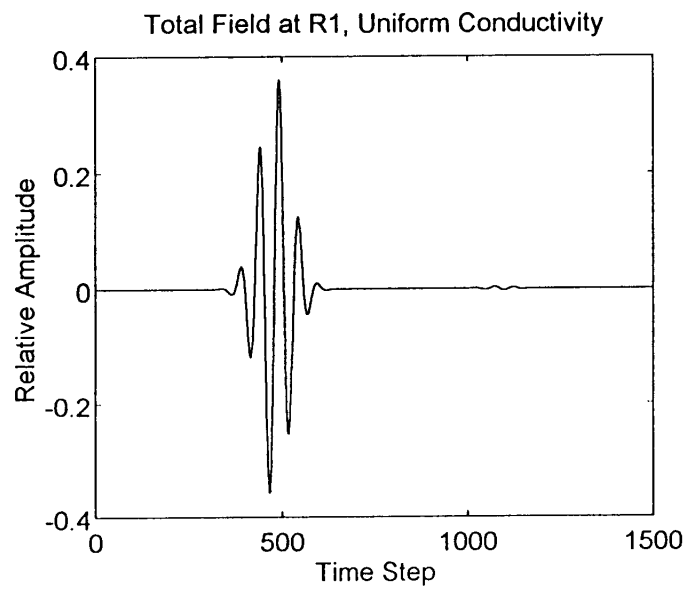


Figure 4 Received signals at point *R1* for soil $d = 1.4$, $m = 10\%$, with frequency independent $\sigma_c = 0.032$ S/m, $\epsilon' = 6.8$; and for conductivity model.

A General Method for FDTD Modeling of Wave Propagation in Arbitrary Frequency-Dispersive Media

William H. Weedon, *Member, IEEE*, and Carey M. Rappaport, *Senior Member, IEEE*

Abstract— A general formulation is presented for finite-difference time-domain (FDTD) modeling of wave propagation in arbitrary frequency-dispersive media. Two algorithmic approaches are outlined for incorporating dispersion into the FDTD time-stepping equations. The first employs a frequency-dependent complex permittivity (denoted Form-1), and the second employs a frequency-dependent complex conductivity (denoted Form-2). A Padé representation is used in Z-transform space to represent the frequency-dependent permittivity (Form-1) or conductivity (Form-2). This is a generalization over several previous methods employing either Debye, Lorentz, or Drude models. The coefficients of the Padé model may be obtained through an optimization process, leading directly to a finite-difference representation of the dispersion relation, without introducing discretization error. Stability criteria for the dispersive FDTD algorithms are given. We show that several previously developed dispersive FDTD algorithms can be cast as special cases of our more general framework. Simulation results are presented for a one-dimensional (1-D) air/muscle example considered previously in the literature and a three-dimensional (3-D) radiation problem in dispersive, lossy soil using measured soil data.

Index Terms— Dispersive media, FDTD methods.

I. INTRODUCTION

THE finite-difference time-domain (FDTD) technique is one of the most popular computational electromagnetic techniques for modeling time-domain wave propagation [1]. One of the advantages to using the FDTD technique is that all of the frequency components of the scattered field resulting from a broadband transmitted pulse may be computed simultaneously. The standard Yee FDTD algorithm [2] models inhomogeneous materials using spatially varying, but frequency-independent permittivity and conductivity. To model dispersive media, the standard Yee time-stepping equations need to be modified.

Several techniques have been introduced recently to incorporate frequency dispersion into FDTD models [3]–[17]. These dispersive FDTD algorithms have been shown to be important for computing optical pulse propagation [7], [11], predicting wave propagation in magnetized plasmas and in biological tissue [8], [14], [15]. One of our interests is in utilizing dispersive FDTD algorithms to model wave propagation in lossy, dispersive, inhomogeneous soils. This type of modeling is extremely useful for predicting the performance of ground-penetrating radar (GPR) systems in specified inhomogeneous environments, computing the electromagnetic fields scattered by various types of scattering objects such as buried waste drums, metallic or dielectric pipes, and pollution plumes, and for testing GPR detection and imaging algorithms with synthetic data.

The existing frequency-dependent FDTD methods can roughly be categorized into three types: 1) methods that efficiently implement a discrete convolution of the dispersion relation $D(\mathbf{r}, t) = \epsilon(t) * E(\mathbf{r}, t)$ [3], [4], [12]–[15]; 2) methods that discretize a differential equation relating $D(\mathbf{r}, t)$ to $E(\mathbf{r}, t)$ [5]–[9], [16]; and 3) Z-transform methods [10], [11]. The discrete convolution methods [3], [4] rely on either a Debye model involving single [3] or multiple [14] first-order poles, a Lorentz model utilizing second-order poles [4], or a Drude model [12], to implement the discrete convolution efficiently with a minimal storage requirement. The differential equation methods offer a more general representation for the dispersion relation, but they typically reduce to using either a Debye or Lorentz model [1]–[9]. One disadvantage of the differential equation method is that it is not clear which discretization formula to use (forward difference or backward difference) when high-order derivatives are involved [16]. In the Z-transform method proposed by Sullivan [10], [11], [17], a Debye or Lorentz dispersion model is converted to the time domain and sampled, a process analogous to impulse-invariance filter design in the signal processing literature.

Here, we present a more general approach for dispersive FDTD modeling, utilizing a Padé approximation of the complex frequency-dependent permittivity or conductivity in Z-transform space. This new method characterizes the dispersive medium as a bank of digital pole-zero infinite-impulse response (IIR) filters [18], which leads directly to FDTD implementations. The generation of appropriate filter functions

Manuscript received April 4, 1996; revised September 30, 1996. This work was supported by the U.S. Department of Energy under Grant DE-FC07-95ID13395. Computer time was provided by the National Center for Supercomputer Applications (NCSA) at the University of Illinois at Urbana-Champaign and the NAS Systems Division of the NASA Ames Research Center.

The authors are with the Center for Electromagnetics Research, Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA.

Publisher Item Identifier S 0018-926X(97)02289-8.

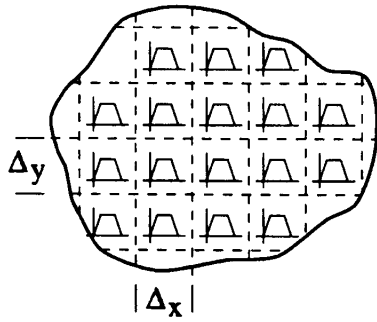


Fig. 1. Representation of a dispersion medium as a subwavelength spatial array of linear time-invariant filters.

for specific dispersive media is reduced to a digital-filter design problem, where both the magnitude and phase characteristics are specified.

Two new dispersive FDTD algorithms employing Padé approximations in Z-transform space are developed here. The first method, which we shall refer to as Form-1, utilizes a Padé model of dielectric permittivity $\epsilon(Z)$, while the second, Form-2, utilizes a Padé model of conductivity $\sigma(Z)$. We discuss some subtle differences between the two methods, particularly in terms of their stability and flexibility in fitting measured data.

The design problem of finding optimal filter coefficients to match measured permittivity and conductivity data, that result in stable, accurate FDTD algorithms is touched on briefly here, but shall be considered in more detail in a future paper.

II. DEVELOPMENT OF FORM-1 AND FORM-2 DISPERSIVE FDTD ALGORITHMS

Formulation of the dispersive FDTD algorithms begins with the assumption of a general inhomogeneous, isotropic, and causal dispersive medium. Throughout this paper, attention is restricted to the case of dispersive dielectrics, where we assume that $\mu = \mu_0$. All of the results, however, may be extended to the case of dispersive magnetic media in a straightforward manner.

If a volume V of the dispersive dielectric medium is divided into subwavelength cells of volume $\Delta_x \Delta_y \Delta_z$, as in Fig. 1, each subwavelength cell in the dispersive medium can be thought of as a linear time-invariant filter that acts on the electric field $\mathbf{E}(\mathbf{r}, t)$. The entire medium may be viewed as a spatial filter array, which can be implemented in two particular ways: 1) as a complex permittivity filter $\mathbf{D}(\mathbf{r}, Z) = \epsilon(\mathbf{r}, Z)\mathbf{E}(\mathbf{r}, Z)$ (Form-1 algorithm) or 2) as a complex conductivity filter $\mathbf{J}_c(\mathbf{r}, Z) = \sigma(\mathbf{r}, Z)\mathbf{E}(\mathbf{r}, Z)$ with constant real permittivity $\epsilon(\mathbf{r})$ (Form-2 algorithm).

A. Permittivity Filter (Form-1) Algorithm

In developing the permittivity filter (Form-1) algorithm, we include dispersion in the Maxwell equations as a frequency-dependent permittivity. In the time-domain Ampere law

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \frac{\partial \mathbf{D}(\mathbf{r}, t)}{\partial t} + \mathbf{J}_s(\mathbf{r}, t) \quad (1)$$

$\mathbf{J}_s(\mathbf{r}, t)$ represents a driving source current and

$$\mathbf{D}(\mathbf{r}, t) = \epsilon(\mathbf{r}, t) * \mathbf{E}(\mathbf{r}, t). \quad (2)$$

In (2), the frequency-dependent dispersion is represented as a temporal convolution of the electric field with a time-dependent permittivity. The convolution integral implied by (2) has limits $t = 0$ to $t = \infty$ since we consider only causal media and do not consider spatial dispersion. The vector \mathbf{r} in (2) is treated only as a parameter.

To derive the dispersive finite-difference algorithms, we first define the Z-transform of the temporal vector field

$$\mathbf{A}^n(\mathbf{r}) = \mathbf{A}(\mathbf{r}, t = n \Delta_t), \quad n = 0, 1, \dots, N_t - 1 \quad (3)$$

with $\mathbf{A}(\mathbf{r}, t)$ representing either $\mathbf{E}(\mathbf{r}, t)$, $\mathbf{H}(\mathbf{r}, t)$, $\mathbf{D}(\mathbf{r}, t)$, $\epsilon(\mathbf{r}, t)$, or $\sigma(\mathbf{r}, t)$ as¹

$$\mathbf{A}(\mathbf{r}, Z) = \Delta_t \sum_{n=0}^{\infty} \mathbf{A}^n(\mathbf{r}) Z^{-n}. \quad (4)$$

Equation (2) becomes

$$\mathbf{D}(\mathbf{r}, Z) = \epsilon(\mathbf{r}, Z)\mathbf{E}(\mathbf{r}, Z) \quad (5)$$

in Z-transform space.

The permittivity Z-transform function $\epsilon(Z)$ may be effectively modeled using a Padé approximation in Z; that is given in

$$\frac{\epsilon(\mathbf{r}, Z)}{\epsilon_0} = \frac{b_0(\mathbf{r}) + b_1(\mathbf{r})Z^{-1} + b_2(\mathbf{r})Z^{-2} + \dots + b_{N_b-1}(\mathbf{r})Z^{-(N_b-1)}}{1 + a_1(\mathbf{r})Z^{-1} + a_2(\mathbf{r})Z^{-2} + \dots + a_{N_a}(\mathbf{r})Z^{-N_a}}. \quad (6)$$

One major advantage to using a Padé approximation in Z is that such a dispersion relation leads directly to a finite-difference implementation.

Consider the one-dimensional (1-D) case of a plane wave (E_x, H_y) traveling in the z direction. The Form-1 dispersive FDTD algorithm may be summarized as follows:

$$H_{y,k}^{n+1/2} = H_{y,k}^{n-1/2} - \frac{\Delta_t}{\mu_0 \Delta_z} [E_{x,k+1/2}^n - E_{x,k-1/2}^n] \quad (7)$$

$$D_{x,k+1/2}^{n+1} = D_{x,k+1/2}^n - \frac{\Delta_t}{\Delta_z} [H_{y,k+1}^{n+1/2} - H_{y,k}^{n+1/2}] - \Delta_t J_{s,x,k+1/2}^{n+1/2} \quad (8)$$

$$E_{x,k+1/2}^{n+1} = \frac{1}{b_{0,k+1/2}} \left\{ \frac{1}{\epsilon_0} [D_{x,k+1/2}^{n+1} + a_{1,k+1/2} D_{x,k+1/2}^n + \dots + a_{N_a,k+1/2} D_{x,k+1/2}^{n+1-N_a}] - b_{1,k+1/2} E_{x,k+1/2}^n - b_{2,k+1/2} E_{x,k+1/2}^{n-1} - \dots - b_{N_b-1,k+1/2} E_{x,k+1/2}^{n+2-N_b} \right\} \quad (9)$$

¹The Δ_t factor [10], [11] in (4) allows us to approximate the continuous-time Fourier transform of $\mathbf{A}(\mathbf{r}, t)$ as

$$\mathbf{A}(\mathbf{r}, \omega) = \int_0^{\infty} dt \mathbf{A}(\mathbf{r}, t) e^{-j\omega t} \approx \mathbf{A}(\mathbf{r}, Z = e^{j\omega \Delta_t})$$

where we assume that $\mathbf{A}(\mathbf{r}, t) = 0$ for $t < 0$.

where the electric field E_x and flux density D_x are evaluated at $t = n \Delta_t$ and $z = (k + 1/2) \Delta_z$, and the magnetic field H_y is evaluated at $t = (n + 1/2) \Delta_t$ and $z = k \Delta_z$ on the staggered Yee grid. The two-dimensional (2-D) and three-dimensional (3-D) Form-1 dispersive FDTD algorithms may be derived in a similar manner.

B. Conductivity Filter (Form-2) Algorithm

For the conductivity filter (Form-2) algorithm, we rewrite Ampere's law (1) as

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \epsilon_0 \epsilon_r(\mathbf{r}) \frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t} + \mathbf{J}_c(\mathbf{r}, t) + \mathbf{J}_s(\mathbf{r}, t) \quad (10)$$

where $\mathbf{J}_c(\mathbf{r}, t)$ represents an induced source current given by

$$\mathbf{J}_c(\mathbf{r}, t) = \sigma(\mathbf{r}, t) * \mathbf{E}(\mathbf{r}, t). \quad (11)$$

and $\mathbf{J}_s(\mathbf{r}, t)$ is again a driving source current. In (10), $\epsilon_r(\mathbf{r})$ is not time-dependent and, therefore, not dispersive. The dispersion is included only in the time-dependent conductivity $\sigma(\mathbf{r}, t)$. In Z-transform space, (11) becomes

$$\mathbf{J}_c(\mathbf{r}, Z) = \sigma(\mathbf{r}, Z) \mathbf{E}(\mathbf{r}, Z). \quad (12)$$

The conductivity Z-transform functions may be effectively modeled using a Padé approximation in Z

$$\sigma(\mathbf{r}, Z) = \frac{d_0(\mathbf{r}) + d_1(\mathbf{r})Z^{-1} + d_2(\mathbf{r})Z^{-2} + \dots + d_{N_d-1}(\mathbf{r})Z^{-(N_d-1)}}{1 + c_1(\mathbf{r})Z^{-1} + c_2(\mathbf{r})Z^{-2} + \dots + c_{N_c}(\mathbf{r})Z^{-N_c}}. \quad (13)$$

Discretizing (10) and, again, examining the 1-D case for brevity

$$\begin{aligned} E_{x,k+1/2}^{n+1} &= E_{x,k+1/2}^n \\ &- \frac{\Delta_t}{\epsilon_0 \epsilon_r, k+1/2 \Delta_z} [H_{y,k+1}^{n+1/2} - H_{y,k}^{n+1/2}] \\ &- \frac{\Delta_t}{\epsilon_0 \epsilon_r, k+1/2} J_{c,x,k+1/2}^n \\ &- \frac{\Delta_t}{\epsilon_0 \epsilon_r, k+1/2} J_{s,x,k+1/2}^{n+1/2}. \end{aligned} \quad (14)$$

The $J_{c,x,k+1/2}^{n+1/2}$ term can be obtained by the usual averaging of the conduction current at time steps n and $n+1$. In Z-transform space, the above becomes

$$\begin{aligned} E_{x,k+1/2} F_{k+1/2}(Z) &= - \frac{\Delta_t}{\epsilon_0 \Delta_z} [H_{y,k+1} - H_{y,k}] \\ &- \frac{\Delta_t}{\epsilon_0} J_{s,x,k+1/2} \end{aligned} \quad (15)$$

where

$$\begin{aligned} F_{k+1/2}(Z) &= \epsilon_r, k+1/2 (1 - Z^{-1}) \\ &+ \frac{\Delta_t}{2\epsilon_0} (1 + Z^{-1}) \sigma_{k+1/2}(Z). \end{aligned} \quad (16)$$

The rational function $F(\mathbf{r}, Z)$ may be written explicitly in the form

$$F(\mathbf{r}, Z) = \frac{e_0(\mathbf{r}) + e_1(\mathbf{r})Z^{-1} + \dots + e_{N_e-1}(\mathbf{r})Z^{-(N_e-1)}}{1 + c_1(\mathbf{r})Z^{-1} + c_2(\mathbf{r})Z^{-2} + \dots + c_{N_c}(\mathbf{r})Z^{-N_c}} \quad (17)$$

where $e_0 = \epsilon_r + (\Delta_t/2\epsilon_0) d_0$, $e_1 = \epsilon_r(c_1 - 1) + (\Delta_t/2\epsilon_0)(d_1 + d_0)$, \dots , $e_n = \epsilon_r(c_n - c_{n-1}) + (\Delta_t/2\epsilon_0)(d_n + d_{n-1})$, etc., and $N_e = \max\{N_c + 1, N_d\}$. Combining (15) and (17), the 1-D Form-2 dispersive FDTD algorithm is then

$$I_{x,k+1/2}^{n+1/2} = \frac{\Delta_t}{\epsilon_0 \Delta_z} [H_{y,k+1}^{n+1/2} - H_{y,k}^{n+1/2}] \quad (18)$$

$$\begin{aligned} E_{x,k+1/2}^{n+1} &= \frac{-1}{e_{0,k+1/2}} \left\{ I_{x,k+1/2}^{n+1/2} + c_{1,k+1/2} I_{x,k+1/2}^{n-1/2} \right. \\ &\quad + \dots + c_{N_c,k+1/2} I_{x,k+1/2}^{n+1/2-N_c} \\ &\quad + e_{1,k+1/2} E_{x,k+1/2}^n + e_{2,k+1/2} E_{x,k+1/2}^{n-1} \\ &\quad + \dots + e_{N_e-1,k+1/2} E_{x,k+1/2}^{n+2-N_e} \\ &\quad + \frac{\Delta_t}{\epsilon_0} [J_{s,x,k+1/2}^{n+1/2} + c_{1,s,x,k+1/2} J_{s,x,k+1/2}^{n-1/2} \\ &\quad \left. + \dots + c_{N_c,s,x,k+1/2} J_{s,x,k+1/2}^{n+1/2-N_c}] \right\} \end{aligned} \quad (19)$$

along with the discretized Faraday's law (7). The 2-D and 3-D dispersive algorithms follow in a similar fashion.

C. Comparison of Form-1 and Form-2 Algorithms

Comparing the Form-1 and Form-2 dispersive FDTD algorithms, it may at first appear that the choice of modeling complex permittivity $\epsilon(Z)$ versus complex conductivity $\sigma(Z)$ is merely definitional. However, there are significant differences in terms of stability² and their ability to model arbitrary dispersion functions.

Beginning with a permittivity function $\epsilon(\mathbf{r}, Z)$ of the form (6), and substituting (5) in (8), (8) may then be written in the form of (15) if we let

$$F_{k+1/2}(Z) = \frac{\epsilon_{k+1/2}(Z)}{\epsilon_0} (1 - Z^{-1}). \quad (20)$$

Then, equating (20) with (16), we find that $\epsilon(Z)$ and $\sigma(Z)$ are related by the bilinear³ form

$$\epsilon(Z) = \epsilon_0 \left[\epsilon_r + \frac{\Delta_t}{2\epsilon_0} \frac{(1 + Z^{-1})}{(1 - Z^{-1})} \sigma(Z) \right]. \quad (21)$$

²Stability relations for Form-1 and Form-2 algorithms are given in Section III. An air/muscle example is shown in Section V-B, whereby stability is improved by converting from a Form-1 to a Form-2 algorithm.

³Equation (21) may also be obtained from the continuous Maxwell equations in the Laplace transform (s) domain with the substitution $s = (2/\Delta_t) [(1 - Z^{-1})/(1 + Z^{-1})]$. This technique, known as the bilinear transform method, is commonly used in signal processing to convert analog filters to digital filters.

Equation (20) allows us to convert between Form-1 and Form-2 dispersive FDTD algorithms. Note, however, that (21) typically increases both the numerator and denominator polynomials by one degree, whether it is used to convert $\sigma(Z) \rightarrow \epsilon(Z)$ or $\epsilon(Z) \rightarrow \sigma(Z)$.

The Form-1 and Form-2 algorithms differ in their ability to match measured data with fixed-order polynomials in the numerator and denominator of $\epsilon(Z)$ and $\sigma(Z)$. This may be seen by examining their asymptotic properties. Setting $Z = e^{j\omega\Delta t}$ and examining the limit $\omega \rightarrow 0$, $\lim_{\omega \rightarrow 0} \Im m[\epsilon(Z)] = \omega \approx 0$ while $\lim_{\omega \rightarrow 0} \Re e[\sigma(Z)] = \text{Constant}$. The imaginary part of $\epsilon(Z)$ gives rise to an effective conductivity, which is zero in the dc limit. Many materials containing polar molecules, such as soil with any significant moisture content, have a nonzero dc conductivity. This finite dc conductivity may be modeled with a Form-1 algorithm with a pole on the real Z-axis, but is more naturally modeled with a Form-2 algorithm.

III. STABILITY CRITERIA

One of the most important considerations for a time-stepping algorithm such as the Form-1 and Form-2 dispersive FDTD algorithms is its stability. Stable physical phenomena can lead to unstable time-stepping algorithms, if algorithmic parameters are not set properly.

The stability formulas for the Form-1 and Form-2 dispersive FDTD algorithms are more complicated than the standard Yee algorithm due to the presence of the IIR filter arrays modeling $\epsilon(\mathbf{r}, Z)$ or $\sigma(\mathbf{r}, Z)$. As shown below, the stability formula is not determined by the poles of $\epsilon(Z)$ or $\sigma(Z)$ alone, but is complicated by the frequency of the wave component and the direction of wave propagation inside the FDTD grid. A similar stability analysis was presented recently for the differential equation methods employing the Debye and Lorentz formulas, although we derived our stability analysis independently.

For a homogeneous dispersive medium, the Ampere and Faraday laws may be written in Z-transform space as

$$\begin{aligned} E_{x,i,j,k+1/2}F(Z) &= \frac{\Delta t}{\epsilon_0} \left\{ \frac{1}{\Delta_y} [H_{z,i,j+1,k} - H_{z,i,j,k}] \right. \\ &\quad \left. - \frac{1}{\Delta_z} [H_{y,i,j,k+1} - H_{y,i,j,k}] \right\} \end{aligned} \quad (22)$$

$$\begin{aligned} (1 - Z^{-1})H_{y,i,j,k} &= \frac{\Delta t}{\mu_0} Z^{-1} \left\{ \frac{1}{\Delta_x} [E_{z,i+1/2,j,k} - E_{z,i-1/2,j,k}] \right. \\ &\quad \left. - \frac{1}{\Delta_z} [E_{x,i,j,k+1/2} - E_{x,i,j,k-1/2}] \right\} \end{aligned} \quad (23)$$

and similarly for the other field components. We now define the time-space transform, analogous to the temporal Z transform as

$$\tilde{H}_y(U, V, W, Z) = \sum_{m=1}^{N_i} \sum_{n=1}^{N_j} \sum_{p=1}^{N_k} H_{y,m,n,p}(Z) U^m V^n W^p \quad (24)$$

where (N_i, N_j, N_k) represent the number of space points in the FDTD lattice. The space-time transform of the other field components follow in a similar fashion. Equations (22) and (23) then become

$$\tilde{E}_x F(Z) = \frac{\Delta t}{\epsilon_0} \left[\frac{1}{\Delta_y} (V^{-1} - 1) \tilde{H}_z - \frac{1}{\Delta_z} (W^{-1} - 1) \tilde{H}_y \right] \quad (25)$$

and

$$(Z - 1) \tilde{H}_y = \frac{\Delta t}{\mu_0} \left[\frac{1}{\Delta_x} (1 - U) \tilde{E}_z - \frac{1}{\Delta_z} (1 - W) \tilde{E}_x \right]. \quad (26)$$

Similarly, we find that

$$(Z - 1) \tilde{H}_z = \frac{\Delta t}{\mu_0} \left[\frac{1}{\Delta_y} (1 - V) \tilde{E}_x - \frac{1}{\Delta_x} (1 - U) \tilde{E}_y \right]. \quad (27)$$

Substituting (26) and (27) into (25)

$$\begin{aligned} (Z - 1) \tilde{E}_x F(Z) &= \frac{\Delta_t^2}{\mu_0 \epsilon_0} \left[\frac{1}{\Delta_y^2} (V - 2 + V^{-1}) \tilde{E}_x \right. \\ &\quad + \frac{1}{\Delta_z^2} (W - 2 + W^{-1}) \tilde{E}_x \\ &\quad - \frac{1}{\Delta_x \Delta_y} (V^{-1} - 1)(1 - U) \tilde{E}_y \\ &\quad \left. - \frac{1}{\Delta_x \Delta_z} (W^{-1} - 1)(1 - U) \tilde{E}_z \right]. \end{aligned} \quad (28)$$

Equation (28) is essentially the time-space transform of the \hat{x} component of the vector wave equation. For a source-free homogeneous medium,⁴ $\nabla \cdot \mathbf{E} = 0$. Enforcing the divergenceless condition on the electric field, we can rewrite (28) as

$$\begin{aligned} (Z - 1) \tilde{E}_x F(Z) &= \frac{\Delta_t^2}{\mu_0 \epsilon_0} \left[\frac{1}{\Delta_x^2} (U - 2 + U^{-1}) + \frac{1}{\Delta_y^2} (V - 2 + V^{-1}) \right. \\ &\quad \left. + \frac{1}{\Delta_z^2} (W - 2 + W^{-1}) \right] \tilde{E}_x. \end{aligned} \quad (29)$$

Setting $\Delta_x = \Delta_y = \Delta_z$, we can rewrite the above as

$$(Z - 1)F(Z) = -4r^2 s^2 \quad (30)$$

where $r = \Delta_t / \Delta_z \sqrt{\mu_0 \epsilon}$ is the usual Courant number and

$$\begin{aligned} s^2 &= -\frac{1}{4} [(U - 2 + U^{-1}) + (V - 2 + V^{-1}) \\ &\quad + (W - 2 + W^{-1})]. \end{aligned} \quad (31)$$

From (24), the choice $U = e^{jk_x \Delta_x}$, $V = e^{jk_y \Delta_y}$, $W = e^{jk_z \Delta_z}$ corresponds to a plane wave with wave vector $\mathbf{k} =$

⁴Imposing the $\nabla \cdot \mathbf{E} = 0$ condition can also be justified by assuming plane wave propagation. If the FDTD algorithm is not stable for plane waves, the algorithm is not stable at all.

$\hat{x}k_x + \hat{y}k_y + \hat{z}k_z$. Then,

$$s^2 = \sin^2\left(\frac{k_x \Delta_x}{2}\right) + \sin^2\left(\frac{k_y \Delta_y}{2}\right) + \sin^2\left(\frac{k_z \Delta_z}{2}\right) \quad (32)$$

and it is clear that s^2 varies between zero and N_{dim} , where N_{dim} is the number of spatial dimensions in the simulation. Moreover, s depends on the frequency of the plane wave as well as the propagation direction, since $\mathbf{k} = \omega/c\hat{\mathbf{k}}$, where $\hat{\mathbf{k}} = \mathbf{k}/|\mathbf{k}|$.

Equation (30) is the discrete dispersion relation for the dispersive FDTD algorithm. The Z variable can be thought of as the gain of the E_x field from time step n to $n+1$. Stability requires that this gain satisfy $|Z| < 1$. Indeed, the discrete dispersion relation (30) will have several roots Z_l , $l = 1, 2, \dots, N_{\text{roots}}$ and the stability criterion for the dispersive FDTD algorithm requires that all of the roots Z_l to (30) satisfy $|Z_l| < 1$, $l = 1, 2, \dots, N_{\text{roots}}$.

We now consider certain special cases of (28). First, for a nondispersive, lossless medium, (16) gives

$$F(Z) = \epsilon_r(1 - Z^{-1}) \quad (33)$$

which reduces (30) to the usual Courant condition. For a Form-1 dispersive FDTD algorithm

$$F(Z) = \frac{\epsilon(Z)}{\epsilon_0}(1 - Z^{-1}). \quad (34)$$

For a Form-2 algorithm

$$F(Z) = \epsilon_r(1 - Z^{-1}) + \frac{\Delta_t}{2\epsilon_0}(1 + Z^{-1})\sigma(Z). \quad (35)$$

The above development of the stability criterion involved the source-free wave equation. The stability criterion given by (28) may also be derived from the poles of the source-driven wave equation. That is, if we add a free source current $J_{s,x}$ term to (22), and solve for the \tilde{E}_x field in terms of the free source current, we find that (28) is exactly the denominator polynomial containing the poles of the transfer function mapping $J_{s,x}$ to E_x .

Because the dispersion relation given by (28) is dependent on the model parameters, the maximum allowable value of r cannot be independently specified as it is with nondispersive media. Given a set of parameters for a second-order filter, the model can be tested for the stability of a particular choice of Δ_t and Δ_x by numerically finding the roots Z_l of the resulting polynomial when s^2 is swept from zero to N_{dim} , and ensuring that all roots are within the unit circle. For a Form-2 dispersion with $N_c = N_d - 1$, $N_{\text{roots}} = N_c + 2$.

IV. MINIMUM STORAGE IMPLEMENTATIONS

The goal of a computationally efficient FDTD algorithm is to perform accurate field calculations with minimal computer time and data storage requirements. The Form-1 and Form-2 dispersive FDTD algorithms require only a few additional

calculations per time step over the standard Yee algorithm. The modification of the standard Yee FDTD algorithm to include dispersion can significantly increase the storage requirement. We can minimize storage while maintaining the accuracy of the computations in two ways: 1) by choosing filter functions $\epsilon(Z)$ and $\sigma(Z)$ that contain just sufficient order to accurately model the wave physics for the frequency range of interest and 2) ensuring that the underlying computer algorithms do not contain unnecessary storage arrays.

Equation (9) appears at first to require storage of the previous time values of both the E_x and D_x fields. This observation was made by Joseph *et al.* [7], and later claimed by Luebbers [4] to be an advantage of the discrete convolution method over the differential-equation method. Similarly, (19) appears to require previous time storage of both the E_x and I_x fields. However, storage of both E_x and D_x or E_x and I_x is unnecessary, as was pointed out by Young [9] using a state-space approach. Other filter implementations may also be used that eliminate the need for storing both D_x and E_x or E_x and I_x .

We illustrate minimum storage implementations of the Form-1 and Form-2 dispersive algorithms. For the Form-1 algorithm, we assume for demonstration a second-order permittivity model. Dropping the space dependence for notational convenience, (9) becomes

$$E_x^{n+1} = \frac{1}{b_0} \left\{ \frac{1}{\epsilon_0} [D_x^{n+1} + a_1 D_x^n + a_2 D_x^{n-1}] - b_1 E_x^n - b_2 E_x^{n-1} \right\}. \quad (36)$$

A direct implementation of (36) above would require five storage arrays: E_x^n , E_x^{n-1} , D_x^{n+1} , D_x^n , and D_x^{n-1} . Storage is not required for E_x^{n+1} since E_x^{n+1} and E_x^n overwrite their previous values stored in E_x^n and E_x^{n-1} , respectively.

Equation (36) above may be rewritten in the form of the following three equations:

$$E_x^{n+1} = \frac{1}{b_0} \left[\frac{1}{\epsilon_0} D_x^{n+1} + W_1^n \right] \quad (37)$$

$$W_1^{n+1} = \frac{a_1}{\epsilon_0} D_x^{n+1} - b_1 E_x^{n+1} + W_2^n \quad (38)$$

$$W_2^{n+1} = \frac{a_2}{\epsilon_0} D_x^{n+1} - b_2 E_x^{n+1}. \quad (39)$$

The filter structure represented by (37)–(39) is known in the signal processing literature as a transposed direct Form-2 structure (not to be confused with a Form-2 dispersive FDTD algorithm). Note that (37)–(39) do not require previous time storage of the E_x or D_x field, but do require storage of W_1^l and W_2^l at time step n only, since W_1 and W_2 overwrite their own previous time values. Hence, (36) may be implemented using four storage arrays instead of five. As usual, $H_y^{n+1/2}$ must also be stored to implement Faraday's law. Although the savings does not appear significant, it would increase with higher order permittivity functions.

For the Form-2 dispersive FDTD algorithm, with a second-order conductivity model, (19) may be rewritten as

$$E_x^{n+1} = \frac{-1}{\epsilon_0} \left\{ I_x^{n+1/2} + c_1 I_x^{n-1/2} + c_2 I_x^{n-3/2} + e_1 E_x^n + e_2 E_x^{n-1} + e_3 E_x^{n-2} + \frac{\Delta t}{\epsilon_0} [J_{s,x}^{n+1/2} + c_1 J_{s,x}^{n-1/2} + c_2 J_{s,x}^{n-3/2}] \right\}. \quad (40)$$

The direct implementation of this equation requires five storage arrays: E_x^n , E_x^{n-1} , E_x^{n-2} , $I_x^{n-1/2}$, $I_x^{n-3/2}$, since only the three most recent E fields need be stored. Equivalently, $H_y^{n-1/2}$ and $H_y^{n-3/2}$ may be stored in place of $I_x^{n-1/2}$ and $I_x^{n-3/2}$, respectively, by (18). The minimal storage representation of the above is then

$$E_x^{n+1} = \frac{-1}{\epsilon_0} \left[I_x^{n+1/2} + \frac{\Delta t}{\epsilon_0} J_{s,x}^{n+1/2} + W_1^n \right] \quad (41)$$

$$W_1^{n+1} = c_1 \left(I_x^{n+1/2} + \frac{\Delta t}{\epsilon_0} J_{s,x}^{n+1/2} \right) + e_1 E_x^{n+1} + W_2^n \quad (42)$$

$$W_2^{n+1} = c_2 \left(I_x^{n+1/2} + \frac{\Delta t}{\epsilon_0} J_{s,x}^{n+1/2} \right) + e_2 E_x^{n+1} + W_3^n \quad (43)$$

$$W_3^{n+1} = e_3 E_x^{n+1}. \quad (44)$$

Four storage arrays are required in the above: E_x^{n+1} , W_1^{n+1} , W_2^{n+1} , and W_3^{n+1} . Again, one additional array is required for $H_y^{n+1/2}$ in implementing Faraday's law. Note that although $N_e = \max\{N_c + 1, N_d\}$, the Form-1 and Form-2 algorithms require the same storage for this case.

V. RELATION TO PREVIOUS METHODS

We now demonstrate that nearly all of the dispersive FDTD methods that have been considered in the literature may be cast as either Form-1 or Form-2 dispersive FDTD algorithms. The advantage to using our approach here is that special algorithms do not have to be developed for Debye, Lorentz, Drude or other media models. Once the Form-1 and Form-2 algorithms are developed, specific $\epsilon(Z)$ or $\sigma(Z)$ model parameters may be generated for a particular situation, without the need to develop a new computer code.

A. Differential Equation-Based Algorithms

We begin by considering the differential equation methods [5]–[8]. The idea behind these methods is that the time-domain dispersion relation is modeled as a differential equation relating $D(\mathbf{r}, t)$ to $E(\mathbf{r}, t)$

$$h_0 E(\mathbf{r}, t) + h_1 \frac{\partial}{\partial t} E(\mathbf{r}, t) + h_2 \frac{\partial^2}{\partial t^2} E(\mathbf{r}, t) + \dots = D(\mathbf{r}, t) + g_1 \frac{\partial}{\partial t} D(\mathbf{r}, t) + g_2 \frac{\partial^2}{\partial t^2} D(\mathbf{r}, t) + \dots \quad (45)$$

TABLE I
COEFFICIENTS OF FORM-1 DEBYE AND LORENTZ MEDIUM-DISPERSIVE FDTD ALGORITHMS CONSIDERED BY JOSEPH *et al.* [7]

Debye Model	Lorentz Model
$b_0 = \frac{\epsilon_s \Delta_t + 2\tau \epsilon_\infty}{\Delta_t + 2\tau}$	$b_0 = \frac{\omega_p^2 \Delta_t^2 \epsilon_s + 2\delta \Delta_t \epsilon_\infty + 2\epsilon_0}{\omega_p^2 \Delta_t^2 + 2\delta \Delta_t + 2}$
$b_1 = \frac{\epsilon_s \Delta_t - 2\tau \epsilon_\infty}{\Delta_t + 2\tau}$	$b_1 = \frac{-4\epsilon_\infty}{\omega_p^2 \Delta_t^2 + 2\delta \Delta_t + 2}$
$a_1 = \frac{\Delta_t - 2\tau}{\Delta_t + 2\tau}$	$b_2 = \frac{\omega_p^2 \Delta_t^2 \epsilon_s - 2\delta \Delta_t \epsilon_\infty + 2\epsilon_0}{\omega_p^2 \Delta_t^2 + 2\delta \Delta_t + 2}$
	$a_1 = \frac{-4}{\omega_p^2 \Delta_t^2 + 2\delta \Delta_t + 2}$
	$a_2 = \frac{\omega_p^2 \Delta_t^2 - 2\delta \Delta_t + 2}{\omega_p^2 \Delta_t^2 + 2\delta \Delta_t + 2}$

In Fourier space, the permittivity function can be seen as a Padé approximation in $j\omega$:

$$\epsilon(\omega) = \frac{h_0 + h_1(j\omega) + h_2(j\omega)^2 + \dots}{1 + g_1(j\omega) + g_2(j\omega)^2 + \dots}. \quad (46)$$

The differential equation is then discretized, using either forward, backward, or central difference approximations for the partial derivatives. There is some discretization error that is introduced by the discretization process, resulting in error between the discrete spectrum and the continuous spectrum. This discretization error has also shown to give rise to a misrepresentation of the actual relaxation time of the medium and to produce phase errors.

One of the benefits of using this technique is that (46) seems to offer flexibility in fitting arbitrary permittivity functions. However, we must keep in mind that when (45) is discretized, a difference equation will result, and the actual discretized permittivity function can be represented as a Padé model in Z-transform space. Hence, for fitting arbitrary media, the differential equation methods offer no advantage over our technique, and it is preferable to optimize the Padé model coefficients in Z-transform space to avoid introducing discretization error.

We have developed the equivalent Form-1 Padé models for $\epsilon(Z)$ directly from the difference equations presented by Joseph *et al.* [7]. For the first-order Debye algorithm, the permittivity function is of the form

$$\epsilon(\omega) = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + j\omega\tau}. \quad (47)$$

For the second-order Lorentz algorithm

$$\epsilon(\omega) = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{\omega_0^2 + j2\omega\delta - \omega^2}. \quad (48)$$

The coefficients of the Form-1 dispersive FDTD algorithms corresponding to (47) and (48) are listed in Table I.

B. Z-Transform Algorithms

The Z-transform algorithms [10] are the most similar to our formulation of all dispersive FDTD algorithms. The essence of the Z-transform method is that one begins with a specific frequency-domain analytic form for the dispersion relation,

such as the Debye or Lorentz model, which is inverse Fourier transformed to obtain a time-dependent permittivity function $\epsilon(t)$. The time-dependent permittivity is then sampled by setting $t = n\Delta_t$, and its Z-transform is computed analytically.

This procedure is exactly analogous to the design of digital filters via the impulse invariance method. One of the advantages to the impulse-invariance technique is that a stable continuous filter always results in a stable discrete filter $\epsilon(Z)$ [18]. However, note that a stable permittivity filter does not guarantee that the resulting FDTD algorithm will be stable. Aliasing problems encountered in the design of high-pass digital filters using impulse invariance are avoided because of the stringent FDTD $\lambda/10$ minimum grid sampling requirement, which ensures that the FDTD algorithms always operate far below the Nyquist frequency.

Sullivan considered a Debye model with an added conductivity term as an approximate model for muscle tissue in the frequency range 10–915 MHz [10]. The model is given as

$$\epsilon(\omega) = \epsilon_r + \frac{\epsilon_1}{1 + j\omega t_1} + \frac{\sigma}{j\omega\epsilon_0} \quad (49)$$

where $\epsilon_r = 50$, $\epsilon_1 = 110$, $\sigma = 0.62$ S/m, $t_1 = 5.88$ ns.

From (49), we apply the impulse-invariance procedure [18] to obtain

$$\epsilon(Z) = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2}}{(1 - Z^{-1})[1 - e^{-\Delta_t/t_1} Z^{-1}]} \quad (50)$$

where $b_0 = \epsilon_r + \epsilon_1 \Delta_t/t_1 + \Delta_t \sigma/\epsilon_0$, $b_1 = -[\epsilon_r(1 + e^{-\Delta_t/t_1}) + \epsilon_1 \Delta_t/t_1 + \Delta_t \sigma/\epsilon_0 e^{-\Delta_t/t_1}]$, $b_2 = \epsilon_r e^{-\Delta_t/t_1}$. We have tried implementing this permittivity filter as a Form-1 dispersive FDTD algorithm, but have found that the algorithm is unstable due to the fact that the denominator polynomial of (50) contains a pole on the unit circle. Sullivan was able to implement a stable algorithm by isolating the pole on the unit circle in a parallel filter.

An alternate, more convenient way to implement the dispersive model given by (50) is to convert the Form-1 algorithm to an equivalent Form-2 algorithm, since the conversion factor contains a pole at $Z = 1$. Employing (20)

$$F(Z) = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2}}{1 - e^{-\Delta_t/t_1} Z^{-1}}. \quad (51)$$

Note that the pole at $Z = 1$ in (51) has been cancelled by a zero at $Z = 1$ that is inherent in the Form-2 algorithm. The Form-2 algorithm with $F(Z)$ given by (51) was found to be numerically stable.

We simulated wave propagation in a medium consisting of a 1-D air/muscle interface. For this problem we chose a 1000 cell simulation and chose time and space step sizes of $\Delta_t = 6.65$ ps and $\Delta_z = 2.0$ mm. Sullivan [10] reported using a grid-space size of $\Delta_z = 10.0$ mm. However, this space step size is too large for this problem since it does not satisfy the $\Delta_z < \lambda_{\min}/10$ grid-dispersion requirement for this medium. For example, at 915 MHz, the permittivity may be determined from (49) to be $\epsilon(\omega) = 50.096 - j15.448$, which corresponds to a wavelength of $\lambda = 44.78$ mm in the medium.

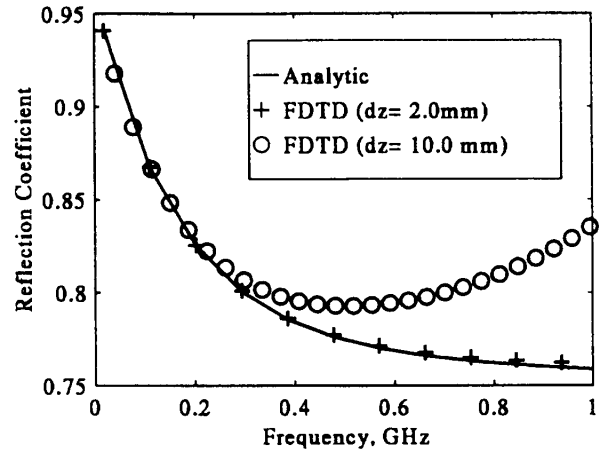


Fig. 2. Magnitude of reflection coefficient for air/muscle interface computed using analytic solution and Form-2 dispersive FDTD algorithm. Solid line is analytic solution, Form-2 dispersive FDTD algorithm results are represented by (+) symbols for $\Delta_z = 2.0$ mm and (o) for $\Delta_z = 10.0$ mm. Sullivan's algorithm [10] gives identical results to Form-2 dispersive algorithm.

The reflection coefficient for the air/muscle interface computed using the Form-2 dispersive FDTD algorithm is shown in Fig. 2 and is compared against the analytic frequency-domain normal-incidence reflection coefficient, where the permittivity was determined from (49). Note in Fig. 2 that we calculated the reflection coefficient for both $\Delta_z = 2.0$ mm and $\Delta_z = 10.0$ mm. The reflection coefficient with $\Delta_z = 2.0$ mm is very close to the exact result, while the result for $\Delta_z = 10.0$ mm contains significant error at the higher frequencies. We have implemented Sullivan's model for this case, and it yields identical results.

C. Discrete Convolution Algorithms

In the discrete convolution approach, the dispersive FDTD problem is solved by first assuming a particular analytical form for the permittivity function $\epsilon(\omega)$ in the frequency-domain, and implementing a discrete convolution using only a very small time history. We examine two discrete convolution approaches: one based on the Debye model [3], and the other based on a multiple Lorentz approximation [4].

The algorithm involving a first-order Debye model [3] may be implemented exactly using our general framework. The algorithm that involves multiple Lorentz poles [4] involves a complex filter operation. We have successfully modeled the air/fourth-order medium using a Form-1 dispersive algorithm with a real filter. However, we shall defer discussion of this example to another paper since the implementation is somewhat involved and the example does not illustrate how to model arbitrary media.

We have converted first-order Debye model algorithm into the form of our Form-2 dispersive algorithm, with the parameters of $F(Z)$ listed in Table II. In the limit $\Delta_t \ll t_0$, which is generally true when the dispersion occurs within the operating frequency band of the FDTD algorithm, we have that $\chi_0 \approx (\epsilon_s - \epsilon_\infty) \Delta_t/t_0$. The above algorithm is then identical to the algorithm that we used to model the air/muscle interface without the conductivity term.

TABLE II
PARAMETERS FOR $F(Z)$ FOR FORM-2 DISPERSIVE FDTD ALGORITHM
CORRESPONDING TO LUEBBER'S FIRST-ORDER DEBYE MODEL ALGORITHM [3]

$$\begin{aligned} b_0 &= \chi_0 + \epsilon_\infty & a_1 &= -e^{-\Delta t/t_0} \\ b_1 &= -\left[\chi_0 + \epsilon_\infty(1 + e^{-\Delta t/t_0})\right] & \chi_0 &= (\epsilon_s - \epsilon_\infty)\left[1 - e^{-\Delta t/t_0}\right] \\ b_2 &= \epsilon_\infty e^{-\Delta t/t_0} \end{aligned}$$

TABLE III
MODEL COEFFICIENTS FOR FORM-2 DISPERSIVE FDTD ALGORITHM MODELING
PUERTO RICO CLAY LOAM WITH $D = 1.4$ g/cc AND $M = 10\%$

$$\begin{aligned} d_0 &= .0901332 & c_1 &= -.87 \\ d_1 &= -.0792821 & c_2 &= .06 \\ d_2 &= -.00947176 & \epsilon_r &= 6.2 \end{aligned}$$

VI. DISPERSIVE SOIL MODELING

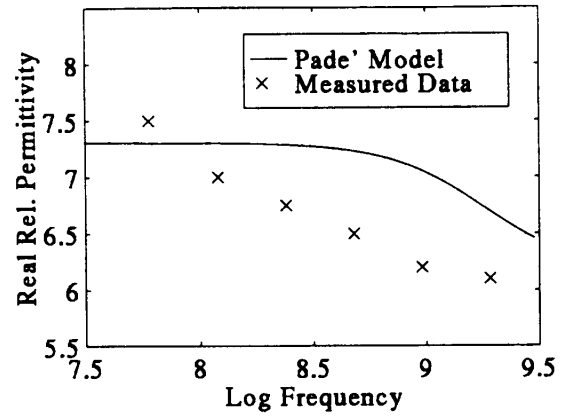
We now consider the problem of modeling wave propagation in an arbitrary dispersive dielectric based on measured data. In particular, we investigate the modeling of wave propagation in dispersive, lossy soils. The permittivity and conductivity of soils at microwave frequencies is known to vary with density D and moisture M as well as frequency. Representative variations for San Antonio and Puerto Rico clay loam, as functions of D , M , and f , are given in [19]. This soil medium is notoriously difficult to model with a low-order Debye, Lorentz, or Drude model. Finding parameters which provide adequate simultaneous agreement for both the real and imaginary permittivity functions is problematic.

To model the wave propagation, we use a Form-2 dispersive FDTD algorithm that is second-order in both numerator and denominator. A conductivity function of the form given by (13) is used with $N_d = 3$, $N_c = 2$. The specific model coefficients that we have obtained for Puerto Rico clay loam with $D = 1.4$ g/cc and $M = 10\%$ are given in Table III. The time and space step values used are $\Delta t = 20$ ps and $\Delta x = \Delta y = \Delta z = 5$ mm. The design problem of obtaining the model coefficients will be discussed in a following paper.

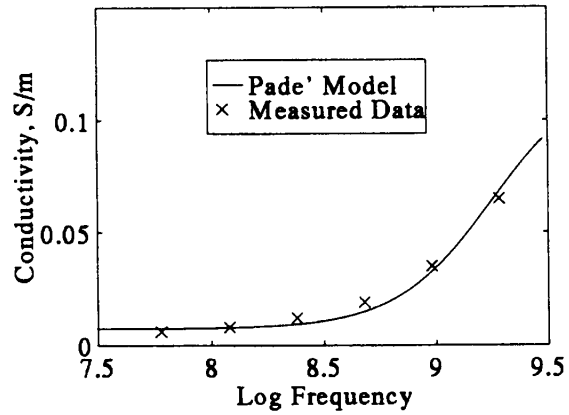
Fig. 3 shows the real components of the effective relative permittivity (a) and conductivity (b) for the clay loam based on the Padé approximation given by (6) and the data in Table I (solid lines), where we have substituted $Z = e^{j2\pi f \Delta t}$ in (13). The data measured by Hipp is indicated by the crosses (x) in Fig. 3.

It is interesting to note in Fig. 3 that although we specified a real frequency-independent relative permittivity ϵ_r (see Table III), there is some variation in the Padé model's effective permittivity. The reason is that the conductivity filter $\sigma(Z)$ has an imaginary component that gives rise to an effective permittivity. This is a direct result of the fact that it is not possible to design a causal, stable IIR filter with zero phase [18].

To illustrate that our Form-2 dispersive FDTD algorithm is accurate and stable for 3-D wave propagation, we investigated the fields radiated by a short electric dipole. The electric field



(a)



(b)

Fig. 3. (a) Real component of relative permittivity. (b) Conductivity for clay loam with density $D = 1.4$ g/cc and moisture $M = 10\%$. Padé model (solid) is compared to measured data (x) [19].

radiated by the dipole is given as

$$\mathbf{E}(\mathbf{r}, \omega) = -j\omega\mu\left(\bar{\mathbf{I}} + \frac{\nabla\nabla}{k^2}\right) \cdot \hat{\mathbf{z}} I l \frac{e^{-jk r}}{4\pi r} s(\omega) \quad (52)$$

where $\bar{\mathbf{I}}$ is the identity operator and $s(\omega)$ is the source frequency spectrum. The time-domain analytic solution for the 3-D radiated fields from the small dipole was generated by numerically inverse Fourier transforming the field computing using (52). The source function $s(t)$ was derived from a Kaiser wavelet with a passband extending from 30.0 MHz to 2.5 GHz.

Fig. 4(a) shows the E_y component of the 3-D Form-2 FDTD solution and the analytic solution for the fields radiated by the short dipole located at the origin, and observed at $(x = 1.10$ m, $y = .420$ m, $z = .420$ m) in the clay loam. Note in the figure that only the time window [10 ns, 18 ns] is shown, since the fields are insignificant outside this region. Fig. 4(b) shows the difference between the numerical and analytic solutions. The E_x and E_z components were computed and have similar characteristics, but are not shown here.

VII. CONCLUSION

We have developed a new method for modeling wave propagation in arbitrary frequency-dispersive dielectric media using a FDTD methodology. This new method involves

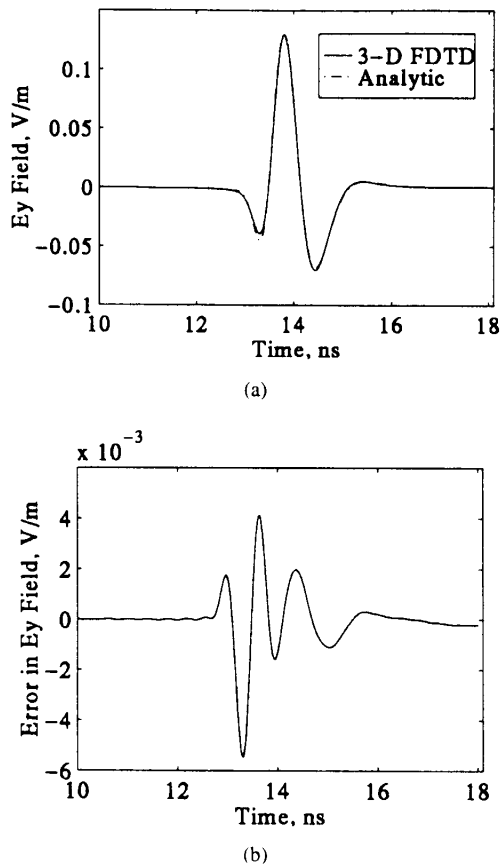


Fig. 4. E_y component of electric field radiated by a small z -directed electric dipole in dispersive clay loam with $D = 1.4$ g/cc and $M = 10\%$. Dipole is located at origin, and E_y computed at $(x = 1.10$ m, $y = .420$ m, $z = .420$ m). (a) Numerical solution (solid) computed using 3-D Form-2 dispersive FDTD and analytic solution (dashdot) superimposed. (b) E_y field error for numerical solution.

modeling the simulation region as a spatial array of IIR digital filters, whereby a Padé approximation in Z-transform space is used to model arbitrary permittivity and conductivity functions. Two algorithms were developed, one involving a Padé approximation for $\epsilon(Z)$ (Form-1), and the other for $\sigma(Z)$ (Form-2). Differences between the Form-1 and Form-2 algorithms were discussed in terms of their stability and ability to model arbitrary data. Stability formulas were derived for both the Form-1 and Form-2 algorithms. Minimum storage implementations were also presented.

Our algorithms were compared to several previous methods for FDTD modeling of dispersive media. These algorithms were classified as discrete convolution methods, differential equation methods, and Z-transform methods. We have demonstrated that all of the previously published dispersive FDTD algorithms of which we are aware, with the exception of the discrete convolution algorithm involving a complex filter operation, may be cast in our more general framework using their exact implementation. In this sense, the Form-1 and Form-2 dispersive FDTD algorithms may be seen as a generalization of previous algorithms.

The main advantage of our new formulations is in their ability to more directly model arbitrary dispersive dielectrics based on measured data. We demonstrated that the permittivity

and conductivity filters lead directly to FDTD implementations without introducing discretization error. An accurate model for Puerto Rico clay loam was presented and demonstrated in a 3-D FDTD simulation. The radiation from a small vertical electric dipole was considered, and the FDTD solution compared against the analytic solution. The agreement was found to be excellent.

ACKNOWLEDGMENT

The authors would like to thank D. Silevitch for his assistance with the stability analysis.

REFERENCES

- [1] K. L. Shlager and J. B. Schneider, "A selective survey of the finite-difference time-domain literature," *IEEE Antennas Propag. Mag.*, vol. 37, no. 4, pp. 39–57, Aug. 1995.
- [2] K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Trans. Antennas Propag.*, vol. AP-14, pp. 302–307, May 1966.
- [3] R. Luebbers, F. P. Hunsberger, K. S. Kunz, R. B. Standler, and M. Schneider, "A frequency-dependent finite-difference time-domain formulation for dispersive materials," *IEEE Trans. Electromagn. Compat.*, vol. 32, pp. 222–227, Aug. 1990.
- [4] R. J. Luebbers and F. Hunsberger, "FDTD for Nth-order dispersive media," *IEEE Trans. Antennas Propag.*, vol. 40, pp. 1297–1301, Nov. 1992.
- [5] T. Kashiwa, N. Yoshida, and I. Fukai, "A treatment by the finite-difference time-domain method of the dispersive characteristics associated with orientation polarization," *Inst. Electron. Inform. Communicat. Eng. Trans.*, vol. E73, no. 8, pp. 1326–1328, June 1990.
- [6] T. Kashiwa and I. Fukai, "A treatment by the FD-TD method for the dispersive characteristics associated with electronic polarization," *Microwave Opt. Tech. Lett.*, vol. 3, no. 6, pp. 203–205, 1990.
- [7] R. M. Joseph, S. C. Hagness, and A. Taflov, "Direct time integration of Maxwell's equations in linear dispersive media with absorption for scattering and propagation of femtosecond electromagnetic pulses," *Opt. Lett.*, vol. 16, no. 18, pp. 1412–1414, 1991.
- [8] O. P. Gandhi, "A frequency-dependent finite-difference time-domain formulation for general dispersive media," *IEEE Trans. Microwave Theory Tech.*, vol. 41, pp. 658–665, Apr. 1993.
- [9] J. L. Young, "Propagation in linear dispersive media: Finite difference time-domain methodologies," *IEEE Trans. Antennas Propag.*, vol. 43, pp. 422–426, Apr. 1995.
- [10] D. M. Sullivan, "Frequency-dependent FDTD methods using Z transforms," *IEEE Trans. Antennas Propag.*, vol. 40, pp. 1223–1230, Oct. 1992.
- [11] ———, "Nonlinear FDTD formulations using Z transforms," *IEEE Trans. Microwave Theory Tech.*, vol. 43, pp. 676–682, Mar. 1995.
- [12] R. J. Luebbers, F. Hunsberger, and K. Kunz, "A frequency-dependent finite-difference time-domain formulation for transient propagation in plasmas," *IEEE Trans. Antennas Propag.*, vol. 39, pp. 29–34, Jan. 1991.
- [13] F. Hunsberger, R. J. Luebbers, and K. Kunz, "Finite-difference time-domain analysis of gyrotropic media—I: Magnetized plasma," *IEEE Trans. Antennas Propag.*, vol. 40, pp. 1489–1495, Dec. 1992.
- [14] R. Pontalti, L. Cristoforetti, R. Antolini, and L. Cescatti, "A multi-relaxation (FD)²-TD method for modeling dispersion in biological tissues," *IEEE Trans. Microwave Theory Tech.*, vol. 42, pp. 526–528, Mar. 1994.
- [15] D. M. Sullivan, "A frequency-dependent FDTD method for biological applications," *IEEE Trans. Microwave Theory Tech.*, vol. 40, pp. 532–539, Mar. 1992.
- [16] P. G. Petropoulos, "Stability and phase error analysis of FD-TD in dispersive dielectrics," *IEEE Trans. Antennas Propag.*, vol. 42, pp. 62–69, Jan. 1994.
- [17] D. M. Sullivan, "Z-Transform theory and the FDTD method," *IEEE Trans. Antennas Propag.*, vol. 44, pp. 28–34, Jan. 1996.
- [18] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, 2nd ed. New York: Macmillan, 1992.
- [19] J. E. Hipp, "Soil electromagnetic parameters as functions of frequency, soil density, and soil moisture," *Proc. IEEE*, vol. 62, pp. 98–103, Jan. 1974.



William H. Weedon (S'86-M'94) was born in [redacted]. He received the B.S. (*summa cum laude*) and M.S. degrees in electrical engineering from Northeastern University, Boston, MA, in 1989 and 1990, respectively, and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, in 1994.

From 1985 to 1989, he worked as a co-op student at Raytheon Submarine Signal Division, Portsmouth, RI, in surface systems development. During the Master's degree program at Northeastern (1989-1990), he worked on the problem of selecting optimal measurement configurations for microwave reflectometry. He spent the summer of 1990 at the MITRE Corporation, Bedford, MA, working on phased-array radar tracking. From 1991 to 1994, he was a Research Assistant at the University of Illinois, where he developed a broadband microwave inverse-scattering imaging system, including an 11-element Vivaldi antenna array—a network analyzer programmed to operate as a step-frequency radar—and computational inverse-scattering algorithms. He also developed several computational electromagnetic algorithms, including a 3-D FDTD code employing the Berenger perfectly matched layer (PML) implemented on the Connection Machine CM-5 massively parallel supercomputer and a 2.5-D spectral Lanczos decomposition method (SLDM) code. From 1994 to 1996, he was a Visiting Assistant Professor in the Electrical and Computer Engineering Department at Northeastern University, where he taught undergraduate electrical engineering and worked as a Research Scientist in the Center for Electromagnetics Research. His research focused on ground-penetrating radar modeling and data interpretation. He was the Technical Director on research grants from the United States Department of Energy, United States Army Research Laboratory, and the New England Transportation Consortium. Currently, he is a Technical Consultant specializing in electromagnetic and acoustic sensor modeling and data interpretation.

Dr. Weedon is a member of Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi honor societies.



Carey M. Rappaport (SM'96) received the S.B. degree in mathematics, the S.B., S.M., and E.E. degrees in electrical engineering, all in 1982, and the Ph.D. degree in electrical engineering, in 1987, all from the Massachusetts Institute of Technology (MIT), Cambridge.

He has worked as a Teaching and Research Assistant at the Massachusetts Institute of Technology (MIT) from 1981 until 1987, and (during the summers) at COMSAT Labs, Clarksburg, MD, and the Aerospace Corporation, El Segundo, CA. He joined the faculty at Northeastern University, Boston, MA, in 1987, and has been an Associate Professor of electrical and computer engineering since 1993. During the fall of 1995 he was a Visiting Professor of electrical engineering at the Electromagnetics Institute of the Technical University, Lyngby, Denmark, as part of the W. Fulbright International Scholar Program. He has authored more than 80 technical papers in the areas of microwave antenna design, electromagnetic scattering computation, and bioelectromagnetics, and has received two reflector antenna patents and one biomedical device patent.

Dr. Rappaport was awarded the IEEE Antenna and Propagation Society's H. A. Wheeler Award for the best applications paper of 1985. He is a member of Sigma Xi and Eta Kappa Nu professional honorary societies.

A Common Framework of Subsurface Sensing and Imaging

M. B. Silevitch*, S. W. McKnight, and C. Rappaport

Center for Subsurface Sensing and Imaging Systems (CenSSIS)
Northeastern University, Boston, MA 02115

ABSTRACT

Subsurface sensing and imaging problems arise in a variety of contexts: underground, underwater, inside the human body, and inside a cell or a collection of cells. All of these problems require reconstruction of internal structures or functions from a highly distorted probe or wave sampled outside of an obscuring surface. There is an emerging common framework of physics-based signal processing which will allow progress in any of these areas to be applied to create advances in the other areas. The recognition of the essential similarity of these problems and the development of the common framework is a key to the next generation of environmental and biomedical imaging systems.

Keywords: Subsurface sensing and imaging, tomography, inverse problems, underground, underwater, and medical imaging, 3D microscopy.

1. INTRODUCTION

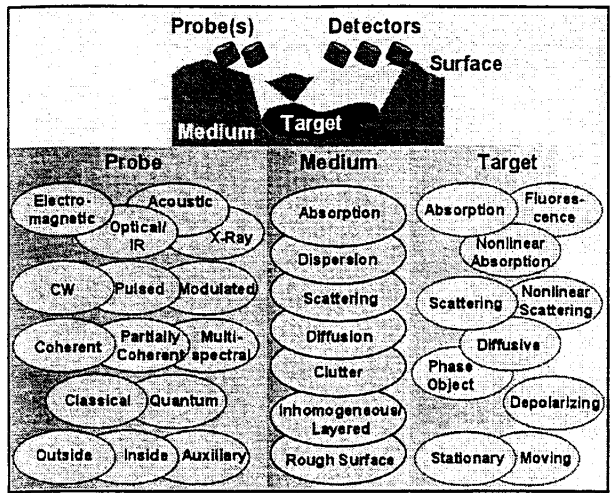
The problem of imaging under a surface arises in a wide variety of contexts, and these problems are among the most difficult and intractable system challenges known. Spread one hundred plastic landmines on top of a farmer's field and they can be safely removed in hours by a worker with a minimum of training. Bury them under one centimeter of soil, and you have a problem that has been the subject of intensive research for over half a century and remains far from solved. State-of-the-art inductive sensors in the hands of an experienced operator can detect non-metallic mines from the signal received from the firing pin and other small metal parts. In typical operation, however, over 300 false alarms are recorded for every mine identified, each requiring lengthy and delicate examination. In the end, operational mine detection systems have little, if any, advantage over probing each square centimeter of the ground with a titanium rod, a process that can clear a field at a rate of 1 meter by 25 meters of ground per person per day. No one has any idea how the three million landmines buried in Bosnia or the 10 million in Cambodia can be removed at any reasonable cost.¹

De-mining, in common with nearly all subsurface sensing and imaging problems, is an *information* problem. If we knew where the mines were buried, world-wide humanitarian de-mining would require relatively few physical resources. Yet in an Information Age, when the cost of computation and communications is reduced by a sizable fraction each year, the full potential of applying our exponentially expanding information technology sector to subsurface problems has not been realized because of lack of equivalent progress in subsurface detection and identification.

In addition to the technical problems of probes and processing that we will discuss below, we identify two major systems obstacles to progress in subsurface sensing and imaging:

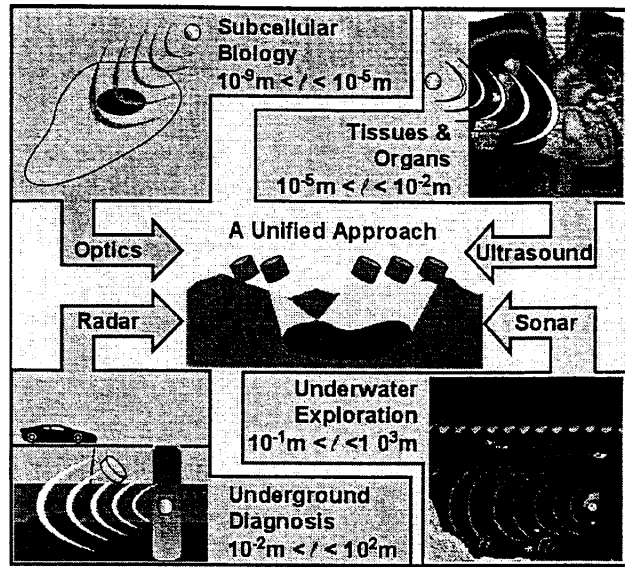
- 1) the problems of sensor design, modeling, image processing, and recognition have been compartmentalized, viewed as separate disciplines rather than as integrated parts of a *system* optimization problem.
- 2) the subsurface problems in different media and different length scales are commonly viewed as unrelated problems and addressed with *ad hoc* solutions. Lessons learned in one subsurface technology are rarely applied to other problems, and no overarching theory exists to identify fundamental limitations, predict what can be detected and the optimal way to do it.

* E-mail: msilevit@lvnx.neu.edu; CenSSIS Tel. (617)373-5110; WWW: <http://www.censsis.neu.edu>



NU-011

Figure 1. Subsurface problems can be classified on the basis of the probe-medium-target interactions.



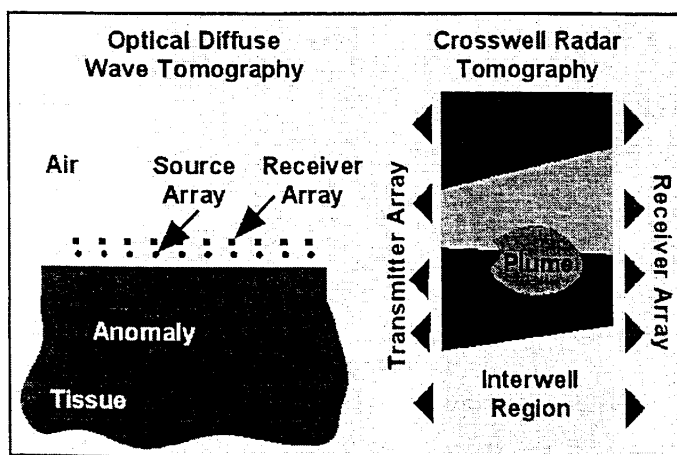
NU-001

Figure 2. A unified physics-based approach can unite subsurface problems from many domains at different length scales.

The subsurface sensing and imaging problem is to extract information about a subsurface target from scattered and distorted waves received above the surface. Imaging techniques, whether ultrasound sensors in tissue or electromagnetic probes in soil, can be described by the properties of probe wave, the wave propagation characteristics of the medium and surface, and the nature of target/probe interaction as shown in Figure 1.

The framework of Figure 1 describes not only underground imaging, but also underwater imaging, medical imaging inside the body, and 3D biological microscopies inside a cell or collection of cells. A unified theory of subsurface sensing and imaging, as illustrated in Figure 2, should encompass all of these applications and permit progress in one domain to be transparently applied in other domains with similar elements in the taxonomy of Figure 1.

For example, diffusive wave optical imaging for medical diagnosis and crosswell radar/EMI tomography for geophysical exploration both involve extracting an image of, or information about, anomalous regions (e.g., diabetic lesions under the skin or oil-bearing rock formations under the ground; see Figure 3). Although the problems occur on vastly different length scales, both require solution of the frequency-domain diffusion equation in the presence of an inhomogeneous, layered medium, and a need to filter large data sets from multiple transmitters and receivers that are, nevertheless, sparse compared to the information set sought. Attacking these two problems within the same framework allows the synergy of the two solutions to be exploited. Thus, even the critical differences between the two problems (lossy vs. lossless propagation, Poisson vs. Gaussian noise statistics, the diffusion equation as a limit of the radiative transfer equation vs. the diffusion equation derived by neglecting the displacement current in Maxwell's Equations) become a basis for more complete understanding of the unified problem, rather than just an obstacle to applying the same specialized algorithm to each problem.



NU-014

Figure 3. The physical/mathematical framework of diverse problems can be very similar.

2. PHYSICS-BASED SIGNAL PROCESSING

It is rarely the case that we cannot get any information from the subsurface region. The concealing media, while not transparent, can usually be penetrated to a considerable depth by a variety of acoustic and electromagnetic wave probes. The problem is that the target signal is distorted by complex absorption, dispersion, diffraction, and refraction of the wave through the media and obscured by surface reflection, subsurface clutter, and scattered energy from unknown inhomogeneities on many scales. The signal received, y , depends on the target information x and various signal-dependent clutter and nuisance parameters z through the function C which describes the physics of the probe-wave generation, propagation, and target and clutter interaction:

$$y = C(x, z) + n \quad (1)$$

The inverse problem of un-encoding the signature of the target object x from the received signal y in the presence of unpredictable clutter signals due to z and noise, n , is the challenge of subsurface sensing and imaging.

Since the mapping from the target to the sensor depends on unknown information about the subsurface media and target, the inversion from the scattered wave to the target properties is a nonlinear mathematical problem.^{2,3} The use of appropriate physical models of the probe/surface/media/target/receiver interaction (C in Equation 1) to assist in the solution of that inverse problem is what is referred to as physics-based signal processing (PBSP). PBSP has been identified in a seminal 1998 review article as a key to progress in image formation in complex media.⁴ Physics-based reasoning through the entire image understanding process and goal-directed processing will produce algorithms which are robust to modeling errors and generate accurate reconstructions of the critical information.

The fundamental problem of subsurface sensing and imaging is to differentiate the target of interest from irrelevant clutter and scatter, to distinguish a landmine from roots, stones, shell-casing, or ground-surface reflections. In the pulse-reflection ground-penetrating radar (GPR) simulation in Figure 4, for example, the signal from the plastic cylinder in the lower figure is obscured by the rough-surface reflection in the upper figure. The task is to extract the signal from the complex scattered field of random surface irregularities. In principle, if the surface profile and the soil dielectric properties were precisely known, one could subtract the background from the received signal to extract the target signature, but a full 3D calculation of the scattered field for a single pulse could take on the order of 10 hours on a 450 MHz desktop Pentium computer. Problems where the target distinguishing features are comparable to the clutter size, such as demining, are among the most challenging subsurface problems that exist.

One of the primary differentiating features is *shape*. Since resolution in the far-field is limited to order of the wavelength, it is desirable to use probes with wavelengths smaller than the size of identifying features. Unfortunately in most subsurface modalities, absorption increases with increasing frequency (smaller wavelengths). For example, the attenuation of medical ultrasound increases at the rate of 1dB/cm for every megahertz of frequency. Arterial plaque, which can be resolved by inserting catheters containing 30 MHz ($\lambda = 50 \mu\text{m}$) ultrasonic probes, cannot be noninvasively imaged from outside the body because the

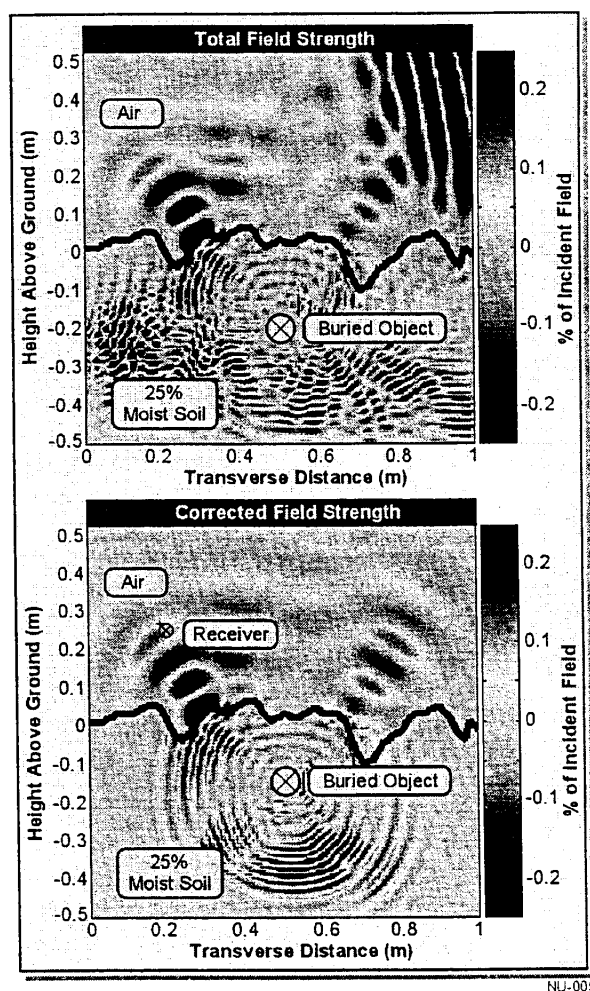


Figure 4. Clutter from rough-surface reflection in the top frame obscures the signature of the buried object in the bottom frame in this pulse GPR simulation.

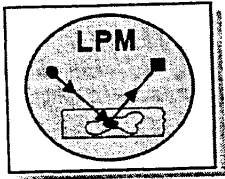
attenuation is too severe (about 30 dB per centimeter of depth at 30 MHz). This range/resolution trade-off is a fundamental limitation on many subsurface modalities including underground seismic imaging and underwater sonar imaging.

Alternatively, probes which are sensitive to target *material properties*, such as to material spectral response (color), conductivity, or magnetic susceptibility, can offer advantages for target differentiation. For example, medical imaging probes such as magnetic resonance imaging or nuclear medicine molecular tags which are sensitive to target chemistry can be used to differentiate targets on the basis of *physiology* (functioning) instead of *anatomy* (structure). Imaging the subtle physiological differences between cancerous cells and normal cells would be a medical breakthrough.

Nonlinear material properties are used for subsurface discrimination in two-photon microscopy^{5,6,7} or ultrasonic harmonic imaging. Harmonic imaging can yield diagnostically useful information on the 25% to 30% of the population that cannot be imaged well by ultrasound due to high clutter levels, distortions, and artifacts. Although harmonic imaging is already commercially available, the physical mechanisms behind it are poorly understood.

3. INFORMATION EXTRACTION STRATEGIES

Despite the bewildering variety of imaging modalities and techniques covered in the Figure 1 taxonomy, subsurface problems can be organized into a relatively small number of *information extraction strategies* which use similar algorithmic tools. Three broad information extraction strategies are discussed here.



Localized probing and mosaicking (LPM) concentrates the probe wave on a local subsurface region by focusing or time-gating and then assembles these individual pieces of information into an information mosaic. Common to these techniques are problems of concentration, aberration, and

registration which may use tools as simple as a lens or as complex as three-dimensional image matching and reconstruction. For example, medical reflection ultrasound and confocal microscopy both collect scattered energy from a subsurface target voxel. In both cases, precise focusing assumes a uniform homogeneous wave velocity, rarely the case in subsurface imaging, and resulting aberrations impede accurate imaging. LPM techniques are subject to obstruction by opaque objects (e.g., bones), and because reflection geometries are sensitive to high spatial frequencies (interfaces) LPM techniques are poor at detecting low-contrast or phase-only objects.

Wide-scan, high resolution LPM imaging usually requires mosaicking of multiple frames. Figure 5 shows an image of an ocean-floor hydrothermal vent assembled in this way by our collaborators at Woods Hole Oceanographic Institution. Errors in image registration and composition techniques contribute to errors on the scale of meters in the large-scale representation of imaged objects, reducing their utility in quantitative oceanography. The use of high-resolution sonar maps to register the optical images is a multi-modality path to the desired capability for high-resolution mapping of hundreds of thousands of square meters with an accuracy of centimeters.

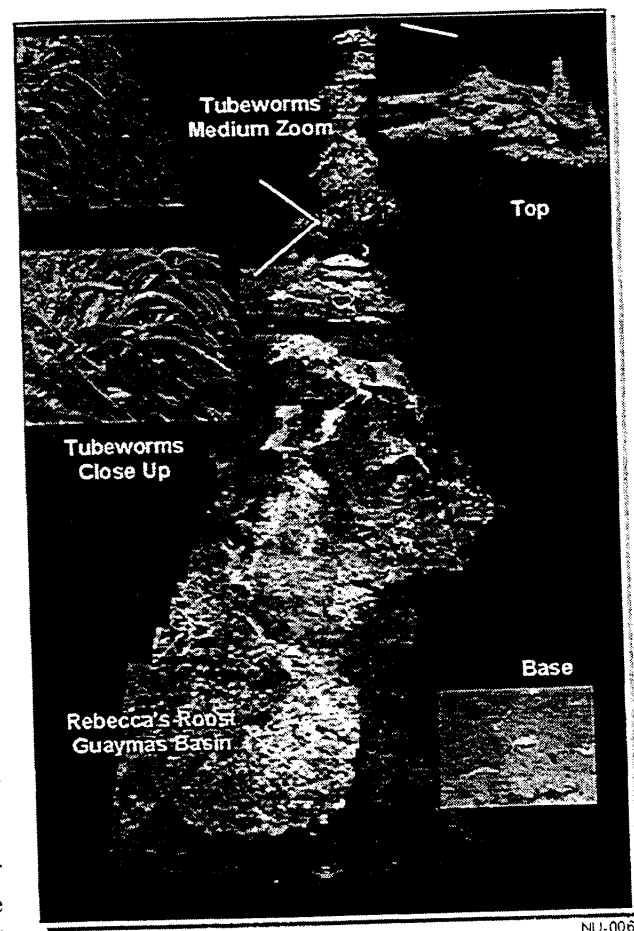
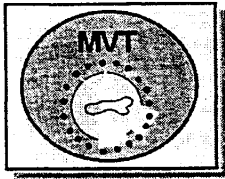


Figure 5. High-resolution underwater optical image of an ocean-floor thermal vent is assembled by 3D photo-mosaicking techniques. (Photo courtesy of Woods Hole Oceanographic Institution.)



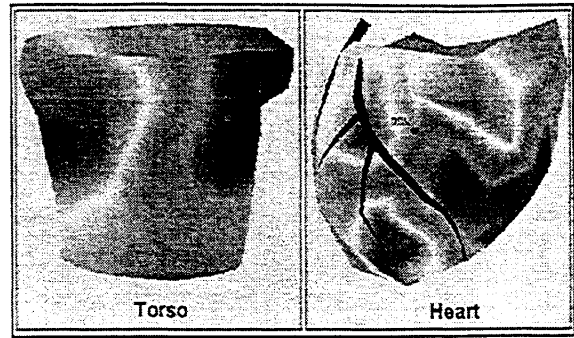
In contrast to LPM where the sensor information is spatially isolated, in *multi-view tomography* (MVT) correlated information from multiple sensors is combined mathematically to create a virtual map of the physical properties of the target. These systems all involve mathematical inversion of integral

equations through linearization, regularization, and integral transforms. Examples include x-ray CAT scanning, diffraction tomography, and synthetic aperture radar.

If multiple view angles are possible, MVT techniques can image obstructed/occluded objects and yield quantitative maps of wave velocity as well as absorption, allowing imaging of phase-only objects. For wavelengths that are short compared to feature dimensions, as in CAT scans^{8,9}, Radon convolution-backprojection algorithms combined with Fast Fourier Transforms (FFT) can achieve 3D feature imaging.

Diffraction tomography¹⁰ is the technique of image reconstruction and resolution enhancement by multiple-view imaging when the wavelength is comparable to feature size. The development of the theory of diffraction tomography by the linearization and Fast Fourier Transform (FFT) inversion of the wave diffraction equations using the filtered back-propagation algorithm was pioneered in the early 1980s.^{11,12} Diffraction tomography has been successfully applied for seismic imaging of near-surface objects, including fossil dinosaur bones¹³ and to ultrasonic imaging^{14,15,16}. Applications of diffraction tomography with limited or obstructed field-of-view or with higher-order, non-linear models is at the forefront of the state-of-the-art.

When the wavelength is long compared to feature size, *near-field* tomographic techniques can still yield useful information in geophysical or medical applications. In Electrical Resistance Tomography and Electromagnetic Induction Tomography, quasi-static probes and models are used to image contaminants in soil and groundwater and leaks from storage tanks on scales much smaller than the electromagnetic wavelength.^{17,18} Applications in medical imaging include Electrical Impedance Tomography¹⁹ and Cardiac Electrical Imaging^{20,21,22}. Figure 6 shows the electric potential on the heart imaged from the measured potential on the torso by Cardiac Electrical Imaging. The potential benefits of the enhanced information gained by this technique over standard electrocardiograms (ECGs) are enormous. ECGs have a rate of false diagnosis of myocardial infarctions ("heart attacks") as high as 30% which results in unnecessary health-care costs in the U.S. estimated at \$4 billion per year²³, while up to 25% of actual heart attacks go unnoticed until evidence of cardiac damage is detected in annual checkups.



NU-008

Figure 6. In Cardiac Electrical Imaging (CEI), near-field MVT inversion yields the electric potential on the heart from measured voltages on the torso.



Multi-spectral discrimination (MSD) adds the element of frequency discrimination to the spatial resolution sought by LPM and MVT giving a 4-dimensional map (3-space plus frequency) of the object. Combinations of MSD with LPM are common (a color photograph or hyperspectral image are examples). Joint methods for MSD and MVT have received little attention. MSD information extraction methods focus on material dispersion, parameter estimation, image registration, and fusion. Multi-sensor fusion can be viewed as an MSD problem involving, in some cases, probes that differ in modality (acoustic and optical, for example) as well as frequency.

For example, the work illustrated in Figure 7 shows that subtractive imaging at two nearby optical wavelengths can map specific chemical concentrations, such as oxygenated/deoxygenated hemoglobin (Hb).²⁴ This use of optical spectroscopy to detect chemical indicators of physiological function *in vivo* is promising for diagnostic discrimination. The rich spectral interaction of IR-VIS-UV light with biological molecules, however, causes absorption and strong scattering in tissue^{25,26,27} and makes the localization of emergent light difficult. Diffusive Wave Imaging^{28,29} in strongly scattering media is the focus of much current research, including optical coherence tomography^{30,31,32} and CenSSIS work in *dual-wave* acousto-phonic imaging which seeks to improve spatial resolution from centimeters to millimeters for precise quantitative diagnosis.

Satellite hyperspectral imaging of the Caribbean Basin has been used to determine the health of coral reef ecosystems and measure coastal erosion³³. Reflected light is strongly scattered in the water column, by the ocean surface, and by the atmosphere, distorting the frequency spectral information as well as the position of underwater objects. This problem is similar to medical diffusive imaging except on a length scale that differs by orders of magnitude. The physics of both are modeled by the radiative transfer equation (RTE)³⁴. However, in diffusive medical imaging, the ratio of absorbed to scattered light is assumed to be small leading to the diffusion equation, while in ocean scattering Beer's law is often applied by assuming the ratio is large. In reality, the physical situation in both cases may be intermediate, and there is a need for more rigorous forward models and more robust inversion algorithms. Current spatial resolution from space-based platforms is approximately 1 meter; processing techniques that take advantage of accurate physical models may improve the resolution limit to 10 centimeters.

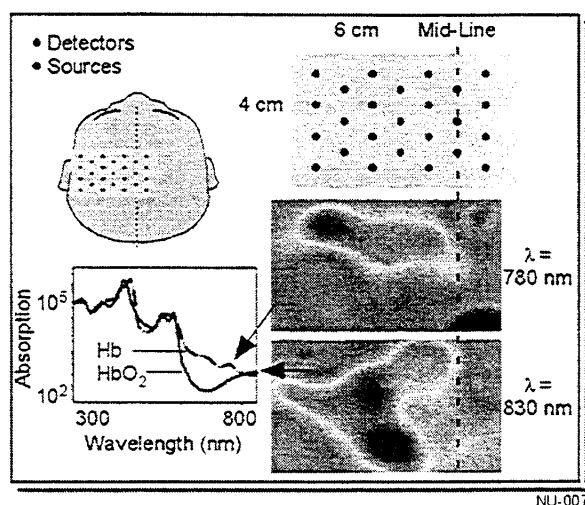


Figure 7 MSD analysis of diffusive optical waves images areas of activity (high blood oxygenation levels) in infant brain.

4. RESEARCH NEEDS AND BARRIERS TO PROGRESS

Progress in subsurface sensing and imaging approaches within these information extraction strategies has been documented in the feature articles in a recent issue of *Science* ("*Imaging: New Eyes on Hidden Worlds*")³⁵. Key elements in these advances include the increase in computation power, the application of new mathematical algorithms and advanced sensing strategies, the exploitation of wave coherence, and the fusion of multiple sensing modalities (e.g., microwave and infrared) to extract increasingly detailed information from physical systems.

Still, the need for new technologies is clear. The General Accounting Office has stated, "the dimensions and potential costs of cleaning up our environment are so great that, without innovative technologies, we may find the solution cost prohibitive and impacting on our ability to address other national needs." By using current technologies, the costs of remediating Superfund and Resource Conservation and Recovery Act sites, Federal facilities, and other known hazardous waste sites may total \$750 billion over the next 30 years³⁶. Humanitarian de-mining remains an unsolved problem. No current imaging technique can adequately detect precancerous cell masses in soft tissue or noninvasively diagnose arteriosclerosis, and there is no way to collect and correlate the images from different modalities to automatically identify incipient health problems.

Barriers to such advanced civil-environmental and biomedical detection systems lie both in unsolved fundamental research problems and in lack of adequate technology tools. Some of the major barriers are:

Barrier 1: Fundamental knowledge is lacking about nonlinear interactions, dual-wave sensing mechanisms, and coherent imaging in scattering media. While linear acoustic and electromagnetic interactions can be modeled and characterized by well-understood linear response functions, advanced imaging techniques using non-linear or dual-wave (e.g., acoustic/optical) probes require fundamental investigations to determine appropriate physical models.

Barrier 2: The present formulation of coherent inverse scattering is inadequate to quantitatively image objects in highly-scattering random inhomogeneous and cluttered environments. In these situations the non-linear character of the inverse problem defeats tomographic reconstruction and adequate alternatives do not yet exist.

Barrier 3: Recognition strategies for obscured and limited-view subsurface applications are not well developed, and we have no theory for combining different sensor inputs to optimize the information obtained.

Barrier 4: Forward modeling of large complex scattering geometries is too slow for real-time inverse-processing applications.

Progress is required in both efficient approximate forward solvers and in hardware/software implementation of processing.

Barrier 5: There are few widely-available test facilities with sufficient flexibility and sensor reconfigurability to permit the optimization of sensor modality/configuration and processing strategies based on recognition and decision objectives.

Barrier 6: Techniques for rapid processing, cataloging, storage and retrieval of large image databases are not sufficiently developed. Data and metadata standards will need be instituted so that processing algorithms can be routinely tested on experimental results from diverse experimental domains.

5. CONCLUSION

The pieces are in place for a major advance in the field of sensing and imaging. The development of a common framework and unified discipline of subsurface sensing and imaging promises to allow the field to emerge as a co-pillar of the Information Age, along with computation and communications. We can look forward to systems-level advances such as integrated, field-tested, algorithmic and computational tools for the entire range of subsurface problems, and standards and criteria for the use of multiple sensing modalities to achieve subsurface sensing system goals. These, in turn, will open the door for the next generation of systems for environmental sensing underground or under the water, medical imaging and automatic diagnosis inside the body, and biological microscopy to reveal fundamental processes inside living cells.

ACKNOWLEDGEMENTS

We would like to acknowledge the invaluable assistance of our many CenSSIS colleagues in developing the concepts presented in this paper. In particular, we would like to thank Bahaa Saleh for the concept of Figure 1, Eric Miller for Figure 3 and the related discussion, Hanu Singh and Jon Howland of Woods Hole Oceanographic Institution for Figure 4, Charles DiMarzio and David Boas for Figure 6, and Dana Brooks for Figure 5. We are especially in the debt of David Castanon for his elucidation of the concept of information extraction strategies. Valuable communications with Ron Roy, Robin Cleveland, and Tom Szabo are gratefully acknowledged.

REFERENCES

1. G. Strada, "Horror of Landmines," *Scientific American*, May 1996, p. 40.
2. D. Colton and R. Kress, "Inverse Acoustic and Electromagnetic Scattering Theory," *Applied Mathematical Science Series*, Vol. 93 (Springer-Verlag) 1992.
3. D. Colton and P. Monk, "The Detection and Monitoring of Leukemia Using Electromagnetic Waves: Mathematical Theory," *Inverse Problems* **10**, 1235-1251(1994).
4. J. A. O'Sullivan, R. E. Blute, D. L. Snyder, "Information Theoretic Image Formation," *IEEE Trans. Information Theory* **44**, 2094-2121(1998).
5. W. Denk, J. H. Strickler and W. W. Webb, "Two-Photon Laser Scanning Fluorescence Microscopy", *Science* **248**, 73-76 (1990).
6. *Handbook of Biological Confocal Microscopy*, J. B. Pawley, ed. (Plenum, New York, 1995), pp. 445-458.
7. W. Denk, "Two-Photon Excitation in Functional Biological Imaging", *Journal of Biomedical Optics* **1**, 296-304 (1996).
8. G. N. Ramachandran and Lakshminarayan (1971), "Three-Dimensional Reconstructions from Radiographs and Electron Micrographs: Applications of Convolution Instead of Fourier Transforms", *Proc. Nat. Acad. Sci.*, **68**, 2636-2240.
9. L. A. Shepp and B. F. Logan (1974), "The Fourier Reconstruction of a Head Section", *IEEE Trans. Nucl. Sci.*, **NS-21**, 21-43.
10. E. Wolf, "Principles and Development of Diffraction Tomography", *Trends in Optics*, Anna Consortini, ed. (Academic Press, San Diego 1996) Chap. 5.
11. A. J. Devaney, "A Filtered Back-Propagation Algorithm for Diffraction Tomography", *Ultrasonic Imag.* **4**, 336 (1982).

12. A. J. Devaney, "A Computer Simulation Study of Diffraction Tomography", IEEE Trans. on Biomed. Eng. BME-30, 377 (1983).
13. A. J. Witten and W. C. King, "Acoustic Imaging of Subsurface Features", J. of Env. Eng. 116, 166 (1990).
14. N. Sponheim, I. Johansen and A. J. Devaney, "Initial Testing of a Clinical Ultrasound Monograph", in Acoustic Imaging, ed. M. Bertero and E. R. Pike (Adam Hilger, Bristol 1991) 401.
15. K. T. Ladas and A. J. Devaney, "Application of ART Algorithm in an Experimental Study of Ultrasonic Diffraction Tomography", Ultrasonic Imaging 15, 48 (1993).
16. J. J. Stamnes, L-J. Gelius, I. Johansen and N. Sponheim, "Diffraction Tomography Applications in Seismics and Medicine", in Inverse Problems in Scattering and Imaging, ed. M. Bertero and E. R. Pike (Adam Hilger, Bristol, 1992) 268.
17. O. Dorn, H. Bertete-Aguirre, J. Berryman, and G. Papanicolaou, "A Nonlinear Inversion Method for 3D-Electromagnetic Imaging Using Adjoint Fields," submitted to Inverse Problems.
18. W. Daily and A. Ramirez, "Environmental Process Tomography in the US," Chem. Eng. Journal 56, 159-165 (1995).
19. M. Cheney, D. Isaacson, and J. C. Newell "Electrical Impedance Tomography," SIAM Review, Vol 41, pp. 85-101(1999).
20. Y. Rudy and B.J. Messinger-Rapport, "The Inverse Solution in Electrocardiography: Solutions in Terms of Epicardial Potentials", CRC Crit Rev Biomed Eng 16, 215-268 (1988).
21. D.H. Brooks and R.S. MacLeod, "Electrical Imaging of the Heart: Electrophysical Underpinnings and Signal Processing Opportunities", IEEE Sig. Proc. Mag. 14, 24-42(1997).
22. R.S. MacLeod and D.H. Brooks, "Recent Progress in Inverse Problems in Electrocardiology", IEEE Eng. in Med. & Biol. Soc. Magazine 17, 73-83(1998).
23. H. P. Selker, "Coronary care unit triage decision aids: how do we know when the work?", Am. J. Med. 87, 491-493(1989).
24. P. J. Dwyer, C. A. DiMarzio and R. R. Anderson (1997), "Imaging of Blood Oxygenation in Skin Using Four-Wavelength Reflectance Spectroscopy", Biomedical Sensing, Imaging, and Tracking Technologies II, 2976, 270-280.
25. S. Wan, R. R. Anderson and J. A. Parrish (1991), "Analytical Modeling for the Optical Properties of the Skin with in Vitro and in Vivo Applications", Photochem. Photobiol., 34, 493-499.
26. R. R. Anderson and J. A. Parrish (1982), "Optical Properties of Human Skin", The Science of Photomedicine, Plenum, New York, 147-194.
27. M.J. C. Van Gemert, Steven L. Jacques, HJ. C. M. Sterenborg and W. M. Star (1989), "Skin Optics", IEEE Transactions on Biomedical Engineering, 36, 12, 1146.
28. A.J. Knuttel, J. Schmitt and J. Knutson (1993), "Spatial Localization of Absorbing Bodies by Interfering Discursive Photon Density Waves", Applied Optics, 32, 381.
29. L. Svaasand, R. Tromberg, T. Haskell and M. Berns (1993), "Tissue Characterization and Imaging Using Photon Density Wave", Optical Engineering, 32, 258.
30. D. Huang, et al (1991), "Optical Coherence Tomography", Science, 254, 1178-1181.
31. M. E. Brazinski, et al (1996), "Optical Coherence Tomography for Optical Biopsy", Circulation, 93, 1206-1213.
32. D. Beneron, et al (1997), "Tissue Optics", Science, 276, 2002-2003.
33. L. O. Jimenez and M. Velez-Reyes, "Subset Selection Analysis for the Reduction of Hyperspectral Imagery," Proc. IEEE International Geosciences and Remote Sensing Symposium, Seattle, WA, 1998.
34. Akira Ishimaru, *Wave Propagation and Scattering in Random Media*, IEEE Press, New York, 1997.
35. T. Appenzeller and C. Norman (1997) "Imaging: New Eyes on Hidden Worlds", Science, 276, 1981-1995.
36. Russel, M., Colglazier, E.W., and English, M.R., 1991, Hazardous waste remediation--The task ahead: Knoxville, Waste Management Research and Education Institute, University of Tennessee..

Statistical fusion of GPR and EMI data

Robert A. Weisenseel^{a*}, W. Clem Karl^a, David A. Castañón^a,
Eric Miller^b, Carey Rappaport^b, Charles A. Dimarzio^b

^aBoston University, Electrical and Computer Engineering Department

^bNortheastern University, Electrical and Computer Engineering Department

ABSTRACT

In this paper, we develop a statistical detection system exploiting sensor fusion for the detection of plastic A/P mines. We design and test the system using data from Monte Carlo electromagnetic induction (EMI) and ground penetrating radar (GPR) simulations. We include the effects of both random soil surface variability and sensor noise. In spite of the presence of a rough surface, we can obtain good results fusing EMI and GPR data using a statistical approach in a simple clutter environment.

More generally, we develop a framework for simulation and testing of sensor configurations and sensor fusion approaches for landmine and unexploded ordnance (UXO) detection systems. Exploiting accurate electromagnetic simulation, we develop a controlled environment for testing sensor fusion concepts, from varied sensor arrangements to detection algorithms. In this environment, we can examine the effect of changing mine structure, soil parameters, and sensor geometry on the sensor fusion problem. We can then generalize these results to produce mine detectors robust to real-world variations.

Keywords: mine, sensor fusion, statistical, radar, induction, model, UXO, detection

1. INTRODUCTION

The need for mine and unexploded ordnance (UXO) detection and removal is growing in both military and humanitarian applications. Civilian casualties from landmines are on the order of hundreds per week, and in some locations around the world, landmines are emplaced faster than they can currently be removed. The current U.N. standard for humanitarian mine detection probability is set at 99.97%. But with any single current buried object detection technology, such a high P_d results in an unacceptably large number of false alarms, particularly when searching for plastic anti-personnel (A/P) mines. However, through information fusion, we can combine the best aspects of multiple sensor technologies to achieve the goals of landmine removal.

To achieve an optimal information fusion system, we must start from fundamental, controlled studies of system design tradeoffs, taking into account every aspect of the problem, from sensor physics and configuration to fusion and detection. Unfortunately, little work has been done to date on such a comprehensive approach to system design. While some researchers have studied individual sensors and detection schemes extensively, sensor fusion researchers have not conducted necessary controlled studies of the effects of design choices on the fusion and ultimate detection problems.

This paper represents a first effort toward filling in this gap in our knowledge. Our goal is to develop a framework for mine detection system research and design that includes the critical aspects of a sensor fusion scenario. We have chosen to focus on detecting plastic A/P mines because their detection represent a particularly challenging problem for which there is currently no satisfactory fielded solution. Information fusion appears to be the solution for this difficult detection problem.

In this work, we concentrate on fusing information from two sensors, a Ground Penetrating Radar (GPR) and an Electro-Magnetic Induction Spectroscopy (EMIS) sensor. We demonstrate that even a simple sensor fusion design can be quite successful over either individual sensor when the physics of their operation provides complementary information. However, we show that individual sensors can be strongly affected by variations in soil parameters and sensor configuration, and this in turn can affect our design choices sensor fusion and detection, which supports our contention that more extensive controlled studies are necessary.

*Correspondence: Email:lorax@bu.edu; WWW: <http://mdsp.bu.edu/lorax.html>

2. BACKGROUND

Our current approach to detecting plastic A/P mines is based on fusing data from two sensors, a Ground Penetrating Radar (GPR) array similar to that operated by Geocenters¹ and an Electro-Magnetic Induction Spectroscopy (EMIS) sensor such as those developed by Geophex.² Each type of sensor has its advantages and drawbacks, but neither alone is sufficient to the problem of reliably detecting plastic A/P mines with sufficiently low false alarm rates. A visualization of the basic GPR sensing concept appears in Figure 1(a) and a photo of one kind of EMIS sensor appears in Figure 1(b).

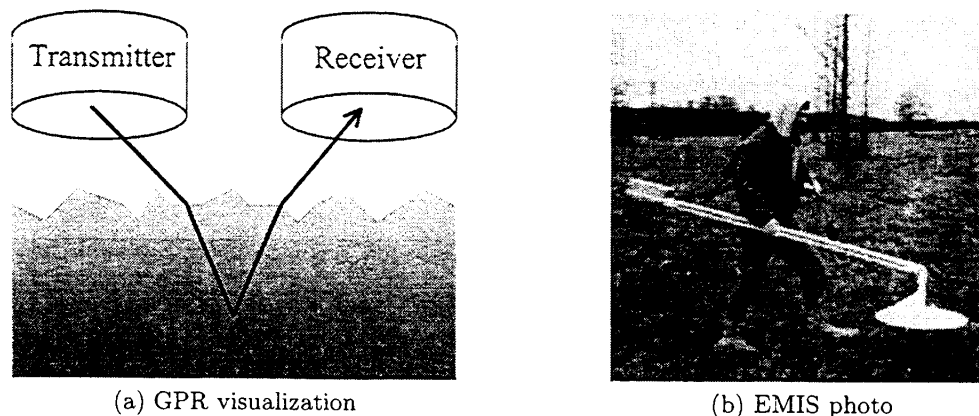


Figure 1. Ground penetrating radar and electromagnetic induction spectroscopy sensors

To date, GPRs have been the favored electromagnetic technology for detecting plastic A/P mines.³ GPRs operate by emitting wideband electromagnetic pulses, which propagate into the earth and reflect from subsurface structures. The reflections are then measured and, in the case of arrays, fused. GPRs are good for detecting shallow dielectric cavities, of which plastic A/P mines are examples. Unfortunately, plastic A/P mines are not the only examples of dielectric cavities, with rocks representing a significant potential confusing class of objects. This potential confusion is especially true since random reflections from a rough air/ground interface can swamp the mine or discrete clutter signatures, making distinguishing between these two classes difficult. Similarly, soil inhomogeneity, such as soil compaction or moisture content, can lead to numerous false alarms. Also, very small metallic objects, such as the firing pin found in many plastic A/P mines, are almost invisible to GPR, which would remove a significant potential characteristic for discerning plastic landmines if GPR were used alone.

An EMIS sensor can complement a GPR by detecting and recognizing the small metal content of many plastic A/P mines. By emitting sinusoidal signals and sweeping across multiple frequencies, an EMIS sensor can obtain information that depends on metal material properties and spatial structure, even for very small amounts of metal.² Also, EMIS sensors are less likely to be affected by soil variations, unlike GPR. However, at gains high enough to see the small metal content of many plastic A/P mines, such an inductive sensor is also highly sensitive to, and frequently confused by, discrete metallic clutter such as small shell fragments, pop-tops from soda cans, or many other kinds of metallic clutter. Additionally, EMIS sensors cannot distinguish landmines with zero metallic content. Thus, an EMIS sensor alone also cannot successfully discriminate plastic A/P mines. But many potential false alarms in the GPR domain are not the same as those in the EMIS domain, and vice versa, i.e. we expect that in most cases, dielectric cavities and metallic objects that coincide in space are likely to be plastic A/P mines and not random clutter.

3. METHOD

As stated in Section 1, our goal is to develop a framework for testing a wide range of possible sensor and fusion scenarios for landmine detection, far more cases than is possible with currently available landmine data sets. By providing an environment in which we can control every aspect of the problem, from rough soil surfaces and soil inhomogeneity to variations in mine pose and sensor geometry to different clutter environments, we can conduct

fundamental tradeoff studies for sensor fusion. With this approach, we will be able to find optimal solutions that account for the wide range of possible situations that can arise in real, practical problems. In this paper, we present an initial study that uses this approach to examine array GPR and EMIS data fusion.

3.1. Sensors

Our detection system consists of two sensors, one a GPR array and the other an electromagnetic induction spectroscopy (EMIS) sensor, and a set of detection and fusion algorithms. Figure 2 shows the basic sensor and landmine geometry.

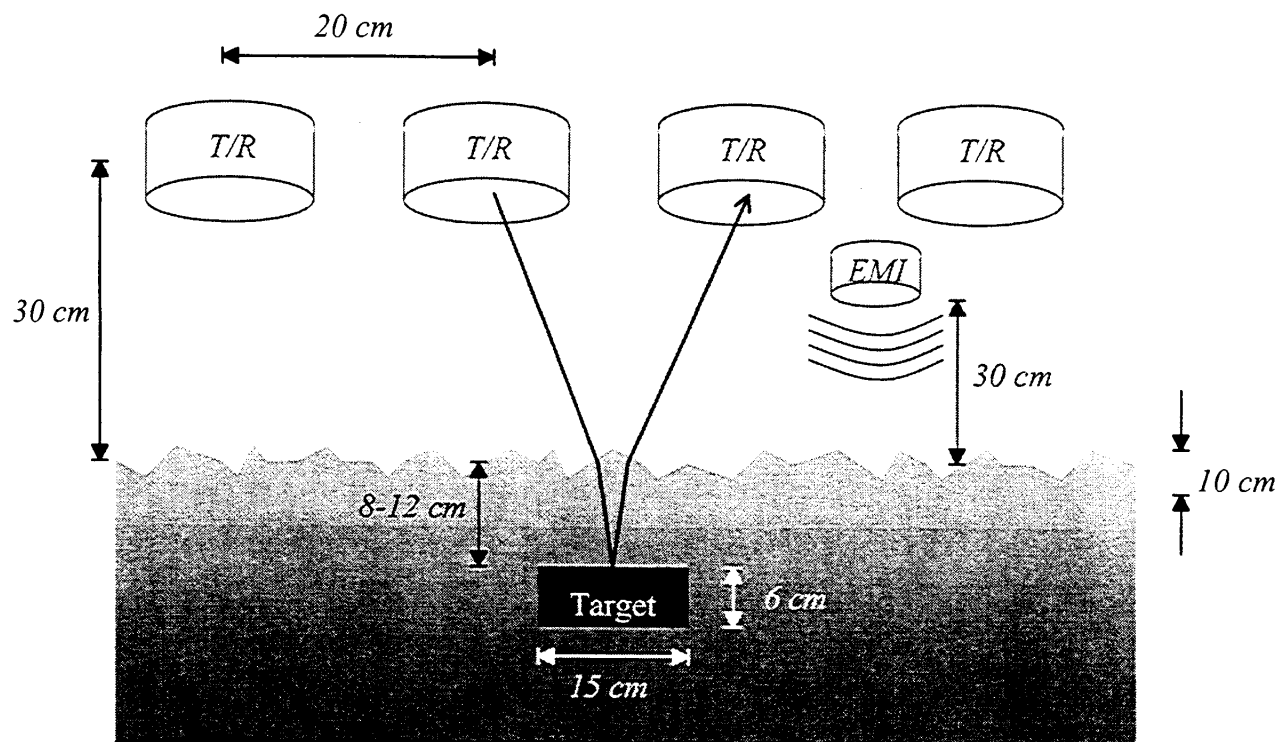


Figure 2. Geometry of the sensor/mine configuration

Our GPR array uses four antennae with transmit/receive capability. We locate these sensors approximately 30cm from the nominal soil surface, corresponding to a vehicular mounting arrangement. The antennae are spaced at 20cm intervals. We choose to use a short, wideband pulse. Our pulse width is about 0.8ns, corresponding to a 1.25GHz band width. The observed signal from the GPR consists of 16 time traces, one for each transmitter/receiver pair. It is important to note here that we are not explicitly focusing the array. We simply transmit a pulse from a single antenna while receiving at all four antennae, and then transmit from the next antenna while again receiving with all four, and so on. This approach maximizes the amount of information available to the detection algorithm. We show a noiseless example set of time traces for one transmit/receive pair in Figure 3. The two overlaid traces correspond to the cases with and without a mine in dispersive soil for a single example of a rough ground surface.

Our EMIS sensor is a single square induction coil located approximately 30cm above the nominal soil surface. The frequency is swept over a 30Hz-20kHz range, which we have sampled logarithmically. Multiple such sensors could be placed in an array, but here we have only considered the single coil case. Our observed data is a vector of samples of the observed EMF signal, logarithmically spaced over the frequency range.

3.2. Physical Models

To generate our data, we use accurate electromagnetic software tools. In the case of the GPR array, we use a Finite Difference Time-Domain (FDTD) algorithm.⁴ With FDTD, we can generate data corresponding to a variety of

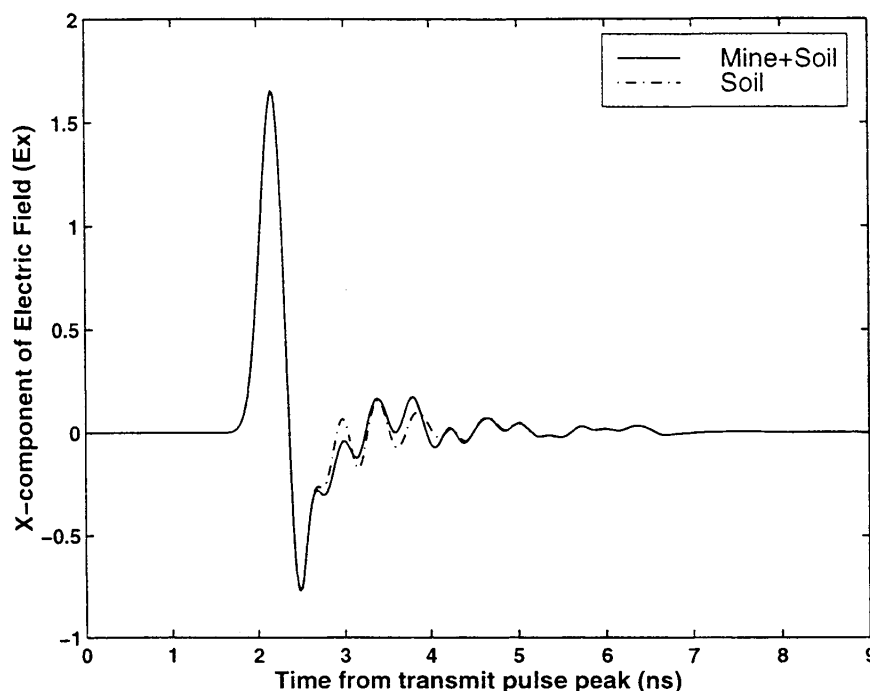


Figure 3. Noiseless GPR time traces; transmitter 2, receiver 3

situations that we can control explicitly: varying rough ground surfaces, different soil parameters, and alternative sensing geometries, for cases with plastic A/P landmines or discrete clutter such as rocks or no subsurface objects at all. For the EMIS sensor, we use a quasi-static magnetic moment physical model, for which we can control sensing geometry and the properties of mine or clutter.⁵ For this study, we choose to assume that our sensors are centered over the section of ground we are testing, and that we are merely trying to test whether or not a landmine is located directly below the sensors at a known depth.

For the GPR, we include a random, but correlated, air/ground interface as a source of uncertainty, as well as white, Gaussian sensor noise of about 20dB per channel signal to noise ratio. In our physical GPR model, we construct three cases. In the first, we include a plastic landmine of known shape, using a relative permittivity of 2.9, corresponding to TNT. In the second, we include a discrete clutter object, with a relative permittivity of 2.8, that is approximately an ellipsoid of roughly the same size as the landmine. In the last case, we modeled no discrete object at all, measuring the response from the empty, homogeneous soil. The soil's relative permittivity is 2.6, and we studied both dispersive, lossy and non-dispersive, lossless cases.

For the EMIS sensor, we model the plastic landmine as having a small, steel firing pin centered in the mine, approximately 1cm in size, with a relative permeability of 5000 and a conductivity of 10^7 S/m. We assume the earth is essentially invisible to the sensor at such low frequencies. We also model a discrete metal fragment as approximately 4cm in size, with the same material properties as the firing pin. For sensor noise, we include white, Gaussian noise for 14dB signal to noise ratio. Note here that, at this noise level, the resulting clutter object looks indistinguishable from the firing pin. In addition to the mine and clutter situations, we include empty soil by simply using the sensor noise with zero mean.

In modeling this system, we only include single objects within the field of view. Thus, the landmine and discrete clutter cases are mutually exclusive. To stress this physical model of clutter versus target, we include the chart of Figure 4, showing how the GPR and EMI clutter characteristics relate. We also never vary the pose of the objects; we assume they lie flat. We will explore more complicated models in future work.

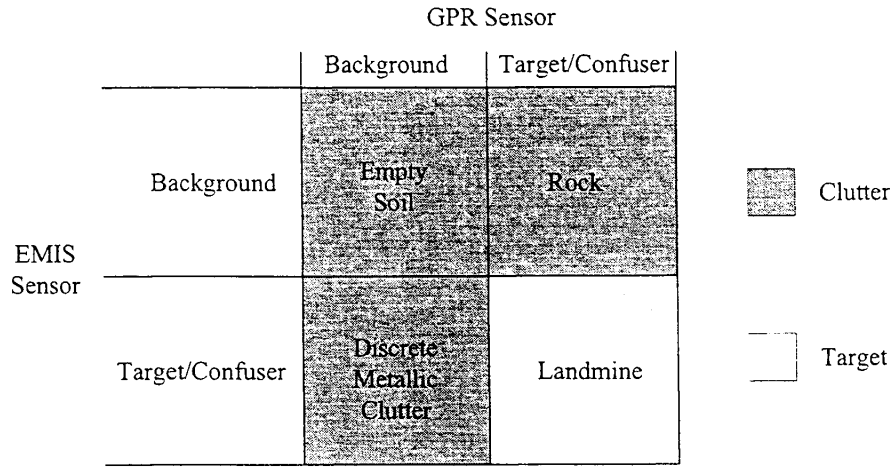


Figure 4. Clutter versus target domains

3.3. Statistical Models

Our statistical models are related to, but quite distinct from, our physical models. Our goals in doing statistical modeling are threefold. First, by allowing statistical error into the problem, we can make our approach robust to modeling error. Thus, the statistics can absorb any differences between our physical model and reality, particularly difficult-to-manage nonlinearities. Second, statistical approaches usually can provide measures of error, so that we can estimate how well we are doing. Third, by using a statistical approach, we open a wealth of available techniques and knowledge for sensor fusion and detection.

For the GPR, our statistical model assumes that the data comes from three additive vector sources: the rough ground return, \underline{g} ; white Gaussian sensor noise, \underline{w} ; and possibly a buried landmine, \underline{s} . We formulate our two detection hypotheses by the presence or absence of the landmine signal, \underline{s} :

$$\begin{aligned} H_{r,0} : \quad \underline{y}_r &= \underline{g} + \underline{w} \\ H_{r,1} : \quad \underline{y}_r &= \underline{s} + \underline{g} + \underline{w} \end{aligned} \quad (1)$$

where \underline{s} , \underline{g} , and \underline{w} are simply the time signals corresponding to the mine signal, the ground return, and the white sensor noise, fused at the data level by vectorization:

$$\underline{s} = [s_{j,k}(n)] \quad \underline{g} = [g_{j,k}(n)] \quad \underline{w} = [w_{j,k}(n)] \quad \begin{cases} j, k = 1, 2, 3, 4 \\ n = 1, \dots, 750 \end{cases} \quad (2)$$

where j and k denote the specific transmit/receive antenna pair, and n is the time sample. The sum of these signals, \underline{y}_r , is what we can observe as output from the antenna array. We assume that these signals are independent Gaussian random vectors, distributed as follows:

$$\underline{s} \sim N(\underline{\mu}_s, \Sigma_s) \quad \underline{g} \sim N(\underline{\mu}_g, \Sigma_g) \quad \underline{w} \sim N(\underline{0}, \Sigma_w) \quad (3)$$

where $\underline{\mu}_s$ and $\underline{\mu}_g$, and Σ_s and Σ_g , are the means and covariances, respectively, of their distributions, and the elements of \underline{w} are zero-mean and identically distributed, so $\Sigma_w = \sigma_w^2 I$. For computational tractability, we currently assume Σ_s and Σ_g are diagonal, though in future work we may introduce tractable correlation structure.

For EMIS, we have a single sensor, with data from a sweep over frequency, and this data is corrupted by Gaussian noise. Going immediately to vector notation, we assume that the signal \underline{c} is a known deterministic signal and that \underline{u} is independent and identically distributed, zero-mean, Gaussian noise, so our two cases are simply:

$$\begin{aligned} H_{e,0} : \quad \underline{y}_e &= \underline{u} \\ H_{e,1} : \quad \underline{y}_e &= \underline{c} + \underline{u} \end{aligned} \quad (4)$$

where

$$\underline{u} \sim N(0, \Sigma_u) \quad \Sigma_u = \sigma_u^2 I \tag{5}$$

It is important to note that in this preliminary study, we elected not to explicitly model the discrete clutter in our statistical detection problem, but we do include such clutter in the physical model for generating the data. Nowhere in any of the statistical models do we include a discrete clutter case, i.e. a rock or metallic clutter. We did this because we wanted to explore the problem of model mismatch and the advantages of detection-level fusion when discrete clutter models are unknown or ignored.

3.4. Detectors

Given the statistical models above, it is relatively easy to formulate an “optimal” maximum likelihood detector, based on a likelihood ratio test. In the GPR case, we assume the three Gaussian signals are independent, so we can simply add their means and covariances

$$\begin{aligned} H_{r,0} : \quad & \underline{y}_r \sim N(\underline{\mu}_g, \Sigma_g + \Sigma_w) = N(\underline{\mu}_0, \Sigma_0) \\ H_{r,1} : \quad & \underline{y}_r \sim N(\underline{\mu}_s + \underline{\mu}_g, \Sigma_s + \Sigma_g + \Sigma_w) = N(\underline{\mu}_1, \Sigma_1) \end{aligned} \tag{6}$$

The detector resulting from this model and a likelihood ratio test is the full quadratic classifier:

$$\left(\underline{y}_r - \underline{\mu}_0\right)^T \Sigma_0^{-1} \left(\underline{y}_r - \underline{\mu}_0\right) - \left(\underline{y}_r - \underline{\mu}_1\right)^T \Sigma_1^{-1} \left(\underline{y}_r - \underline{\mu}_1\right) \underset{H_0}{\overset{H_1}{\geq}} \gamma_r \tag{7}$$

We estimate the parameters from a large number of Monte Carlo runs by estimating $\underline{\mu}_g$ and Σ_g for the case with empty soil, and $\underline{\mu}_s + \underline{\mu}_g$ and $\Sigma_s + \Sigma_g$ for the landmine case. These parameters are averaged over many different, random rough soil/air interfaces. Because we control the sensor noise, we can add the known white noise covariance to each case ourselves.

Similarly, we develop the maximum likelihood detector for the EMIS case. However, because the signal is a known deterministic signal and we control the noise, we can simplify both the equations and the detector:

$$\begin{aligned} H_{e,0} : \quad & \underline{y}_e \sim N(0, \Sigma_u) \\ H_{e,1} : \quad & \underline{y}_e \sim N(\underline{c}, \Sigma_u) \end{aligned} \tag{8}$$

$$\underline{c}^T \Sigma_u^{-1} \underline{y}_e \underset{H_0}{\overset{H_1}{\geq}} \gamma_e \tag{9}$$

While this approach is adequate for our current simulations, in practice, our detector and noise parameters will need to be estimated from data.

From these models, we perform three detection experiments to demonstrate the advantages of fusion. First we test the GPR alone, then we test the EMIS sensor alone. Lastly, we fuse the two sensing modalities with detection-level fusion, using an AND detector. To build an appropriate optimal detector, we need to explicitly model the confusing clutter classes statistically. The advantage of the AND detector is that it can approximate the optimal multi-class detector without explicit clutter models. While it is always better to obtain the optimal detector by explicitly modeling clutter, in cases where this is infeasible, an AND detection rule will win. In future work, we intend to formulate an optimal detector by determining and using appropriate multi-class clutter models.

4. RESULTS

We conducted three major experiments, over which we varied soil parameters and mine depth. In each of our experiments, we ran 400 Monte Carlo runs with different rough air/soil interface and sensor noise realizations. In 100 cases, we placed a landmine. In another 100 cases we placed the GPR clutter object (the “rock”). In the next 100 cases, we used the EMIS clutter object (the small metal fragment). In the last 100 cases, we included no discrete clutter or mine at all, only empty soil. In each of the three experiments, we used each of the three detection methods: GPR alone, EMIS alone, and their detection-level fusion.

For each of the 400 Monte Carlo runs, we generated the random soil surface using a correlated Gaussian random model with approximately $\pm 5\text{cm}$ of deviation from nominal, maximum, and an 8cm correlation length. In every case, we set the soil's relative permittivity to 2.63, but in two cases we used a lossless, non-dispersive soil model, while in the third, we implemented a lossy, dispersive model based on data from the Bosnia Steele Castle site. In this case, the moisture content was 4.7% and the density was 1.181 g/cm^3 , and we modeled the loss and dispersion using a 2-1 Padé approximation.

In one of the two lossless, non-dispersive cases, we placed a landmine with its upper surface at a depth of 8cm below the nominal soil surface, while in the other we placed it 12 cm below the nominal soil surface. In the lossy, dispersive case we placed the landmine 8cm below the surface. Our landmine was 6cm in thickness, with a diameter of 15cm. Except for the firing pin, we assumed the mine was filled with TNT, with a relative permittivity of 2.9, relative permeability 1, and conductivity 0. We also assumed that the mine contained a small, 1cm firing pin with a relative permeability of 5000 and a conductivity of 10^7S/m .

For our clutter, we assumed that the rock was roughly the same size and shape as the mine, but approximately ellipsoidal with major axis 17cm and minor axis 7cm, oriented similarly to the landmine. We assumed the rock had a relative permittivity of 2.8, with relative permeability of 1 and no conductivity. We assumed that the metal fragment clutter about 4cm in size with the same material properties as the landmine firing pin.

Figure 5 shows the ROC curves for the case of a mine at a depth of 8cm in lossless, non-dispersive soil. First, note that the ROC curve for the EMIS sensor alone shows good correspondence to what we expect for the case that the metal clutter object is nearly indistinguishable from the landmine. The straight line ROC is indicative that guessing is as good as any other technique when trying to distinguish the metal clutter from the firing pin when using the EMIS sensor. The GPR is better at distinguishing between the rock and the landmine than the EMIS sensor is for the metal clutter, but it still has a significant number of false alarms for $P_d \approx 0.99$. However, the detection-level fusion scheme has a much lower false alarm rate for comparable P_d .

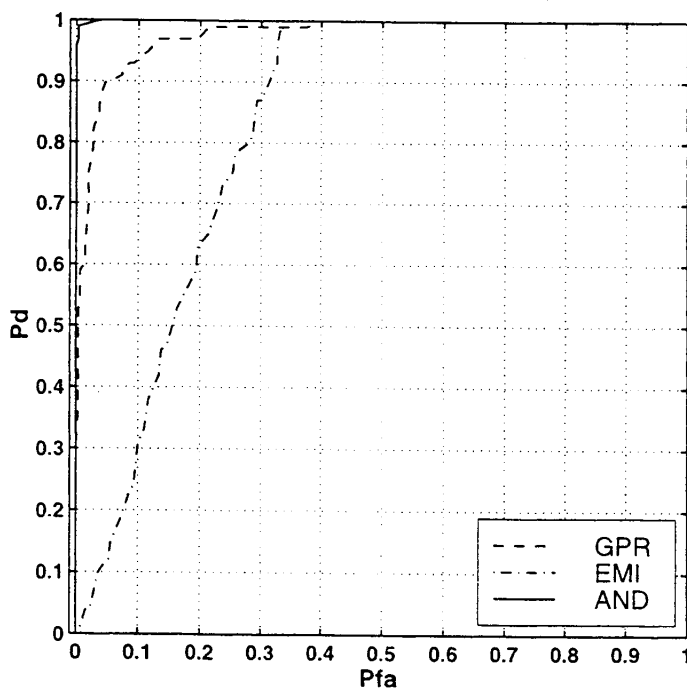


Figure 5. ROC curves for mine at 8cm depth, non-dispersive soil

In Figure 6, we see the ROC curves for the mine at a depth of 12cm, with the same soil parameters. As we expect, the increased distance makes the EMIS sensor perform slightly worse than before. The GPR alone, however,

gets significantly improved performance. This improvement comes from the time delay of the mine signal. While the mine signal is attenuated due to the increased distance, the distance attenuation of the reflection from the rough ground surface is much stronger, resulting in a much smaller interfering signal relative to the mine signal at this time delay. This improvement would be mitigated somewhat by lossy soil. Also, as expected, the fused detector performs better than either sensor alone, despite its simplicity.

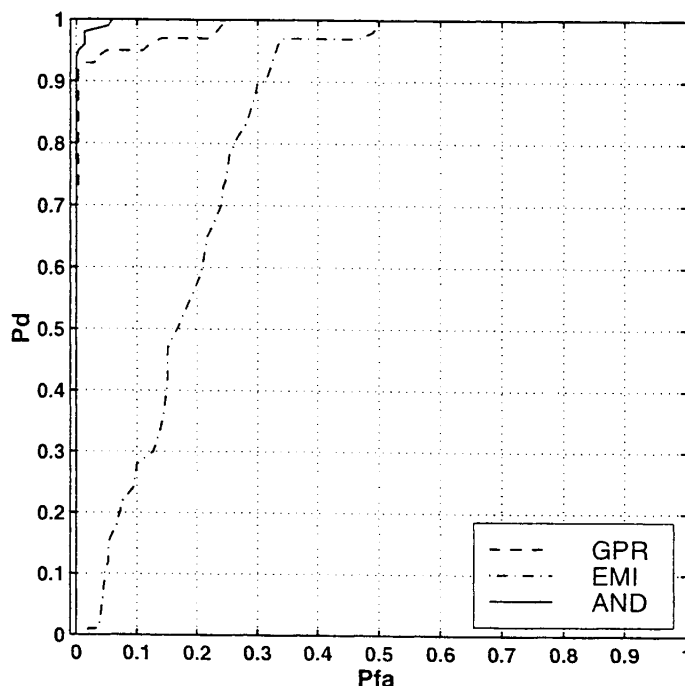


Figure 6. ROC curves for mine at 12cm depth, non-dispersive soil

Lastly, in Figure 7, we show the ROCs for the case of a mine buried at a depth of 8cm in lossy, dispersive soil. The EMIS ROC is essentially the same as the first case, but the GPR curve is much different. Here we begin to see problems with model mismatch. Because we have not explicitly modeled clutter statistically, we have found a situation where the GPR clutter signal is much stronger than the mine signal. Thus, despite the differences in the two signals, the clutter signal is detected as a mine more often than the mine is. This poor GPR detector performance suggests that we can do much better if we model the clutter explicitly in our statistics. The AND detection rule again does quite well; in fact it appears to do perfectly. This is an artifact of ROC estimation from discrete, experimental data and the small number of Monte Carlo runs. More Monte Carlo samples, or another approach, such as importance sampling, would help rectify this issue. With more samples, our ROC estimates and confidence in them would improve.

5. CONCLUSIONS

We have developed a framework for studying the system design tradeoffs in sensor fusion for landmine detection applications. We can now conduct controlled experiments that examine the effects that soil or landmine parameters, sensor geometries, clutter, and fusion algorithms have on the landmine detection problem. With this approach, we aim to identify the optimal sensor configurations and fusion algorithms for general, real-world landmine problems.

We have used this framework to study the effect of variations in soil parameters and mine depth on GPR and EMIS data fusion, particularly for the case of plastic A/P mines. We also focused on the issue of model mismatch and the advantages of accurate clutter modeling for detection. We found that even simple detection-level sensor fusion techniques can provide a large improvement in our ability to detect plastic A/P mines.

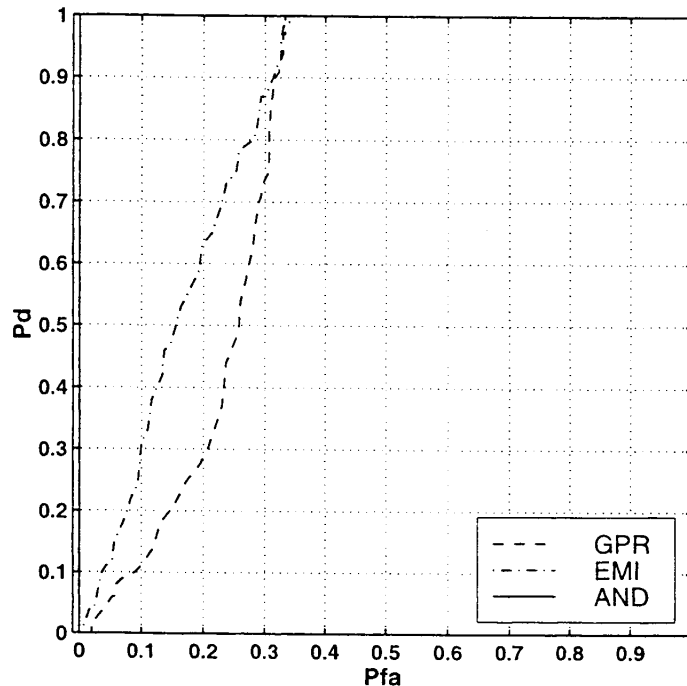


Figure 7. ROC curves for mine at 8cm depth, dispersive soil

6. ACKNOWLEDGMENTS

This work is partially supported OSD MURI program, the Air Force Office of Scientific Research under Grant F49620-96-1-0028, and the National Institutes of Health under Grant NINDS 1 R01 NS34189. We would also like to thank Scott Winton and Mustafa Özdemir of Northeastern University for their assistance with the electromagnetic simulations.

REFERENCES

1. C.M. Rappaport, D.M. Reidy, "Focused array radar for real-time imaging and detection", *Radar Sensor Technology*, Gerald S. Ustach, editor, Proc. SPIE V 2747, p202-213, SPIE, Orlando (1996)
2. I.J. Won, D. Keiswetter, "Electromagnetic Induction Spectroscopy", *Detection and Remediation Technologies for Mines and Mine-Like Targets III*, A.C. Dubey, J.F. Harvey, J. Broach, editors, Proc. SPIE V 3392, p14-22, SPIE, Orlando, (1998)
3. T.R. Witten, "Present state-of-the-art in ground penetrating radars for mine detection", *Detection and Remediation Technologies for Mines and Mine-Like Targets III*, A.C. Dubey, J.F. Harvey, J. Broach, editors, Proc. SPIE V 3392, p576-585, SPIE, Orlando, (1998)
4. G.A. Tsibrantzis, C.M. Rappaport, S.C. Winton, P.M. Johansen, "Statistical modeling of rough surface scattering for ground-penetrating radar applications", *Detection and Remediation Technologies for Mines and Mine-Like Targets III*, A.C. Dubey, J.F. Harvey, J. Broach, editors, Proc. SPIE V 3392, p735-744, SPIE, Orlando, (1998)
5. M. Özdemir, E. Miller, S. Norton, "Localization and characterization of buried objects from multi-frequency, array inductive data", *Detection and Remediation Technologies for Mines and Mine-Like Targets IV*, A.C. Dubey, J.F. Harvey, J. Broach, R.E. Dugan, editors, Proc. SPIE V 3710, SPIE, Orlando, (1999)

FDTD Wave Propagation in Dispersive Soil Using a Single Pole Conductivity Model

Carey M. Rappaport, Shuang Wu and Scott C. Winton
Center for Electromagnetics Research, Northeastern University, Boston, MA

Abstract—In FDTD modeling of lossy, dispersive soil for subsurface imaging and detection applications, the electric flux and the current are convolutions of $E(t)$ with $\epsilon(t)$ and $\sigma(t)$ respectively. To avoid these memory-intensive computations, the convolutions can often be accurately and simply modelled as second order difference equations. In particular, by matching the corresponding Z -transform of the E-field/current relation to frequency-dependent conductivity results in a ratio of polynomials in Z^{-1} (where $Z = e^{j\omega\Delta t}$). A good fit to measured soil data over two decades in frequency is possible using only a single pole, two zero conductivity model. Compared to a similarly accurate three-term Debye model, this one-pole model requires one-third the storage of previously computed field values.

Index terms—FDTD, Soil Modeling, Dispersion, Mine Detection

I. BACKGROUND

There is growing interest in simulating wave propagation for underground microwave and RF sensing and imaging applications. Soil is a good candidate for the finite difference methods since it is usually inhomogeneous, lossy, and has an irregular surface boundary. It has been challenging to accurately compute wideband wave behavior in realistically modelled soil because of its dispersive nature, requiring either many individual frequency domain calculations or a robust deconvolution of $E(t)$ from $D(t)$ in the time domain. A simple dispersive variant of the FDTD algorithm which includes the effects of frequency-dependent conductivity and dielectric constant enables the use of this prevalent and efficient modeling method.

The conventional approach to modeling dispersion in soil approximates the frequency domain dispersive complex dielectric constant with rational functions (Debye or Lorentz models) of $j\omega$ [1,2], multiplies the constitutive relation by the denominator and inverse Fourier transforms the result. We improve on this method by modeling solely the conductivity as a simple rational function of the Z -transform [3,4], based on the observation that the frequency variation of the real dielectric constant does not significantly affect either the real propagation constant β or the decay rate α . By modeling σ in terms of powers of the Z -transform variable Z^{-1} (which readily transforms to time delays), the conversion of the generalized dispersive Ohm's Law $\mathbf{J}(Z) = \sigma(Z)\mathbf{E}(Z)$ to the time domain is particularly straightforward.

Manuscript received June 1, 1998.

This work has been supported by The Army Research Office, Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013

II. MODEL FORMULATION

To preserve the conceptual elegance and simplicity of the FDTD method, it is important to keep the media model to at worst second-order. In previous reports, a quite satisfactory two-pole, two-zero rational function conductivity model was presented. This current work shows that a model with a single pole can often be just as good. The right hand side of Ampere's Law $\nabla \times \mathbf{H} = \partial \mathbf{D} / \partial t + \mathbf{J}$ Z -transforms to: $[(1-Z^{-1})/\Delta t]\epsilon_0\epsilon_{Av}\mathbf{E} + \sigma(Z)\mathbf{E}$ (where it is assumed that ϵ_{Av} is frequency independent). The time-dependent conductivity may thus be modelled as

$$\sigma(Z) = \frac{J(Z)}{E(Z)} = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2}}{1 + a_1 Z^{-1}} \quad (1)$$

Both the real and the imaginary components of $\sigma(Z)$ depend on the frequency sampling interval Δt and the coefficients of the rational function (a_1 , b_0 , b_1 , and b_2). The imaginary component corresponds to an effective permittivity. This model is implemented by fitting $\text{Re}\{\sigma\}$ to measured conductivity and $\text{Im}\{\sigma\}/\omega\epsilon_0$ to measured real dielectric constant less an average value ϵ_{Av} . To simultaneously solve for b_0 , b_1 , and b_2 , an initial guess is made for a_1 . The conductivity and dielectric constant at three representative frequencies for the measured data and the model are equated. Further simple optimization is performed by trial and error, varying either a_1 or one of the three frequencies. Complex wave number $k(\omega)$ is then calculated from both the model and the measured data, and the real propagation constant β and the decay rate α are derived from $k = \omega/c\sqrt{\epsilon_{Av} - j\sigma/\omega\epsilon_0}$.

Once the a and b parameters of σ have been determined, integration into the FDTD algorithm is straightforward. If the temporal average of the current density \mathbf{J} is used, Ampere's law in the Z domain becomes:

$$\nabla \times \mathbf{H}(Z) = \epsilon_{Av} \frac{1 - Z^{-1}}{\Delta t} \mathbf{E}(Z) + \frac{1 + Z^{-1}}{2} \sigma(Z) \mathbf{E}(Z) \quad (2)$$

Using (1) in (2) and noting that Z^{-1} corresponds to the time index shift $n \rightarrow n-1$, the E-field update equation for a 2-D Transverse Magnetic ($\mathbf{E} = \hat{z}E_z$) simulation is:

$$\begin{aligned} E_z^{n+1} &= \frac{1}{e_0} (-e_1 E_z^n - e_2 E_z^{n-1} - e_3 E_z^{n-2} + \Delta H) \\ \Delta H &= \frac{\Delta t}{\epsilon_{Av}} \left(-I_x^{n+\frac{1}{2}} - a_1 I_x^{n-\frac{1}{2}} + I_y^{n+\frac{1}{2}} + a_1 I_y^{n-\frac{1}{2}} \right) \end{aligned} \quad (3)$$

where $e_0 = 1 + b_0 \Delta t / 2\epsilon_{Av}$, $e_1 = a_1 - 1 + (b_0 + b_1) \Delta t / 2\epsilon_{Av}$, $e_2 = -a_1 + (b_1 + b_2) \Delta t / 2\epsilon_{Av}$, $e_3 = b_2 \Delta t / 2\epsilon_{Av}$ and I_x and I_y are the spatial first differences of the H_x and H_y fields with respect to y and x (spatial indices have been suppressed for clarity).

The one-pole model of $\sigma(Z)$ requires storing I_x and I_y (and I_z in a three-dimensional simulation) a single additional time compared with non-dispersive FDTD. For each additional conductivity pole, this same amount of storage space must be additionally allocated. Fig. 1 shows the fit of the model (at $\Delta t = 20$ ps) to the conductivity and dielectric constant data measured from Puerto Rican clay loam (PRCL) with 1.4 g/cm^3 density and 10% moisture [5]. The difference between the imaginary conductivity of the model and the measured ϵ (lower panel) is used to determine the constant real soil permittivity ϵ_{Av} . Fig. 2 shows the fit of the full model to β and α . The maximum amount of error is less than 20%, which is close to the measurement error of the soil sample.

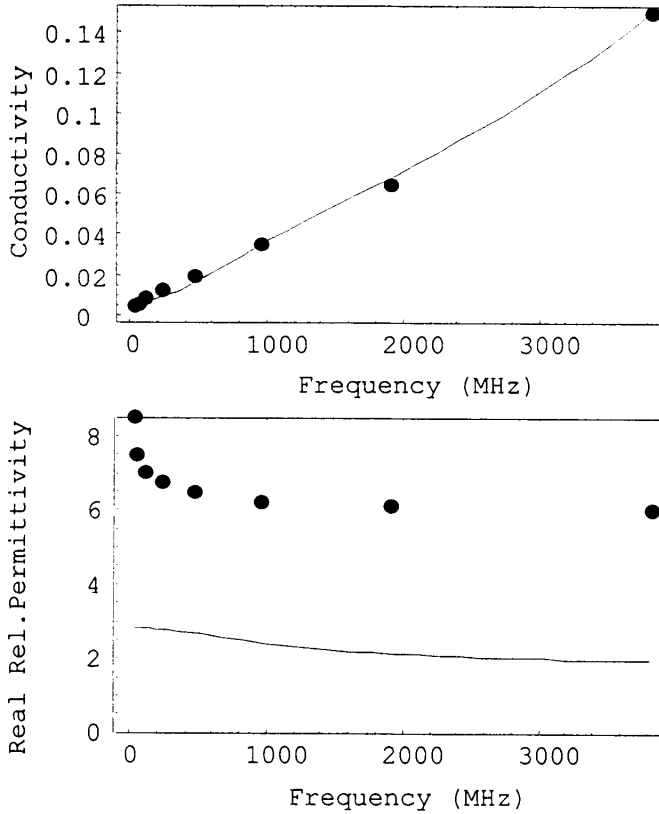


Fig. 1. (a) Measured (circles) versus single pole model (continuous curves) of $\sigma(Z)$: (a) real conductivity and (b) imaginary conductivity (real relative dielectric constant) for Puerto Rican clay loam, 10% moisture, 1.4 g/cm^3 density [5]. The constant real dielectric constant ϵ_{Av} is the average difference between the curve and the measured circles of (b).

The use of the average value of the current density is well established and produces accurate results. However, using only the present value of the current density in the update equation results in the savings of another storage location. To facilitate this savings without sacrificing accuracy, $\sigma(Z)$ is multiplied by $(1 + Z^{-1})/2$ and a new set of model parameters is determined to fit this

product over the frequencies of interest. In this case, (3) is additionally simplified with $e_0 = 1 + b_0 \Delta t / \epsilon_{Av}$, $e_1 = a_1 - 1 + b_1 \Delta t / \epsilon_{Av}$, $e_2 = -a_1 + b_1 \Delta t / \epsilon_{Av}$, and $e_3 = 0$, and this dispersive model requires only one additional storing of each field value compared to non-dispersive FDTD. This compares with two field storings for the (2,2) model and three field storings for the comparably accurate three-term Debye model.

Similar modeling for several different types of soil with widely varying density and moisture [5,6] at $\Delta t = 20$ ps is summarized in Table 1. All models have wave numbers in agreement with measured values within 20%.

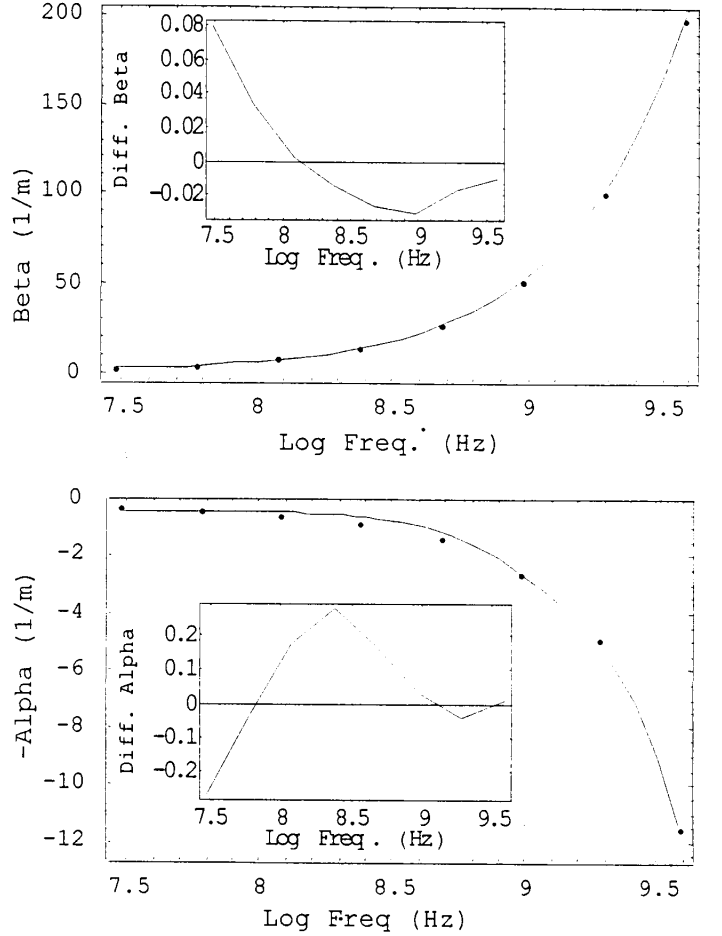


Fig. 2. Measured (circles) versus model for single pole model of $\sigma(Z)$ with constant ϵ_{Av} (continuous curves): (a) Real and (b) imaginary parts of wave number for Puerto Rican clay loam, 10% moisture, 1.4 g/cm^3 density [5] as a function of frequency. Insets show normalized error for this model.

III. STABILITY CONSIDERATIONS

The Courant stability analysis in [3] shows that in a source-free homogeneous medium, the time-space transform of the wave equation for a plane wave ($\Delta x = \Delta y = \Delta z = \Delta$) is:

$$(Z - 1)F(Z) = -4r^2 s^2 \quad (4)$$

where from (2), $F(Z) = \epsilon_{Av}(1 - Z^{-1}) + (\Delta t/2\epsilon_0)(1 + Z^{-1})\sigma(Z)$, and $r = c\Delta t/\sqrt{\epsilon_{Av}}\Delta$, and $s^2 = \sin^2 k_x\Delta + \sin^2 k_y\Delta + \sin^2 k_z\Delta$ with maximum value of s equal to the spatial dimension of the FDTD simulation. Expressing $F(Z)$ in rational form:

$$F(Z) = \epsilon_{Av} \frac{(e_0 + e_1 Z^{-1} + e_2 Z^{-2})}{1 + a_1 Z^{-1}} \quad (5)$$

Using (4), (3) becomes a third order polynomial in Z^{-1} . The solution to this equation yields one real root and two complex roots that are a function of Δ . Stability requires $|Z| < 1$. For Δ within the range of 4.6mm to 120mm, all three roots are within the unit circle.

NUMERICAL TEST CASES

Several numerical experiments were performed to validate this method. The simulations were run in one dimension for simplicity. The choice of Δt is critical since it defines the model, determines the range of usable frequencies and determines the limits on Δ through the Courant condition. For these experiments, the spatial increment was 2.5 mm. The incident field is a modulated gaussian pulse, used to excite many frequencies simultaneously. The modulation frequency is 3 GHz for the Alicia and A.P. Hill soils and 1.5 GHz for the PRCL soil sample.

In order to compare the results of the FDTD simulations with the measured data, the frequency-dependent wave number k must be extracted. Fast Fourier Transforming the time domain fields gives

$$k_{\text{FDTD}} = -\frac{1}{j\ell} \ln \frac{E_z(\omega, \ell)}{E_z(\omega, 0)} \quad (6)$$

Fig. 3 shows the real and imaginary parts of the wave number versus frequency for measured values and either the “Time-Averaged” or “Model-Averaged” k_{FDTD} . These latter “Model-Averaged” simulations use only the current value of the conduction current J^n . The parameters in Fig. 3 correspond to the three separate sites of Table 1. Both the real and imaginary wave numbers agree well for each soil, for both types of models across the entire two decades of bandwidth from 45 MHz to almost 4 GHz. It should be noted that the “Time-Averaged” FDTD computation is more robust, with less sensitivity to excitation function and Courant number.

CONCLUSIONS

A minimal-storage time-domain model for frequency-dispersive soil based on a constant real dielectric constant and a (2,1) rational function of Z^{-1} conductivity function has been developed and tested. By pre-multiplying the measured conductivity data by the Z-transform of the time average function, the resulting model avoids requiring conduction current time averaging in Ampere’s law. Without this time averaging,

TABLE I
ONE-POLE CONDUCTIVITY PARAMETERS

Puerto Rican Clay Loam: $m = 10$, $d = 1.4^\dagger$					
av?	a_1	b_0	b_1	b_2	ϵ_{Av}
no	-0.88	0.91625	-1.67662	0.761072	4.18775
yes	-0.9	0.7983	-1.4695	0.67176	4.282
A.P. Hill, Firing Point 22: $m = 19.3$, $d = 1.527$					
no	-0.975	1.51947	-2.97284	1.45362	5.25834
yes	-0.97868	1.5473	-3.036	1.4888	4.7731
Bosnia, Test site Alicia: $m = 25.3$, $d = 1.263$					
no	-0.925	1.76106	-3.32102	1.56193	5.03815
yes	-0.93	1.6325	-3.0827	1.4521	4.9831

$^\dagger m$ is percent moisture content of the soil, d is soil density (g/cm³).

a one-pole conductivity model requires just one additional store of electric and magnetic field compared to the non-dispersive FDTD algorithm. This is half the required additional storage of the previously reported (2,2) model, and even less than standard Debye and Lorentz models. In addition, fitting parameters to a single, real conductivity function — rather than to both parts of a complex permittivity function — is quite advantageous.

Numerical simulations on measured data show that this simple model is efficient and accurate across a wide frequency band for both real and imaginary parts of wave number, giving good predictions of velocity and decay rate. A stability analysis shows that the model is stable for one, two, and three dimensions, for a wide range of grid spacings.

REFERENCES

- [1] T. Kashiwa and I. Fukai, “A Treatment by the FD-TD Method for the Dispersive Characteristics Associated with Electronic Polarization”, *Microwave and Guided Wave Letters*, vol. 16, pp. 203–205, June 1990.
- [2] O. Gandhi, “A Frequency-Dependent Finite Difference Time Domain Formulation for General Dispersive Media”, *IEEE Transactions on Microwave Theory and Techniques*, vol. 41, pp. 658–665, April 1993.
- [3] W. Weedon, and C. Rappaport, “A General Method for FDTD Modeling of Wave Propagation in Arbitrary Frequency-Dispersive Media”, *IEEE Transactions on Antennas and Propagation*, pp. 401–410, March 1997.
- [4] C. Rappaport and S. Winton, “Modeling Dispersive Soil for FDTD Computation By Fitting Conductivity Parameters”, *12th Annual Review of Progress in Applied Computational Electromagnetics Symposium Digest*, pp. 112–118, March 1997.
- [5] J. Hipp, “Soil Electromagnetic Parameters as Functions of Frequency, Soil Density, and Soil Moisture”, *Proceedings of the IEEE*, vol. 62, pp. 98–103, January 1974.
- [6] J. Curtis, “Dielectric Properties of Soils: Various Sites in Bosnia”, *U.S. Army Corps of Engineers, Waterways Experimental Station Data Report*, 20 August 1996.

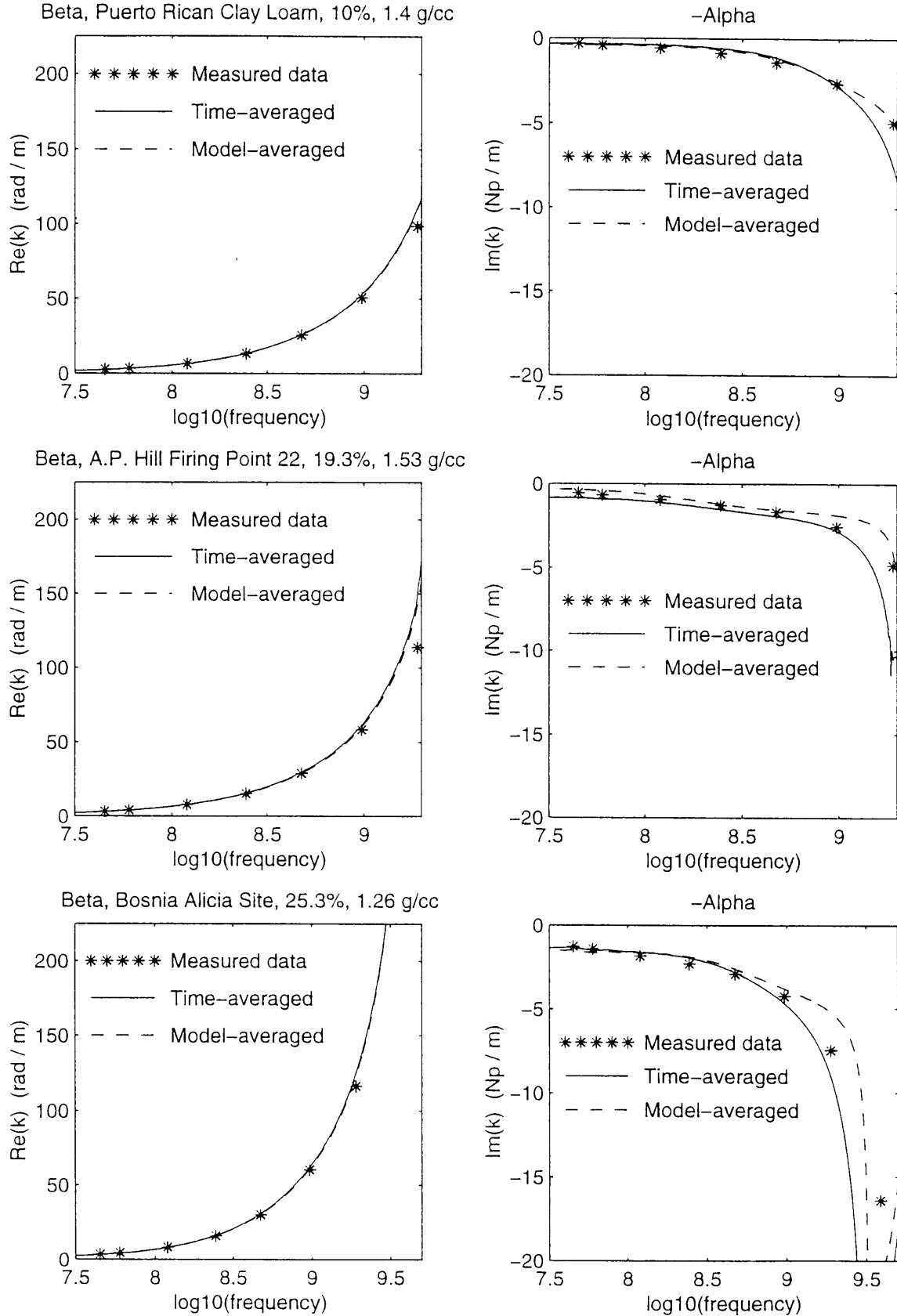


Fig. 3. Measured (*) versus FDTD computations of real and imaginary wave number using "Time-averaged" (solid) and "Model-Averaged" (dashed) models for single pole model of $\sigma(Z)$ with constant ϵ_{Av} for three soils with varying moisture and density.

Statistical sensor fusion analysis of near-IR polarimetric and thermal imagery for the detection of mine-like targets

Robert A. Weisenseel^{a *}

W. Clem Karl^a

David A. Castañón^a

Charles A. Dimarzio^b

^aBoston University, Electrical and Computer Engineering Department

^bNortheastern University, Electrical and Computer Engineering Department

ABSTRACT

We present an analysis of statistical model based data-level fusion for near-IR polarimetric and thermal data, particularly for the detection of mines and mine-like targets. Typical detection-level data fusion methods, approaches that fuse detections from individual sensors rather than fusing at the level of the raw data, do not account rationally for the relative reliability of different sensors, nor the redundancy often inherent in multiple sensors. Representative examples of such detection-level techniques include logical AND/OR operations on detections from individual sensors and majority vote methods. In this work, we exploit a statistical data model for the detection of mines and mine-like targets to compare and fuse multiple sensor channels.

Our purpose is to quantify the amount of knowledge that each polarimetric or thermal channel supplies to the detection process. With this information, we can make reasonable decisions about the usefulness of each channel. We can use this information to improve the detection process, or we can use it to reduce the number of required channels.

Keywords: mine, sensor fusion, statistical, multispectral, polarimetric, thermal, REMIDS, UXO, detection

1. INTRODUCTION

The need for mine and unexploded ordnance (UXO) detection and removal is growing in both military and humanitarian applications. In places like Cambodia, the threat of mines to the general populace is overwhelming. Since the end of the Cold War, there has been a growth in smaller, regional conflicts where threats are often not from high-tech weaponry, but from inexpensive ordnance. Bosnia is just one example. Mines are one of the least expensive weapons available, and their threat often far outlasts the conflict for which they are emplaced. Mine detection is therefore necessary in two roles: detecting minefield obstacles for military intelligence and detecting individual mines for eventual removal.

One specific subject of mine detection involves wide-area surveillance. In one case, military forces need to chart possible impediments to ground movement accurately over broad swaths of territory. In humanitarian applications, surveyors examining terrain for mine clearance can limit the area searched with wide-area surveillance. For both situations, it is highly desirable to conduct mine searches from the air to minimize the danger to personnel. Unfortunately, ground penetrating radar and quasistatic electromagnetic approaches are somewhat limited in range.

Optical techniques can meet many of the requirements of wide-area minefield detection. One area of much interest is polarimetric sensing. In the near infrared domain, surface mines have a highly polarizing characteristic. Disturbed soil, indicative of buried objects, may also have such a polarizing nature. In fact, mine detection rates are high using only polarization information, but reducing the false alarm rate is a harder problem, requiring that we apply additional information. While this feature alone is inadequate to detect mines, it can be a powerful tool when combined with multispectral information, such as thermal imaging or imaging spectrometers.

However, fusing information from multiple sources in a rational, statistical way is often neither simple nor obvious. In previous work, we showed that fusing local spatial information with polarization and thermal observations can reduce false alarms, either through a statistical prior model or local adaptivity.¹ In this paper, we focus solely on the

R.A.W.: Email:lorax@bu.edu; WWW: <http://mdsp.bu.edu/lorax.html>

relative advantages of fusing combinations of polarization, reflectance, and thermal data for distinguishing mine-like targets from background. Counterintuitively, we find that fusing only polarization and thermal information gives better results than any other combination, including the combination of all channels together, even under conditions of low thermal contrast. This finding is likely caused by model mismatch, but it clearly points out the complexity of choosing an appropriate fusion method. System designers should not develop sensors, models, and algorithms in isolation. We must develop these technologies together to achieve optimal solutions.

2. SENSOR BACKGROUND

Our data consist of three infrared imagery channels generated by the U.S. Army Engineer Waterways Experiment Station's Remote Minefield Detection System (REMIDS).^{2,3} The system images the first two channels using an active infrared system at the $1.05\ \mu\text{m}$ wavelength. One channel is percent polarization, $(P - S)/(P + S)$, and the other channel is total reflectance, $(P + S)$, where P is reflectance in parallel polarization and S is reflectance in cross polarization. The third channel is a passive thermal infrared channel operating over the $8\text{-}12\ \mu\text{m}$ range. The sensor is mounted on an airborne platform, represented in Figure 1. The data that we use for this study has a resolution

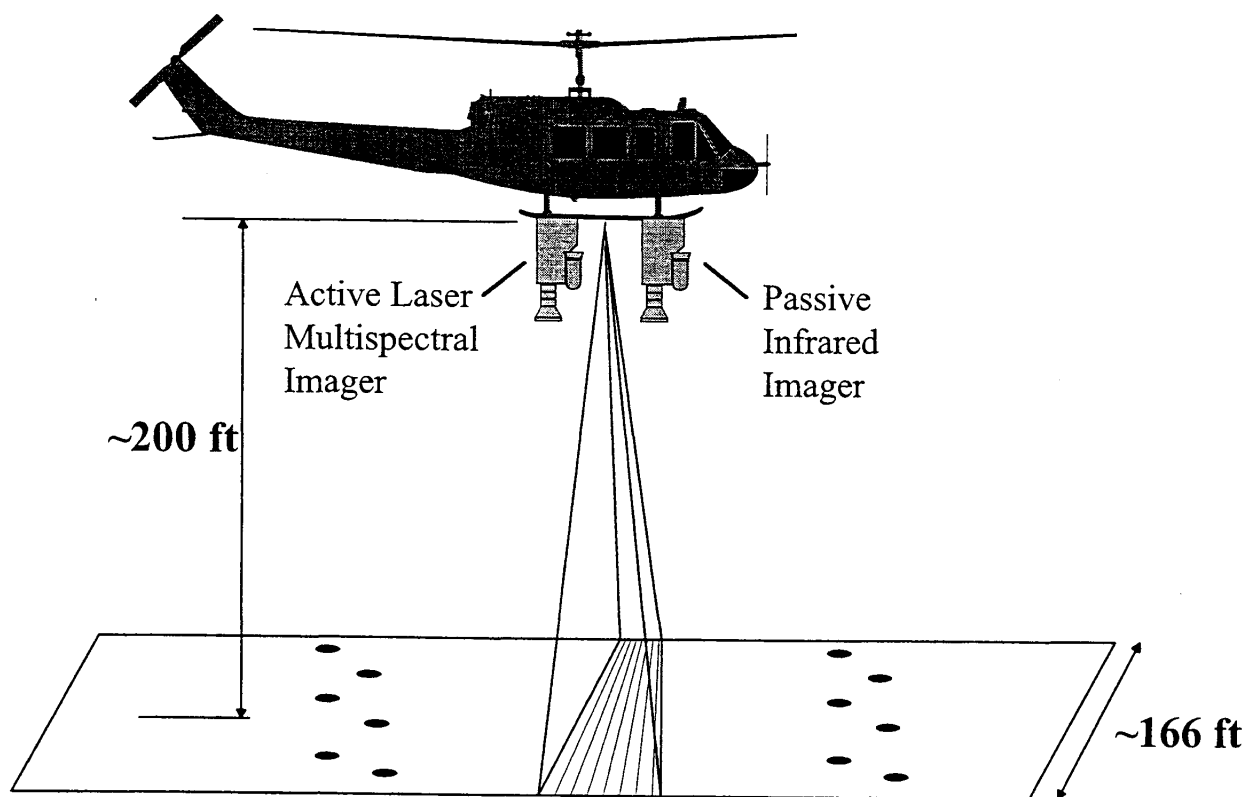


Figure 1. REMIDS Thematic Representation

of 2-3 inches per pixel, depending on altitude, although later system upgrades have increased this resolution.

Several researchers have shown that polarization characteristics are advantageous in the mine detection process.^{2,4,5} Man-made objects such as mines and other unexploded ordnance (UXO) are significant near infrared polarizers compared to natural backgrounds, which tend to be uniformly random polarizers. Disturbed earth can indicate buried objects, and may also be a source of polarization features. REMIDS only uses linear polarization features, but we can extend the same algorithmic techniques to circular polarization as well. Additionally, thermal data alone is sensitive to weather and time of day, especially diurnal thermal crossover.

The active near-IR sensor used in REMIDS is capable of all-weather imaging, but this all-weather capability comes at the price of a laser source, its power supply, and other support equipment. A passive polarization sensor could be implemented cheaply, but would have weather-dependent detection and false alarm rates. An alternative system is a passive hyperspectral polarimetric imager,⁴ but this again increases the system cost. Additionally, the computational requirements of the system increase approximately as the square of the number of channels used in the detection process, so every channel that we can ignore reduces the overall cost of the system. This last cost is the main reason we are fusing different combinations of sensors in this study: to determine if intelligently choosing channels to fuse can provide such savings with little loss in detection capability.

One difficulty with the data set used in this study is that the percent polarization channel had been saturated prior to the signal processing stage to “improve” usage of available dynamic range. Unfortunately, doing so results both in a decrease in signal-to-clutter ratio in the percent polarization channel and an unrealistic decrease in the variance of the mines in this channel. This approach also saturated many background pixels, and thus these pixels cannot be distinguished from mine pixels through the percent polarization channel alone. To distinguish these pixels, we must use data from the other channels or prior information. We will return to this issue in Section 4.

3. METHOD

To analyze and compare the advantages of the various channels and their combinations, we choose to use a simple Maximum Likelihood (ML) approach to detection:

$$\hat{x} = \underset{x}{\operatorname{argmax}} p(y|x) = \underset{x}{\operatorname{argmax}} \ln p(y|x) \quad (1)$$

where y is the multichannel image and x is the hypothesis image of mine locations, each represented as a vector,

$$y = [y_1^T, \dots, y_N^T]^T \quad x = [x_1, \dots, x_N]^T$$

$$y_i = \begin{bmatrix} y_i^p \\ y_i^r \\ y_i^t \end{bmatrix} \in \mathbb{R}^3 \quad x_i = \begin{cases} 1 & \text{if pixel } i \text{ is a mine} \\ 0 & \text{if pixel } i \text{ is not a mine} \end{cases}$$

and N is the total number of pixels. Thus, each element of the observation image, y_i , is a 3-vector where the superscripts p , r , and t denote the three channels, percent Polarization, total Reflectance, and Thermal, respectively. For our channel comparison analysis, we sometimes ignore one or two channels. For these cases, y_i becomes either a 2-vector or a scalar, respectively.

For our observation model, we assume that each pixel of data, y_i , is conditionally independent of all other data pixels, conditioned on knowledge of mine presence at that pixel, x_i . Formally,

$$p(y|x) = p(y_1, \dots, y_N | x_1, \dots, x_N) = \prod_{i=1}^N p(y_i | x_i) \quad (2)$$

This assumption allows us to perform the maximization in Equation 1 individually for each pixel,

$$\hat{x} = \underset{x}{\operatorname{argmax}} \ln p(y|x) = \underset{x}{\operatorname{argmax}} \sum_{i=1}^N \ln p(y_i | x_i)$$

$$\hat{x}_i = \underset{x_i}{\operatorname{argmax}} \ln p(y_i | x_i) \quad (3)$$

Thus, we formulate the problem as a likelihood ratio test,

$$\frac{p(y_i | x_i = 1)}{p(y_i | x_i = 0)} \underset{\hat{x}_i = 0}{\overset{\hat{x}_i = 1}{\geq}} \alpha \quad (4)$$

where α is a threshold chosen by the user.

We also assume that each pixel is a conditionally identically distributed Gaussian, with parameters that are a function mine presence at the pixel of interest,

$$p(y_i|x_i) = \frac{1}{|2\pi\Sigma(x_i)|} \exp \left\{ -\frac{1}{2} (y_i - \mu(x_i))^T \Sigma(x_i)^{-1} (y_i - \mu(x_i)) \right\} \quad (5)$$

where $\mu(x_i)$ and $\Sigma(x_i)$ are the mean and the covariance, respectively, of the observation variable, y_i . For the purposes of this paper, we assume that the mean and covariance are known for both background and mines, and we estimate the mean and a full covariance matrix for both mines and background directly from the data. We make no attempt to model or adapt to nonstationarity of the background in this study. The likelihood ratio test then simplifies as follows,

$$[y_i - \mu(0)]^T \Sigma^{-1}(0) [y_i - \mu(0)] - [y_i - \mu(1)]^T \Sigma^{-1}(1) [y_i - \mu(1)] \underset{\hat{x}_i=0}{\overset{\hat{x}_i=1}{\geq}} \gamma \quad (6)$$

Note that these decision threshold surfaces are quadratic, not linear, because the covariances are not equal.

The simplicity of this model is desirable because it makes the resulting analysis and comparison of the channels easier to understand. One possible drawback, however, is that this very simplicity may not be a good match to reality, and may produce unexpected results.

4. RESULTS

For our test we have access to only two data sets, both from 1991 test flights of REMIDS over Fort Drum, New York. The first flight took place at 9 am on an overcast day, resulting in low thermal contrast between mines and background, but also low thermal variance. The second flight occurred at 3 pm on a clear, sunny day. We show representative segments of all channels of both scenes in Figures 2 and 3. The mines in these images are surface patterned anti-tank mines, but this approach to comparing the relative usefulness of the channels is equally applicable to scattered mines as long as sufficient spatial resolution is available. The applicability to buried mines depends on whether the advantages of polarization information extend to buried mines that disturb the surface of the soil in which they lie. Thermal detection of buried mines is an established technique.

For the purposes of our analysis, we define a detection as any continuously 8-connected region of pixels labeled as a mine by the ML approach. Using an image ground-truthed by hand, we define a correct detection as any single detection that coincides with at least one pixel of any continuously 8-connected region of pixels in the ground-truth image. If two such detection regions coincide with a mine region, we only count one as a correct detection. We define the total number of false alarms as the difference between the total number of detections and the number of correct detections. This definition is imperfect; as the threshold, γ , of Equation 6 increases, an initially continuous detection region can break into multiple disjoint regions, resulting in an increase in the number of false alarms, without a related increase in correct detections. Using a Maximum A Posteriori (MAP) detector with a smoothness prior model can relieve this problem.¹

We show the resulting empirically derived Receiver Operating Characteristic (ROC) curves for the clear and cloudy days in Figures 4 and 5, respectively. Each graph contains multiple curves, one for each possible combination of the three channels: percent polarization, total reflectance, and thermal. We label the graph legends with the same convention used in the superscripts of Section 3: 'p' for percent Polarization, 'r' for total Reflectance, and 't' for Thermal. We denote combinations of channels by multiple letters, e.g. 'rt' represents the combination of the total reflectance and thermal channels. Since these are semilog plots, we assigned a value of 0.99 to any cases of zero false alarms, for display purposes.

There are three unusual features of these ROC curves that we must interpret. We explained the first such feature previously, the occasional increase in false alarms as correct detections decrease, as the breaking of a detection into multiple smaller detections as detection threshold increases. The second is the square ROC curve associated in each graph with the polarization channel alone: a sudden decrease from 1 to 0 in the probability of correct detection as you trace the curve from the right with no change in false alarms, followed by an immediate decrease in the number of false alarms to zero. This odd characteristic is a result of the saturation of the polarization channel described in Section 2. There are a fixed set of false alarms and correct detections that exactly attain the upper limit of the range

of values on which the polarization channel is defined. As the detection threshold approaches this upper limit it rules out all other false alarms. As the threshold crosses the top of the range, both correct detections and false alarms plummet to zero immediately. While the graph we chose to represent this is not unique, it is the closest match to our intuition about the shape and structure of ROC curves.

The third unusual feature involves the order of the ROC curves from lower right to upper left. From a statistical viewpoint, adding channels, and hence adding information, cannot result in a worse detection characteristic unless the model used does not match the data. Yet our ROC curves in both types of weather clearly show an increase in false alarms as the reflectance channel is included with the polarization and thermal channels. Similarly, the reflectance channel alone appears to perform better than the combined thermal and reflectance channels on the overcast data. There are two possible ways that we can explain this counterintuitive result. One is our simplifying assumption of background stationarity, which a cursory perusal of the data will show is clearly inaccurate. Including model adaptivity may fix this problem. Another possibility could be a sensitivity in the reflectance channel to the parameter estimates, particularly its covariance with the thermal channel. Each of the unusual cases involves combining the reflectance channel with the thermal channel.

In the other cases, the ROC curves follow our intuition about fusing channels: fusing one channel with another results in an improvement over either channel separately. The interesting result is the amount of improvement. First, note that the polarization and reflectance false alarm rates, alone and separately, do not change much as a function of the weather conditions. We could anticipate this, since a narrowband active sensor should be influenced little by ambient radiation. The thermal channel and all of its combinations outstandingly improve on a clear, sunny day. Again, we could anticipate this since thermal contrast increases dramatically with input radiation. Surprisingly, even on the overcast day, the thermal channel in conjunction with the polarization channel gives a significant and useful improvement, more than halving the number of false alarms of either channel separately. The reflectance channel, on the other hand, provides a much more limited improvement when used in conjunction with the polarization channel under all conditions.

5. CONCLUSIONS

In this research, we have examined the relative value of polarization, reflectance, and thermal information in the particular context of the mine detection problem. Using a particular simplified model we have shown that polarization and thermal information, when fused in a statistically rational way, can significantly improve the detection process over either alone, primarily through a reduction in false alarm rate. Also the thermal channel provides significant information, even under conditions of low thermal contrast.

The advantages of expending the computational effort to include the reflectance channel appear more dubious. The improvements in false alarm rate that this channel provides are relatively small when compared to the other two channels. Also, it appears from a cursory examination of the results that the reflectance channel may be particularly sensitive to model mismatch. However, a careful analysis of the causes of this difficulty will require further study.

6. ACKNOWLEDGMENTS

We would like to thank Ricky Goodson and Ernesto Cespedes of the US Army Corps of Engineers Waterways Experimental Station's Environmental Laboratory for graciously sharing their REMIDS data. This work is partially supported by the Army Research Office under Grant ARO DAAG55-97-1-0013, the Air Force Office of Scientific Research under Grant F49620-96-1-0028, and the National Institutes of Health under Grant NINDS 1 R01 NS34189.

REFERENCES

1. R.A. Weisenseel, W.C. Karl, D.A. Castañon, C.A. Dimarzio, "Statistical approach to multi-channel spatial modeling for the detection of mine-like targets", *Detection and Remediation Technologies for Mines and Mine-Like Targets III*, A.C. Dubey, J.F. Harvey, J. Broach, editors, Proc. SPIE V 3392, SPIE, Orlando, (1998)
2. B.H. Miles, E.R. Cespedes, R.A. Goodson, "Polarization-Based Active/Passive Scanning System for Minefield Detection", *Polarization and Remote Sensing*, Walter G. Egan, editor, Proc. SPIE V 1747, pp239-252, SPIE, San Diego, (1992)
3. G. Maksymonko, K. Breiter, "ASTAMIDS Minefield Detection Performance at Aberdeen Proving Ground Test Site", *Detection and Remediation Technologies for Mines and Mine-Like Targets II*, Abinash C. Dubey, Robert L. Barnard, editors, Proc. SPIE V 3079, pp726-737, SPIE, Orlando, (1997)
4. M.A. LeCompte, F.J. Iannarilli, D.B. Nichols, R.R. Keever, "Multispectral IR Signature Polarimetry for Detection of Mines and Unexploded Ordnance (UXO)", *Detection Technologies for Mines and Mine-Like Targets*, Abinash C. Dubey, Ivan Cindrich, James M. Ralston, Kelly Rigano, editors, Proc. SPIE V 2496, pp180-192, SPIE, Orlando, (1995)
5. B.A. Barbour, S. Kordella, M.J. Dorset, B.L. Kerstiens, "Mine Detection Using a Polarimetric IR Sensor", *The Detection of Abandoned Landmines: a Humanitarian Imperative Seeking a Technical Solution*, EUREL International Conference, V 431, pp78-82, IEE, Edinburgh, (1996)

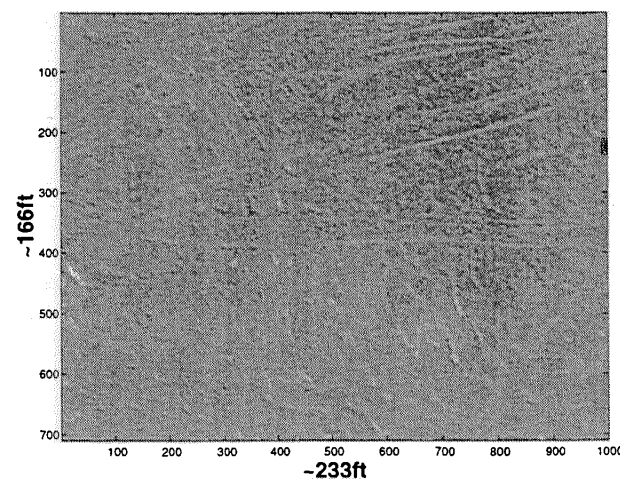
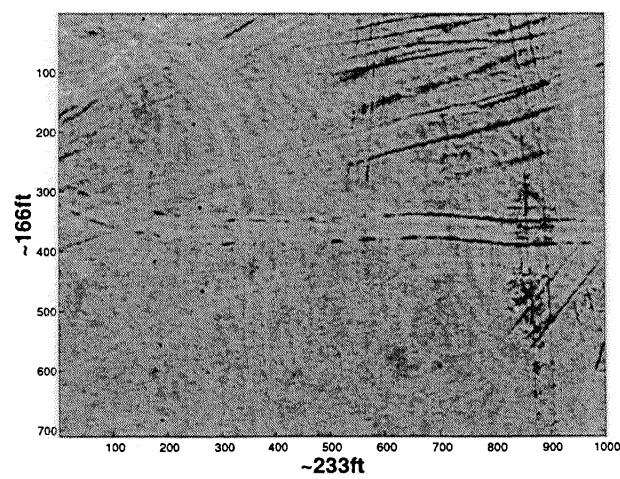
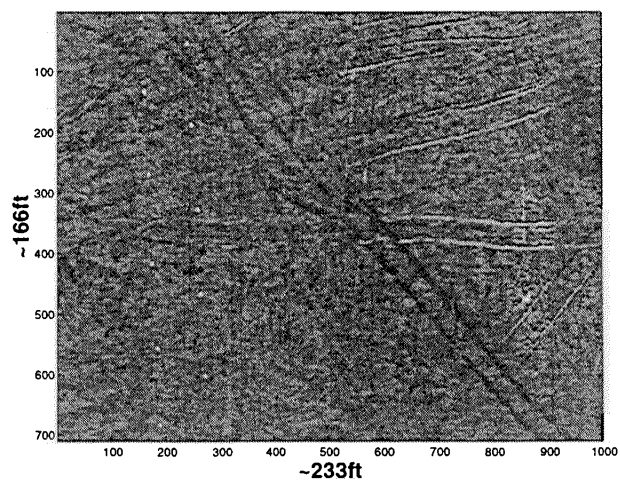


Figure 2. 9 am, Overcast, Percent Polarization, Reflectance, and Thermal

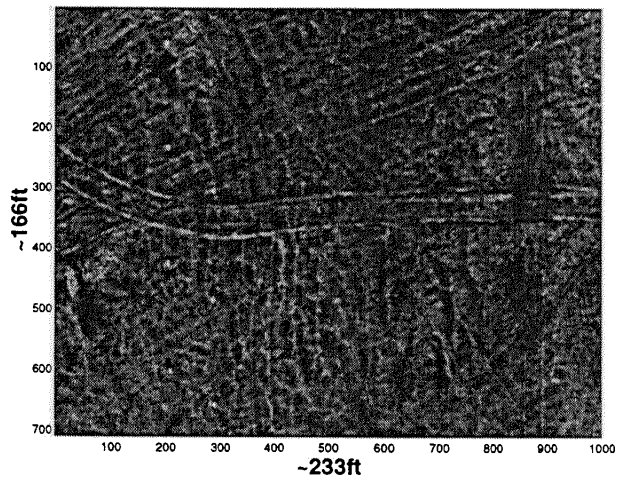
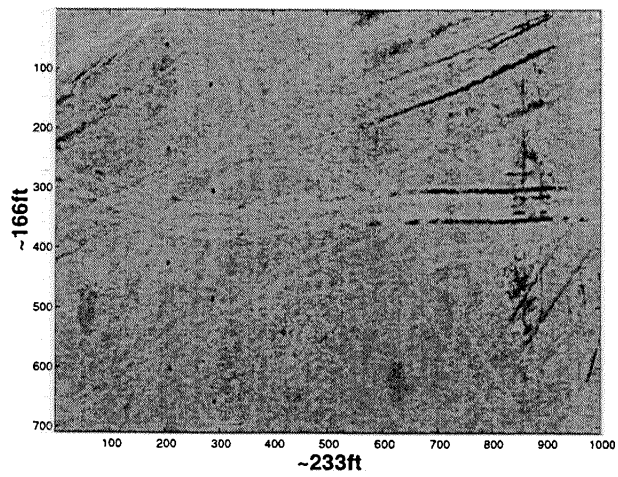
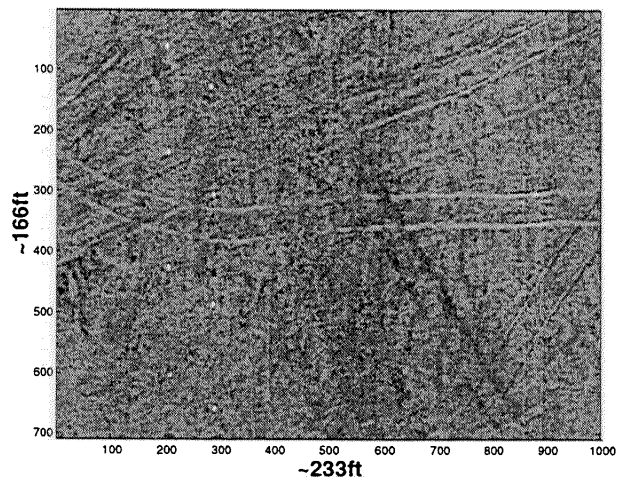


Figure 3. 3 pm, Clear, Percent Polarization, Reflectance, and Thermal

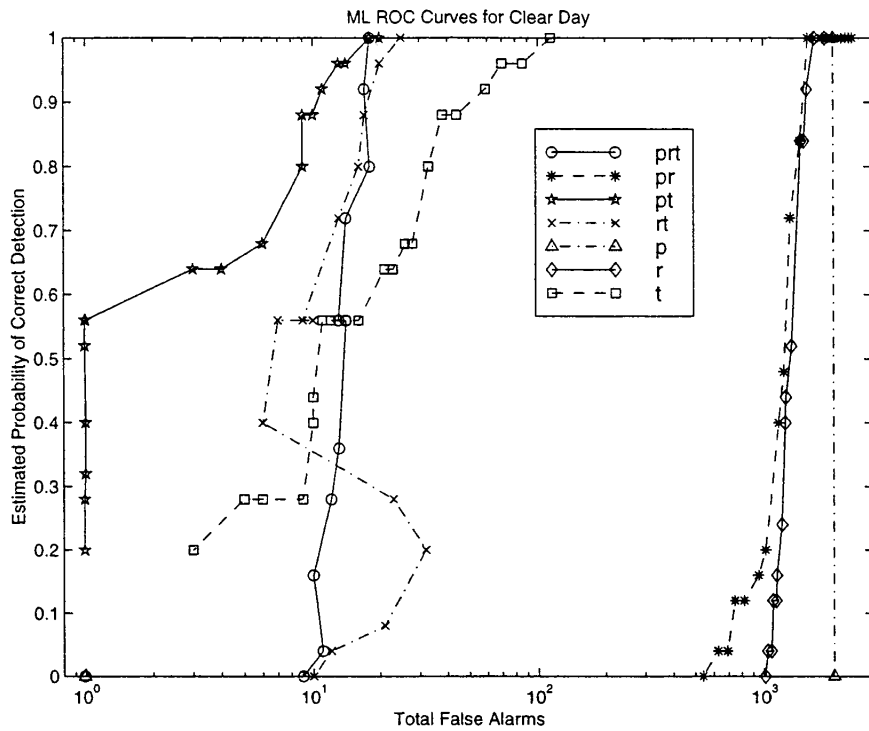


Figure 4. Experimental ROC for ML Detector - Clear Day

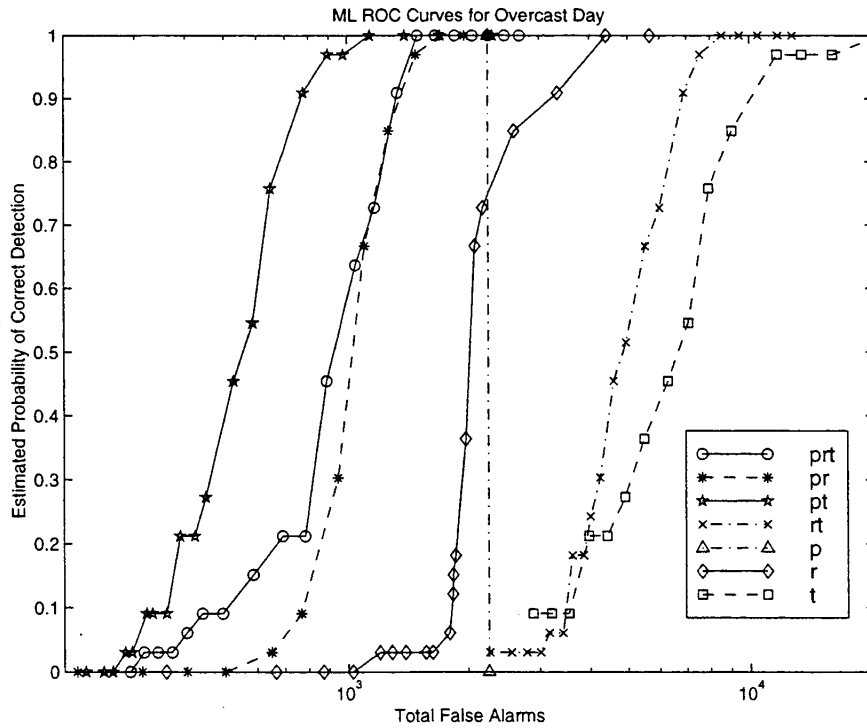


Figure 5. Experimental ROC for ML Detector - Overcast Day

Laser-induced acoustic generation for buried object detection

S. W. McKnight^a, C. A. DiMarzio^a, W. Li^a, R.A. Roy^b

^aCenter for Electromagnetics Research, Northeastern University, Boston, MA 02115

^bDepartment of Aerospace and Mechanical Engineering, Boston University, Boston, MA 02215

ABSTRACT

Mechanisms for the production of acoustic energy in soil by pulsed CO₂ laser excitation of the surface are reported. When the laser pulse is unfocused with a spot size about 1 cm in diameter, a single narrow acoustic pulse is observed with a spectral content near the detector limit of 100 kHz and a velocity of 255 m/s, close to the speed of sound in air. When the laser is focused to a spot size on the order of 1 mm diameter, the audible acoustic intensity is greatly increased and we observe a second broad acoustic feature. This feature has a much lower frequency (near 3 kHz) and velocity (75 m/s). We have tentatively identified the fast mode as a normal compressive mode and the slow mode as a Biot slow-wave. A study of visible light emission when the focused CO₂ laser beam strikes the sand surface indicates ionized nitrogen, oxygen, and silicon are present. This implies that the mechanism for sound production with the focused beam involves ionization by the optical electric field, expansion, and subsequent collapse of the air. The mechanism for sound production by the unfocused beam, which produces better imaging of underground objects, appears to be quite different.

Keywords: laser-induced, acoustic, landmine detection, Biot waves, porous media

1. INTRODUCTION

The use of high-frequency (~30kHz) acoustic waves produced by a pulsed CO₂ incident on the surface of dry sand for imaging of shallow buried objects such as anti-personnel landmines has been demonstrated in the laboratory^{1,2}. The mechanisms for the conversion of the optical pulses into sound, however, is poorly understood. Since the optimization of the process of optical to acoustic conversion in a porous media depends on understanding and modeling the physical processes, we have initiated an experimental investigation of the behavior of the acoustic modes created in soil under different laser pulse focusing conditions. For a broad, unfocused laser spot, modeling the soil as a uniform effective medium as in Figure 1a may be appropriate. If the laser is focused to a small spot comparable to the size of the sand grains as in Figure 1b, a more complex calculation may be necessary which takes into account the random position and orientation of the sand grains and the intervening air spaces. In addition, under focused beam conditions the optical field intensity can become very large, leading to different physical effects. We have observed

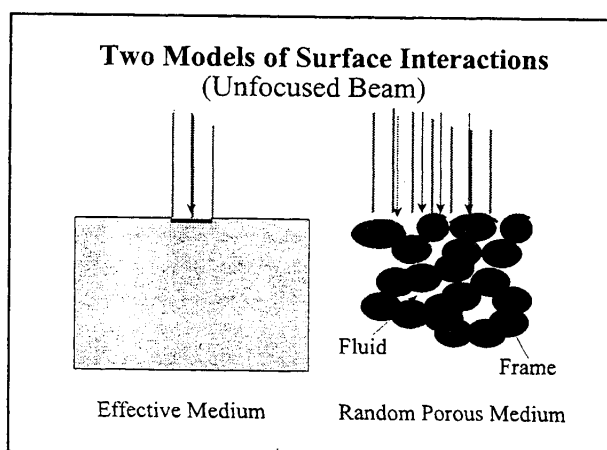


Figure 1 Conceptual models of optical-to-acoustic conversion in soil. a) An effective medium approximation may be adequate when the laser spot is much larger than the structure of the soil, b) The details of the interaction of the light with the discrete particles of the poro-elastic medium may be critical when the beam size is comparable to the particle size.

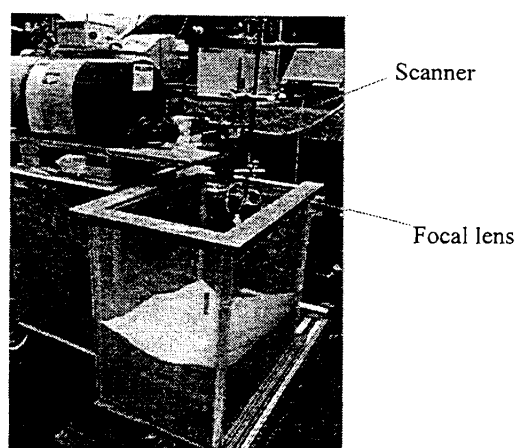


Figure 2 Photograph of experimental setup. The pulsed CO₂ is at top left and the experimental sandbox is below. Objects to be detected or hydrophones to study the laser-to-acoustic coupling are buried in the sand.

that focusing the laser beam to a spot size on the order of 1mm in diameter greatly increases the audible acoustic sound associated with the laser pulse and, in addition, can cause the production of a flash of visible light at the point where the infrared CO₂ pulse impinges on the sand. We have studied the time-dependence of the acoustic pulse at various depths and angles in the soil for both focused and unfocused laser pulses. For unfocused pulses we find a single acoustic mode with a velocity slightly less than the speed of sound in air with a directionality perpendicular to the soil surface. With a focused beam we detect two distinct acoustic features in the time trace: a sharp pulse with a velocity and spectral content similar to that achieved with the unfocused beam, and, in addition, a second broad pulse with a much lower frequency content which propagates at a slower velocity. We suggest that these two modes may be related to the two Biot modes allowed in poro-elastic media³.

2. EXPERIMENT

A photograph of the experimental setup is shown in Figure 2. The source laser is a LSI pulsed 10.6 μ CO₂ TEA laser with a pulse length of 100ns, a pulse energy of 150mJ, and a repetition rate less than 20Hz. The laser is incident on a 1 m² surface of dry sand about 60 cm deep in which we would bury detectors or various types of subsurface targets. In the present studies of the optical-acoustic conversion mechanism, the acoustic signal was measured with a wide-band hydrophone (0-100kHz) which was buried at various depths and positions in the sand. This configuration contrasts with the experimental setup that we used to image underground objects. For imaging, our best results were obtained with a narrow-band (30kHz) tuned PZT detector which was suspended in the air above the position of the laser pulse on the surface of the sand.

The laser was used in two configuration: unfocused or with a focusing lens to reduce the spot size. The unfocused laser spot is an oval approximately 1.1 \times 0.7 cm in size. A germanium lens with a focal length of 25cm was used to focus the beam to a spot size less than 1mm in diameter, comparable to the size of the sand grains. 10.6 μ light is very effectively absorbed in soil with a reflection less than 10% and skin depth of approximately 5 μ . The sound produced by both the focused and the unfocused beams was audible across the room. With the focused beam the sound was considerably louder and there were flashes of visible light emitted and some movement of sand grains at the spot where the laser hit the surface.

The acoustic signal detected by the buried hydrophone for the focused and unfocused beams is shown in Figure 3 and Figure 4.

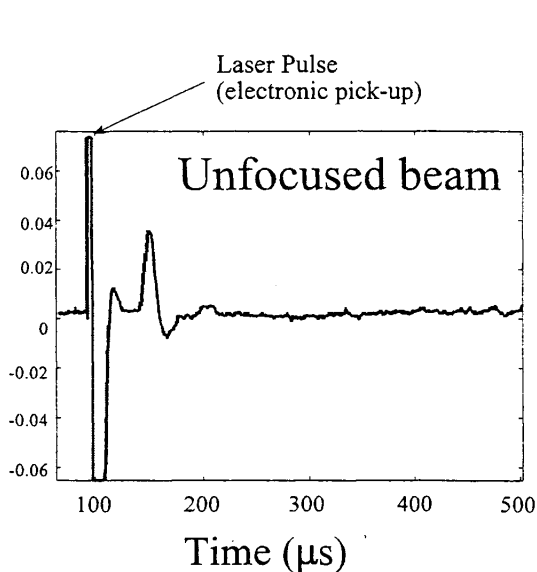


Figure 3 The time trace of the acoustic signal measured by a buried hydrophone as a result of a pulse from the unfocused laser on the sand surface.

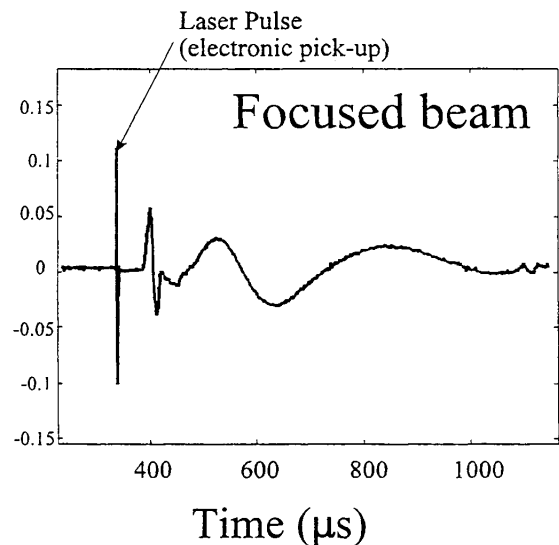


Figure 4 The time trace of the acoustic signal measured by a buried hydrophone as a result of a pulse from the focused laser on the sand surface.

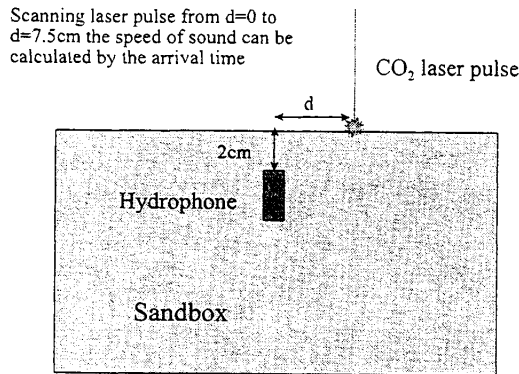


Figure 5 Experiment to measure the velocity in sand of the acoustic features in Figure 3 and 4.

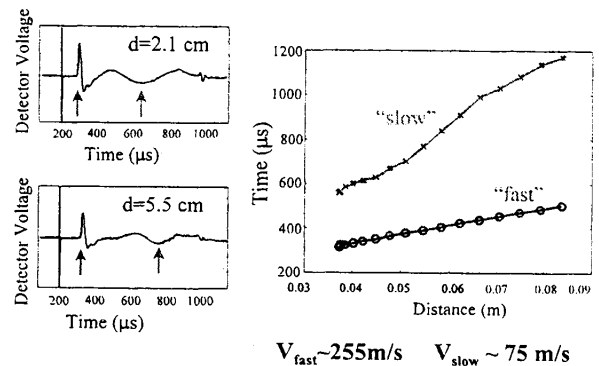


Figure 6 The arrival time vs. distance for the features indicated by the arrows in the inset. The propagation velocity in sand of the two features is determined as indicated from the slope of the lines.

The initial spike is electrical noise caused by the laser discharge. The acoustic pulse is dominated by a sharp peak with a width (about $10\mu\text{s}$) that is probably limited by the bandwidth of the 100kHz detector. For the focused beam only, there is a second, much broader feature with a peak-to-peak width of about $300\mu\text{s}$.

To determine the propagation speed of these two different acoustic features observed with a focused laser beam, we measured the change in the acoustic signal as a function of distance from the laser spot to the detector using the configuration in Figure 5. The time trace of the acoustic signal measured by a buried hydrophone as a result of a pulse from the focused laser at two different positions on the sand surface are shown on the left in Figure 6. Identifying the sharp feature by the initial spike, and the broad feature by the dip between the two broader peaks, in Figure 6 we plot the time delay after the laser fires of these two features as a function of distance from the laser spot to the detector. By taking the slope of these two curves, we find the velocities of the two features are dramatically different. The sharp feature propagates with a velocity around 255 m/s, while the broad feature propagates with a velocity of about 75 m/s.

The observation of acoustic modes with two different velocities in porous media is a well-known phenomena. In the theory of Biot, the solution for acoustic propagation in a random porous elastic media gives rise to two modes: a mode in which the frame and the fluid move in phase (compressional wave) and an additional mode in which the motion of the frame is out of phase with the motion of the fluid. Typically the in-phase mode has a velocity which is close to the velocity of sound in the fluid, while the out-of-phase mode, the "Biot slow wave" has a lower velocity. If we tentatively make an identification of the features we observe with the Biot modes, the sharp feature which travels at 255 m/s, close to the velocity of sound in air, could be the Biot in-phase mode, while it is natural to identify the broader feature, low-velocity feature with the Biot slow wave. Note that we only observe this second, slower wave under excitation with a focused laser beam.

To measure the angular dependence of the magnitude of the acoustic signal from the focused beam, we positioned hydrophones at different angles along an arch buried in the sand centered on the position of the laser spot, as shown in Figure 7. In Figure 8 and 9, we plot the peak amplitude of the sharp "fast-wave" signal and the

Angular Distribution of Acoustic Intensity

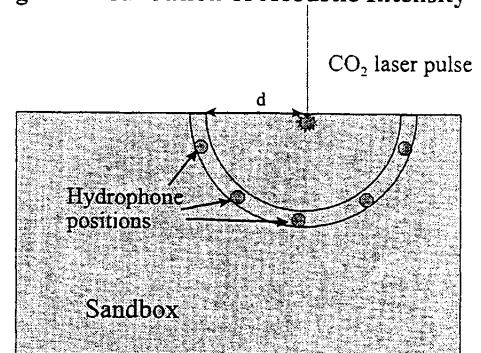
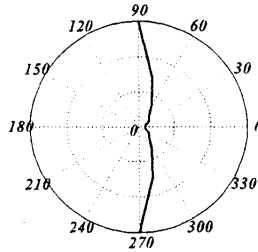
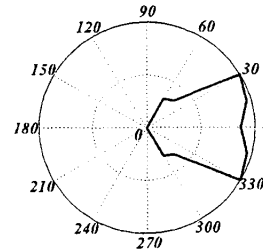


Figure 7 Experimental setup to measure the angular directivity of the acoustic features in Figure 3 and 4.

**Fast Wave Angular Distribution
(Focused laser beam)**



**Slow Wave Angular Distribution
(Focused laser beam)**



$$\text{Rayleigh distance} = D^2/\lambda \lesssim 0.01 \text{ cm}$$

Figure 8 Angular dependence the amplitude of the “fast wave” feature produced with a focused laser beam.

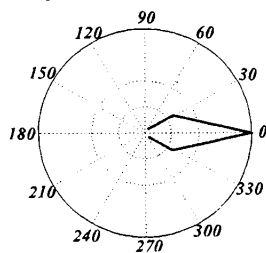
Figure 9 Angular dependence the amplitude of the “slow wave” feature produced with a focused laser beam.

broad “slow-wave” signal as a function of angle from the normal. The two modes show a dramatic difference in angular distribution, with the slow wave signal propagation peaked at angles perpendicular to the surface, while the fast-wave signal strength is peaked along the surface of the sand. The angular dependence of the fast-wave mode is hard to understand, since when the beam is unfocused we observe only a single mode with a shape and velocity similar to the fast-wave signal, but with an angular distribution sharply peaked at a direction normal to the surface, as shown in Figure 10.

The unfocused beam has a Rayleigh length d^2/λ on the order of 1 cm, so for distances into the sand on the order of 1 cm the unfocused beam looks like an extended source. The focused beam, on the other hand, has a Rayleigh length less than a millimeter and is a point source to a good approximation. Since the radius of the detector arch was about 10 cm, it is unlikely, however, the difference between the fast-mode angular dependence seen in Figure 9 and 10 can be accounted for by such physical optics effects. It is more likely that the different acoustic excitation physics between the broad laser beam and the focused beam, such as illustrated in Figure 1, may account for these differences. We will return to this question later.

Table I summarizes the differences between the two acoustic modes that are observed with the focused laser beam. We note that these differences may impact on the utility of the two modes for underground detection and imaging. The fast mode, having a

**Angular distribution
(Unfocused Laser Beam)**



$$\text{Rayleigh distance} = D^2/\lambda \gtrsim 1 \text{ cm}$$

Figure 10 Angular dependence of the amplitude of the fast-wave feature produced by an unfocused laser beam.

Table I: Comparison of Fast-Wave and Slow-Wave laser-induced acoustic features produced with a focused laser beam.

Comparison of Fast and Slow Acoustic Waves

	Fast Wave	Slow Wave
Unfocused Beam	Yes	No
Focused Beam	Yes	Yes
Speed	250 m/s	75 m/s
Frequency peak	100 kHz*	3 kHz
Directivity	Along Surface	Perpendicular to Surface
$\alpha(\omega)$	Unknown	Unknown

* Band-width of detector=100kHz ($1/\Delta t_{\text{laser}} = 10 \text{ MHz}$)

higher spectral frequency content may be more useful for high-resolution imaging where a small wavelength is desired. On the other hand, if we have correctly identified the slow wave as a Biot out-of-phase wave, it will see an exceptionally large contrast between the porous soil where the slow wave is allowed and any subsurface non-porous object where, since there is no longer any differentiation between frame and fluid, its propagation is forbidden. There may also be implications about detection strategies, since to the extent that the energy of the slow-wave is carried in the fluid, techniques that measure the motion of the solid surface, such as laser vibrometry, will be less effective than with an ordinary compressional wave. On the other hand, for the same reason the mismatch between the poro-elastic wave and the wave in the air may be less for the slow-wave than the fast wave, so that detection by means of a microphone above the soil may be more favorable for the slow wave.

It is noteworthy that our attempts so far at subsurface imaging with the focused beam have been less successful than our previous work with the unfocused beam. These imaging studies were carried out with an acoustic detector suspended in the air above the surface. It is not possible to say at present whether the lack of success was due to an inherent effect such as the longer wavelength of the slow wave, less efficient acoustic coupling to or from the soil and air, or some interference effects between the fast and slow acoustic waves, or more incidental effects such as noise from the movement of sand grains and drifting of the laser focus due to heating of the air and movement of the surface.

One clue about the mechanism for acoustic excitation is that a flash of visible light is observed which the CO₂ pulse hit the surface of the sand when the beam is focused, but not for the unfocused beam. Origins of this visible light that we considered were black-body incandescence of the sand surface, photoexcitation and luminescence of the sand material, and electric field ionization and recombination. The energy density of the focused beam is sufficient to heat the sand over 10⁶ °C if the effects of radiant emission and conduction are neglected, but this does not seem likely on the relatively long (0.1 μs) timescale of the pulse. Luminescence of SiO₂-based materials is usually very broad-band and can extend into the ultraviolet.

To gather more information about the light emitted when the focused CO₂ laser pulse strikes the sand surface, we measured the spectrum of the emitted light using a multispectral filter and a photomultiplier detector as in Figure 11. The measured spectrum is reproduced in Figure 12 and Figure 13. The emission peaks are negative, because the electron current of the photomultiplier

Spectral Signature of Laser Flash (Focused beam)

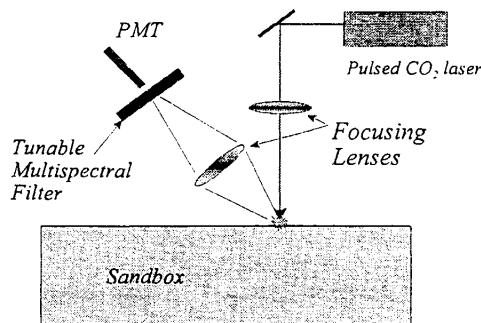


Figure 11 Experimental configuration for measuring the spectrum of visible light produced when a focused laser beam is incident on sand.

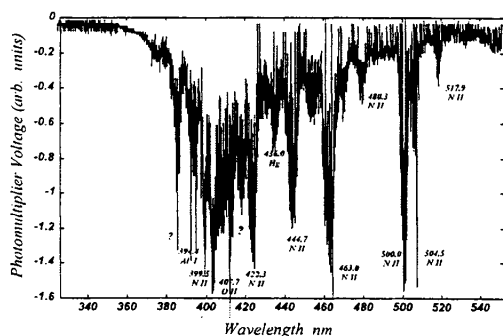


Figure 12 Spectrum of visible light produced by focused laser beam incident on surface of sand (short wavelength). Emission lines are identified from standard handbooks.

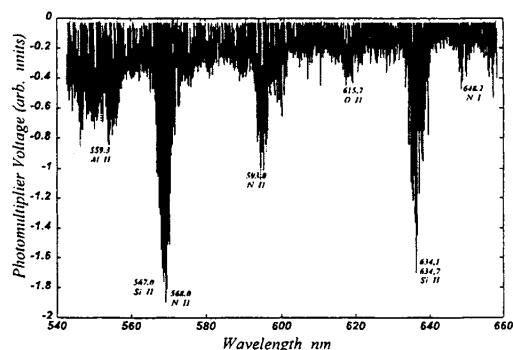
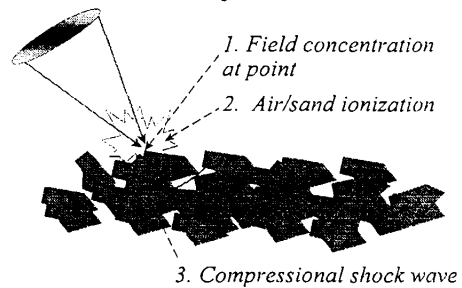


Figure 13 Spectrum of visible light produced by focused laser beam incident on surface of sand (long wavelength). Emission lines are identified from standard handbooks.

is converted to a negative voltage by a voltage-to-current amplifier. Electron photocurrent resulting from the detected light yields a negative voltage.

We have identified emission lines in the spectrum by comparison with the CRC Handbook of Physics and Chemistry. We note that large emission lines associated with ionized nitrogen, oxygen, and silicon are present in the spectrum which implies that the model for optical field ionization is the best model for the production of the light. The concentration of the electric field of the optical signal near the sharp points of the sand grains is the most likely mechanism for the breakdown and ionization of the air and sand.

Focused Beam Surface Interactions



The acoustic production for the focused beam can then be best explained by the collapse of the air following the rapid expansion associated with the ionization and plasma formation as shown in Figure 14. The intense audible noise associated with the focused beam is analogous to the crack associated with an electrical arc or with thunder. One implication of this model is that the sound production for the focused beam is probably primarily as a compressional wave in the air above the soil. The reduced effectiveness of imaging with the focused beam may be a result of the poor coupling of this acoustic compressional wave in air into the soil.

Figure 14 Model for production of light and sound by focused laser beam impinging on soil surface.

The unusual angular dependence of the compressional fast wave for the focused beam seen in Figure 9 may represent the more effective transport of acoustic energy in the air along the surface of the sand than propagation in the sand. In effect, the sound produced by this ionization and collapse travels through the air to a point close to the high-angle, near-surface sensor and then propagates only a short distance into the soil to the detector. Since the velocity of the fast wave is very near the speed of sound in air, it is very difficult to distinguish this propagation channel from propagation through the soil. The slow wave may not have the same channel available to it because a Biot wave will not propagate in a single component fluid such as the air.

3. CONCLUSIONS

The production of acoustic energy in soil by a pulsed CO₂ laser has been studied. When the laser beam is focused to a spot size comparable to the soil grains, a much louder acoustic sound is observed and two different acoustic modes are detected in the soil. One mode, which we identify with a normal compressional acoustic wave has a velocity (255 m/s) close to the speed of sound in air, and a bandwidth of at least 100kHz. The second mode, which we have tentatively identified as a Biot slow-wave phenomena, has a velocity of only 75 m/s and a spectral peak near 3 kHz. By analyzing the visible light emitted by the focused laser beam, we conclude that for the focused laser pulse the acoustic energy is produced by ionization of the air and soil components, with a rapid expansion and subsequent collapse of the air column. Preliminary investigations have indicated that imaging of subsurface objects with the focused beam laser is not as successful as with a broader unfocused beam. This may be a result of the poor coupling of sound produced in air into the soil. Studies to measure the attenuation and dispersion of the acoustic signals produced by the focused and unfocused laser beams are underway.

4. ACKNOWLEDGEMENT

The research described here was supported by the DARPA and ARO under the Multidisciplinary University Research Initiative.

5. REFERENCES CITED

1. S. W. McKnight, C. DiMarzio, W. Li, D. O. Hogenboom, G. Sauermann, "Laser-Induced Acoustic Detection of Buried Objects," SPIE Transactions on Detection of Mines and Mine-like Objects Vol. **3392**, p. 841-847 (1998).
2. S. W. McKnight, W. Li, and C. DiMarzio, "Imaging of Buried Objects by Laser-Induced Acoustic Detection," SPIE Transactions on Detection of Mines and Mine-like Objects, Vol. **3710**, 231-238 (1999).
3. M. A. Biot, "Generalized Theory of Acoustic Propagation in Porous Dissipative Media," J. Acoust. Soc. Am. **34**, 179-191 (1962).

Laser-induced acoustic detection of buried objects

S. W. McKnight, C. DiMarzio, W. Li, D. O. Hogenboom, G. Sauermann

Center for Electromagnetics Research, Northeastern University, Boston, MA 02115

ABSTRACT

We have investigated the use of acoustic energy produced by a pulsed CO₂ laser to detect objects underwater or buried in sand. The CO₂ laser produced 150 mJ pulses of duration 100 ns. The resulting acoustic pulses were detected with an audio microphone with a response to 15 kHz or a PZT transducer with a resonant frequency at 28 kHz. With the laser incident on the surface of a water-filled tank, acoustic echos were observed from the tank walls and from objects in the tank. For objects buried in sand, changes in the acoustic lineshape related to the presence of subsurface objects were observed. Analysis of the data to extract clear signatures of the mine are in progress.

Keywords: laser-induced, acoustic, land mine detection, photoacoustic, optoacoustic

1. INTRODUCTION

The large acoustic mismatch between granular soil and solid subsurface objects suggests that acoustic methods have considerable promise for land mine detection. Sound velocities in sand or granular soil can be on the order of or less than the velocity in air (330 m/s)¹ while in metals or plastics characteristic of land mine casing or fillers, sound velocities are typically greater than 2000 m/s. The velocity and/or density mismatch between soil and mine creates a large acoustic reflection from the mine.

In common with other anomaly detection techniques, the challenge of acoustic detection is to distinguish the effect of a buried mine from stones, roots, and other clutter in the ground. Two separate approaches to acoustic discrimination depend on the identification of acoustic resonances of the target object, or the resolution of the target shape. While mine-like objects may have identifiable acoustical resonances from a few kilohertz to hundreds of kilohertz, these resonances are strongly damped when the object is in soil. Target identification through shape resolution, on the other hand, requires multiple acoustic sources or receivers to correlate the received signal with the position or angle of the source. Coupling acoustic energy from the air into the soil or matching of separate movable ground-contact transducers makes spatial resolution with acoustic transducers difficult.

Optical methods for the detection or generation of sound have advantages for remote position-sensitive acoustic detection. Optical laser vibrometry is a commercially available tool for the detection of sound at a solid surface which has been applied for the detection of subsurface objects.² The photoacoustic effect, production of sound by laser pulses, has been applied for ultrasonic material characterization^{3,4} and for the study of liquid droplet dynamics.^{5,6} In addition, microwave detection of surface acoustic vibrations has also been used for land mine detection.⁷ These optoacoustic techniques create a number of options for the acoustic detection of land mines. First, acoustic transducers can be used both to generate the acoustic signal and to detect the scattered acoustic energy. Second, sound generated by acoustic transducers can be detected by optical (or microwave) signals. This is the approach of References 2 and 7. Alternatively, sound produced by laser pulses can be detected by acoustic transducers. This method is described in this paper.

Figure 1 presents a concept of how a laser-induced acoustic system could be used to locate and identify mines. A pulsed laser is scanned over the ground. The acoustic signal generated by the laser

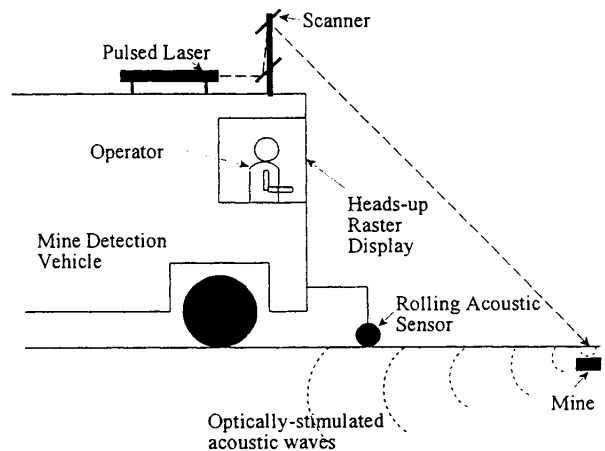


Figure 1. Conceptual laser-induced acoustic land-mine detection technique.

pulses striking the soil is detected by a rolling, ground-contact transducer. Changes in the acoustic signature when the laser pulses are over a subsurface object are encoded into a heads-up display registered with the operator's field-of-view. The correlation of the acoustic signature of an underground object with the position of the laser pulse allows an identification of the location and possibly the shape of the buried object. In concept, this method is similar to location of a structure behind a wall by the change in sound when a percussion source is struck near the target.

It is also possible that sound could be both produced and detected by optical means. This could make possible a remote, non-contact, scanning probe system that could be effective for underground object location and identification. The results of the present investigation will clarify prospects for such a system.

Optical to Acoustic Conversion

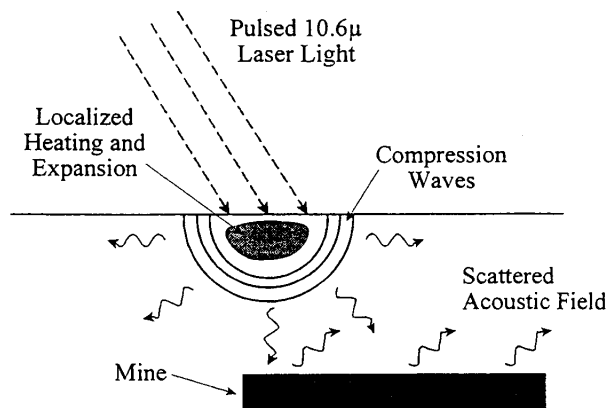


Figure 2. Model for conversion of optical to acoustic energy.

2. EXPERIMENTAL RESULTS

A Laser Science LS150 pulsed CO₂ laser was used to create acoustic pulses on striking a surface. The laser was operating at a wavelength of 10.6 μm and pulse length was 100 ns with an energy per pulse of 150 mJ. Although the pulse repetition rate could be as high as 50 Hz, the data here were taken at a repetition rate of 8 Hz. Acoustic pulses when the laser struck a solid surface were audible at distances of a few meters.

Optical radiation at 10.6 μm is typically over 90% absorbed in soil. Our model for the generation of acoustic energy is that the rapid heating and expansion of the soil in the skin depth of the optical radiation leads to a pressure wave in the solid. This process is illustrated in Figure 2. The thermal expansion is also coupled into an acoustic wave in the air above the surface.

Acoustic energy was detected by two transducers: an audio microphone with a high-frequency cut-off near 15 kHz, and a PZT transducer with a resonant frequency of 28 kHz. Experiments in soil were conducted in a 28x32 cm box filled to a depth of about 7 cm with dry sand. Some experiments were also conducted in a 50x30 cm tank filled to a depth of 30 cm with water. Simulated mine targets were metal disks between 3-10 cm in diameter and 1.5-3 cm in height. While metal targets were chosen for convenience, the acoustic mismatch between soil and plastic is nearly as great, and we expect that equivalent results would be achieved with plastic targets.

Figure 3 shows a typical geometry for experiments in the sand box. The PZT transducer is half-buried in the sand approximately 10 cm from the target, and the laser pulse is scanned in rows across the position of the target. The target is buried 2 cm under the surface of the sand. The experiment is then repeated with the target removed from the sand box. Figure 4 shows the transducer voltage as a function of time after the initial laser pulse when the laser pulse is directly over the target position (Scan 3, y=13.5cm, x=13.5cm). The dotted lines in the same figure show the response when the laser is at the same position, but the target is removed.

While it is clear from Figure 4 that there are differences in the signal caused by the target, it is not clear how to process the data to display these changes. In Figure 5 we have taken

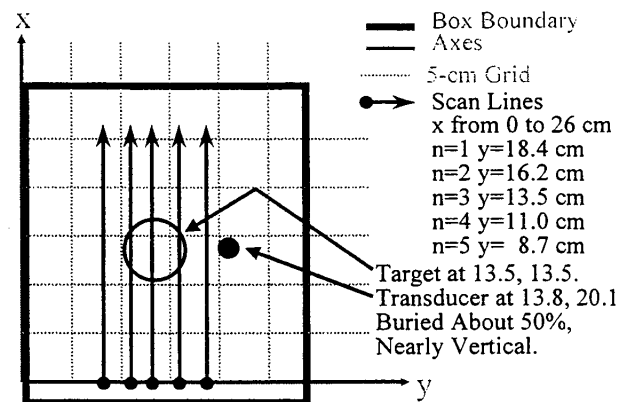


Figure 3. Geometry for experiments in sand.

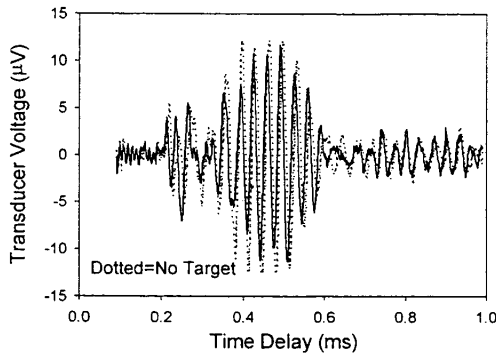


Figure 4. PZT transducer response as a function of time delay following the laser pulse for the geometry of Figure 3 with the laser pulse incident at $y=13.5$ cm, $x=13.8$ cm. The solid line is with the target present and the dotted line is with no target.

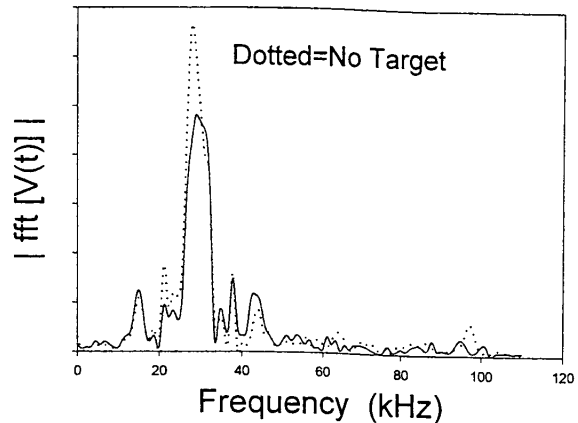


Figure 5. Fourier transform of the data of Figure 4. Solid line is with the target present and the dotted line is with no target.

a Fourier transform of the data in Figure 4. The dominant feature in the plot here is the 28 kHz response peak of the PZT transducer. While there are consistent differences from run to run between the data with the target and without the target, there is no obvious unique identifying feature of the target.

In Figure 6 the difference between the transducer signal with and without the target present is plotted in gray-scale intensity as a function of time delay and laser position for scan number 3 in Figure 3. The effect of the target at $x=13.8$ cm is obscured by the increase in signal strength as the laser excitation approaches and recedes from the point closest to the fixed detector position at $x=13.8$ cm, $y=20.1$ cm. In addition, the difference signal is dominated by changes in the phase of the received signal that may be caused by increased propagation velocity through the target or by slight repositioning error between the data with and without the target.

The arrival time of the initial acoustic pulse is consistent with a propagation velocity near 350 m/s. While as noted above the velocity of sound in sand is highly variable, this is close enough to the velocity of sound in air to leave doubt whether the acoustic signal is propagating primarily through the air or through the sand.

In order to demonstrate conclusively transfer of acoustic energy from the laser pulse into the medium and to improve the chances for detection of a target, we performed experiments on underwater targets in an experimental water tank as illustrated in Figure 7. CO_2 radiation is absorbed very efficiently in water with a reflection coefficient of less than 1% and an absorption length of a few microns. The response of a completely submerged detector was measured as the laser beam was scanned toward an aluminum target extending nearly to the surface of the water.

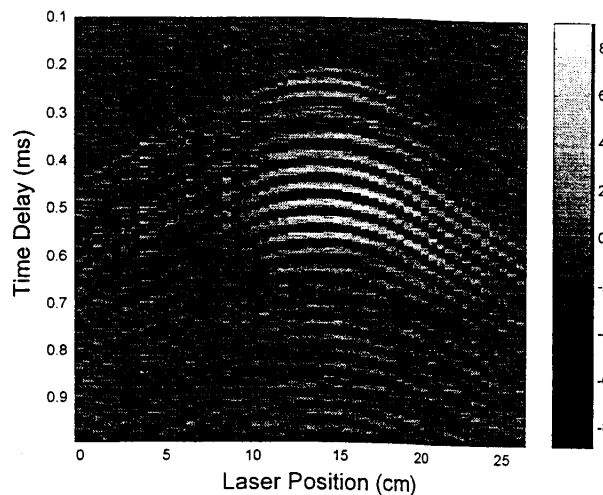


Figure 6. Difference signal for transducer voltage with target present minus voltage with no target present as a function of time delay and laser position x for scan with $y=13.5$ cm (see Figure 3). The grayscale calibration is in microvolts.

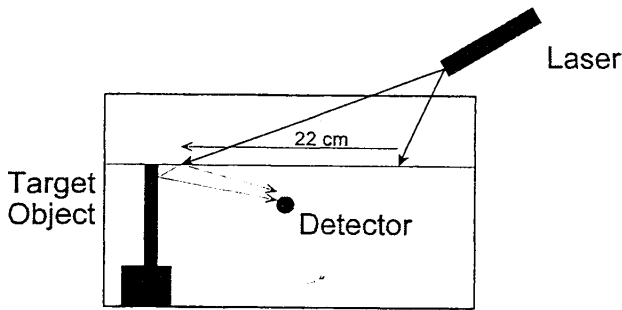


Figure 7. Experimental setup for laser-induced acoustic detection in water tank. As indicated, the time delay between the reflection from the target and the direct reception from the excitation pulse will be expected to decrease as the laser moves closer to the target.

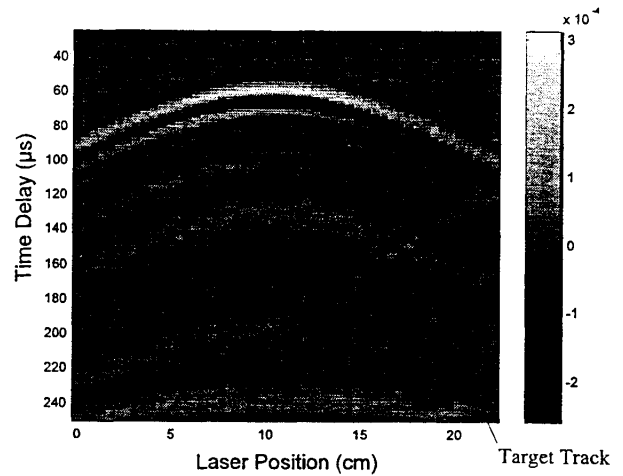


Figure 8. Transducer voltage as function of time delay and laser position for detection of a target in water. The reflection of the target is indicated. The grayscale calibration is in volts.

The target reflection is clearly visible in Figure 8 as the acoustic reflection feature that approaches the direct acoustic reception time delay as the laser excitation approaches the target. From the slope of the target track the propagation velocity can be determined to be consistent with the known speed of sound in water ($v_s = 1500$ m/s). In contrast to the situation in sand or soil, sound propagation in water is only weakly attenuated. The significance of this experiment is that it conclusively demonstrates that laser-induced acoustic signals are transmitted through the target medium (and not just through the air), and that under ideal circumstances reflections from a target can be directly observed.

Acoustic reflections from the enclosure walls have also been observed in experiments in sand. Figure 9 and 10 show the

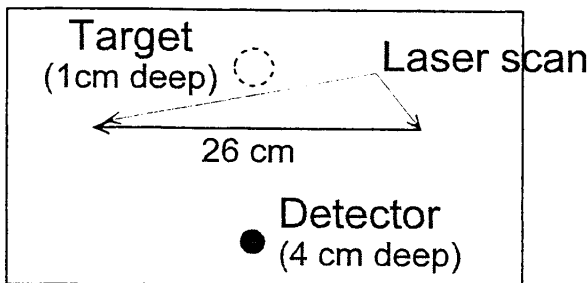


Figure 9. Experimental setup for detection of target in dry sand with acoustic microphone detector.

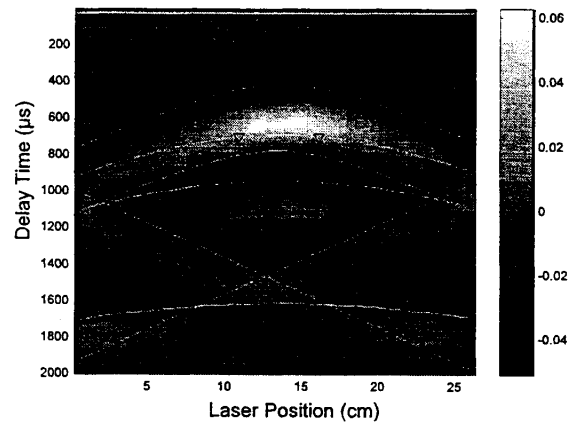


Figure 10. Acoustic microphone signal as function of time delay and laser position. Calculated acoustic arrival times are shown for direct transmission from the laser to the detector (+), for reflection from sides of box (\times , \diamond , \square , \circ), and for reflection from the box bottom (∇). The calculated arrival time for reflections from the target is indicated by the triangles (Δ).

experimental setup and data taken with the acoustic microphone in sand. Reflections from the sides of the box are observed in the diagonal bands in Figure 10. The calculated arrival times for direct transmission from the laser spot to the detector and for reflections from the four sides and bottom of the box are indicated by lines in Figure 10. The calculated lines agree well with the features observed in the data using an acoustic propagation velocity of 260 m/s in dry sand under the soil packing conditions of the experiment, about 75% of the velocity of sound in air. There is no obvious feature at the expected arrival time for reflection from the target, indicated in the figure by the triangles. While there are modifications in the acoustic line shape when the target is removed as shown in Figure 11, it is not easy to identify these subtle changes from the data in the absence of a control data set without a target. What is needed is a way to process the data to extract out the differences in the signal with and without the target present.

One possible processing technique is to use the acoustic signal when the laser spot is far from the target as a control to compare to the signal when the laser spot is near the target. In this way the signal received when the laser is over clear ground could be used as a background to detect the anomaly when the laser is over a buried object. Unfortunately, if the separation between the laser spot and the detector position is changing as the laser is scanned, as in Figure 10, changes in the signal as the laser excitation is over the target are difficult to extract from the changes caused by the spot moving closer to or farther from the detector.

A solution to this problem may be found if the separation between the laser excitation source and the detector can be fixed as the laser spot is scanned. This may be achieved in a practical mine detection system if the detection system as well as the excitation system is laser-based. For example, the CO₂ laser excitation system could be maintained in registration with a laser vibrometer acoustic detector as the two are scanned across the ground.

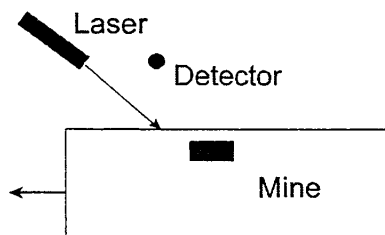


Figure 12. Experimental setup for the acoustic difference experiment. To simulate the registered motion of the laser and detector, the sand box with the mine-simulant target is moved with respect to the laser and detector. The target is buried about 1 cm below the surface of dry sand.

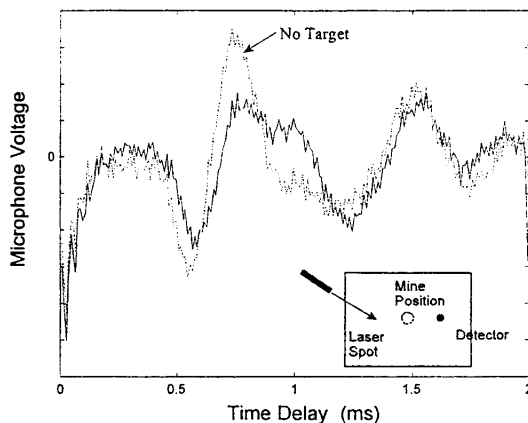


Figure 11. Microphone signal received with and without buried target between laser spot and detector. Experimental configuration is illustrated in inset.

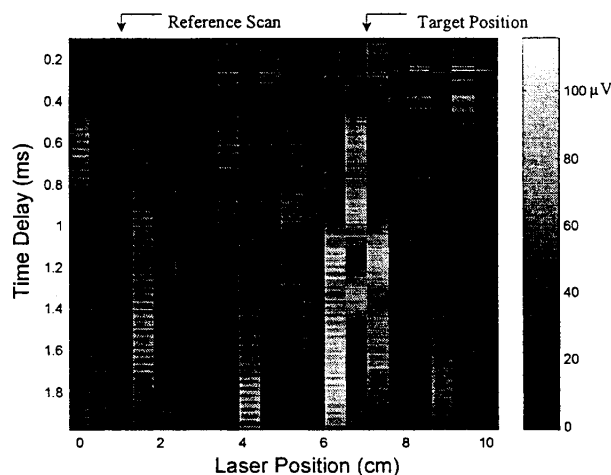


Figure 13. Acoustic anomaly detection as a function of time delay and laser position. The gray scale represents the absolute value of the difference between the time record of the microphone voltage at any laser position and that measured at $x=1$ cm when the laser is far from the target. A clear anomaly is observed near the target position at $x=7$ cm.

As a test of acoustic anomaly detection when the source and detector are maintained in registration, we mounted a microphone in the air over the sand and simulated the coordinated movement of the source and detector by moving the sand container with the detector and laser spot fixed. This experimental configuration is shown in Figure 12. The mine-simulant target was a metal cylinder about 3 cm in diameter and 2 cm high buried with its top about 1 cm under the surface of dry sand. While the acoustic signal will be received in the air above the soil in this configuration, it is reasonable to assume that there will be modification in the sound when the excitation over a near-surface buried object due to changes in the acoustic resonant structure of the soil. In fact, this is the method by which percussion is used to detect objects behind a wall. It is also notable that the configuration in Figure 12 is a non-contact measurement, a considerable advantage in detection of mines.

We processed the data by subtracting the detector signal when the excitation was far from the target from the intensity vs. the time scan at each position. The result of this difference signal is shown in Figure 13, where the data from the third distance bin is subtracted from each of the other data sets and the absolute value of the difference is plotted vs. position and time delay. A clear anomaly is observed at $x=7$ cm when the laser spot is over the position of the mine. Localized anomalies with strong contrast as seen in Figure 12 were not observed in scans taken with no target present. We have not completed experiments to determine if the time delay where the largest difference is observed is correlated to target depth.

3. CONCLUSIONS

Pulsed CO_2 laser-induced acoustic excitation has been shown to be an effective way to get localized acoustic excitation in soil without directly contacting the ground. The acoustic signal detected by a PZT transducer or by an audio microphone has been shown to be sensitive to the presence of near-surface buried objects. While direct target echos have been observed in experiments with laser-induced sound in water, the modification of the acoustic signal by objects under dry sand is complex and not easy to interpret. In experiments with a fixed acoustic receiver in contact with or under the sand, the changes in the acoustic signal when the laser excitation is close to the target position are difficult to distinguish from the changes that result from the change in separation between the scanned laser pulse and the receiver. By maintaining a fixed separation between the laser spot and the acoustic receiver, we have been able to identify acoustic anomalies by taking a difference of the response when the laser is near the target and the response when the laser is far from the target.

These experiments indicate that laser-induced acoustic detection of land mines is a promising area for further study. Our results indicate significant modification of the acoustic signal when the laser is directly over a buried object, so shape discrimination of man-made objects from natural underground objects is possible. Processing of the acoustic signals to identify anomalies from buried objects is complex. Background subtraction has been shown to be feasible, but it is necessary to take account of the changing distance between the scanned laser spot and the microphone. Optical detection by laser vibrometry may be particularly effective in conjunction with laser excitation since the excitation and detection positions can be kept in registration without ground contact or mechanical detector movement.

Additional experiments are required to determine if target depth can be correlated with the time delay associated with the anomaly. In addition, the application of the technique in different types of soils needs to be established. The experiments in dry sand reported here may represent one of the more difficult mine detection scenarios since acoustic propagation in sand is strongly attenuated and depends in a complex way on the packing of the sand grains. Clay soils may allow acoustic detection at greater distances. More sensitive acoustic detectors or accelerometers may improve prospects as well.

4. ACKNOWLEDGEMENTS

This work was supported by the Multidisciplinary University Research Initiative under US Army Research Office Contract number DAAG55-97-1-0013.

5. REFERENCES

1. C. Liu and S. R. Nagel, "Sound in sand," *Physical Review Letters* 68, p. 2301, 1992.
2. W. P. Arnott and J. M. Sabatier, "Laser-doppler vibrometer measurements of acoustic to seismic coupling," *Applied Acoustics* 30, p. 279, 1990.

3. J.-P. Monchalín and R. Héon, "Laser ultrasonic generation and optical detection with a confocal Fabry-Perot interferometer," *Materials Evaluation*, vol. 44, p. 1231, 1986.
4. Monchalín J-P, "Progress towards the application of laser-ultrasonics in industry," *Review of Progress in Quantitative Nondestructive Evaluation*, vol. 12, Edited by D. O. Thompson and D. E. Chimenti (Plenum Publishing Corporation, 1993), p. 495.
5. S. M. Park, M. I. Khan, H. Z Cheng, and G. Diebold, "Photoacoustic Effect in Strongly Absorbing Fluids," *Ultrasonics* 29, 63, 1991.
6. G. Diebold, T. Sun and M. I. Khan, "Photoacoustic Waveforms Generated by Fluid Bodies," in *Photoacoustic and Photothermal Phenomena*, Springer-Verlag, Heidelberg, 1992. (D. Bicanic, ed.).
7. W. R. Scott, "Acousto-electromagnetic sensor for locating land mines," *AeroSense 1998*, SPIE..

Imaging of Buried Objects by Laser-Induced Acoustic Detection

S. W. McKnight*, W. Li, and C. A. DiMarzio

Center for Electromagnetics Research, Northeastern University, Boston, MA 02115

ABSTRACT

We report here on the use of acoustic pulses generated by a pulsed-laser incident on the ground surface for the depth- and shape-resolution of buried objects. The laser-induced acoustic wave has considerable advantages over other acoustic wave generation techniques for landmine detection applications. (1) It is efficient because the sound is generated directly in the ground. (2) The acoustic source can be precisely positioned or scanned by optical redirection of the laser spot. (3) Remote, non-ground-contact detection can be accomplished with a receiving microphone in the air or by using optical vibrometry of the soil surface for detection. Research has been focused on the data acquisition and signal processing applicable to de-mining scenarios. A de-convolution method using a Wiener filter is introduced to the processing of data. By scanning the laser position and filtering the time-trace of the reflected acoustic pulse, we have obtained 3D images of the underground objects. The images give us the clear discrimination of the shapes of underground objects. The quality of the images depends on the mismatch of acoustic impedance of buried objects, the bandwidth of acoustic sensor, and the selection of filter function.

Keywords: Laser-induced acoustic imaging, buried-object detection, landmine detection, photoacoustic, optoacoustic

1. INTRODUCTION

Current de-mining techniques are limited in their ability to resolve sufficient detail to allow the land mine to be identified and differentiated from stones, roots, or stray metal particles. Standard inductive techniques can find low-metal-content mines, but over 300 false alarms must often be investigated for every mine found. These false-alarm rates are unacceptable, even for non-time-critical humanitarian de-mining efforts, due to the enormous areas that need to be cleared as well as the risk of carelessness and inattention caused by a large number of false alarms. A technique that would allow imaging of the shape of a buried object would enable an enormous reduction in the number of underground clutter objects that would need to be investigated. Unfortunately, electromagnetic or acoustic wavelengths that easily penetrate the ground are too long to adequately resolve the shape of scattering objects in the far-field. However, near-field resolution on a sub-wavelength scale is possible with long-wavelength probes in attenuating media if sources can be localized within a few absorption lengths of the object.

Acoustic probes of buried objects are attractive because of the large impedance contrast between solid underground objects and granular soil even for non-metallic targets. Since coupling of sound into and out of the soil is difficult because of the mismatch at the air-ground interface, previous work has used optical¹ or microwave² vibrometry to directly detect the motion of the ground surface. Others have proposed the use of ultrasonic vibrometers to identify landmines by the discrimination of characteristic acoustic non-linear effects.³ We have reported on the use of a short-pulse CO₂ laser to generate acoustic pulses directly in the ground.⁴ Detection was by means of a PZT transducer either buried in the ground or suspended in the air above the laser pulse spot. The acoustic signal was dominated by detector ringing at the 30 kHz resonant frequency, but we were able to detect acoustic anomalies by background subtraction when the laser source was incident on the surface directly over the position of a buried object. In this paper we report on the application of a Wiener filter deconvolution to recover the acoustic pulse echo from our resonant PZT detector and demonstrate that depth resolution can be obtained by time-sampling the filtered signal. Scanning the laser source in registration with the transducer suspended above the ground allows us to map out the shape of a buried object. We demonstrate the application of this technique in determining the shape of a shallow buried non-metallic target, and in differentiating two closely buried objects.

*Correspondence: E-mail: mcknight@neu.edu; Telephone: 617-373-2060; FAX: 617-373-8627

2. EXPERIMENT

The experimental configuration is indicated in Figure 1. The laser source is a Laser Science CO₂ laser with a pulse width of 100 ns and a pulse energy of 150 mJ. Our experiments were carried out with a laser pulse repetition rate of 8 Hz, corresponding to an average power less than 1.2 W. The laser spot is about 1 cm in diameter without focusing.

Our target is a hard rubber hockey puck buried between 0.1 and 2.0 cm under the surface in a box of dry sand. The velocity of sound in hard rubber, about 1600-1800 m/s, creates a good contrast with the velocity in dry sand which we have measured to be 300 m/s, in good agreement with the literature.⁵ The sound velocity in plastic landmine materials are likely to be in the range 1600-2800 m/s.⁶

The laser pulse is strongly absorbed within the first 1 μm of the surface, instantaneously heating the sand about 3°C, and the accompanying pressure wave creates an acoustic pulse. The surface displacement of the sand was measured by a laser vibrometer at about 2 nm, and although we have not measured the energy of the sound pulse, it is clearly audible across a 5 m room.

The detector is a PZT transducer with a resonant frequency of 30 kHz suspended 10 cm over the surface of the sand. Since ambient room noise is small at 30 kHz, the data could be taken without averaging. Data with a lower-frequency ($f \leq 15$ kHz) broad-band microphone were typically averaged over eight laser pulses. The microphone data were not as successful for imaging purposes, either because of the lower signal-to-noise or possibly because of the longer acoustic wavelength. During the experiment the laser spot and the detector are traversed in tandem across the sand box, with the laser spot/detector distance kept as nearly fixed as possible. In some cases the surface of the sand was not level, and that caused the distance to vary as we scanned over the mine location.

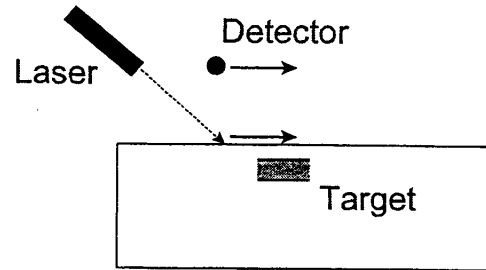


Figure 1. Experimental configuration for laser-induced acoustic experiment with background subtraction or Fourier filtering.

3. RESULTS

In Figure 2 we show a time trace of the detector voltage as a function of time for two single laser pulses—one when the laser and detector are over our target, and one when they are far from the target. The dominant feature in both of these traces is the detector ringing at 30 kHz. While the difference between the traces with and without the target is well above the noise, it is difficult to distinguish characteristics of the target from this data. We have processed this data by using a background subtraction method. In this method, as the laser and detector are scanned across the target we take the acoustic signal as a function of time at each laser position, $y(t, x)$, and subtract the acoustic signal at some reference position where we know there is no mine, $y(t, x_0)$.

An example of the application of this technique is shown in Figure 3. The top panel illustrates on a gray scale the detector signal as a function of time for 25 different positions x as the laser and detector are scanned across the position of a target. The target in this case is a hockey puck buried vertically (with

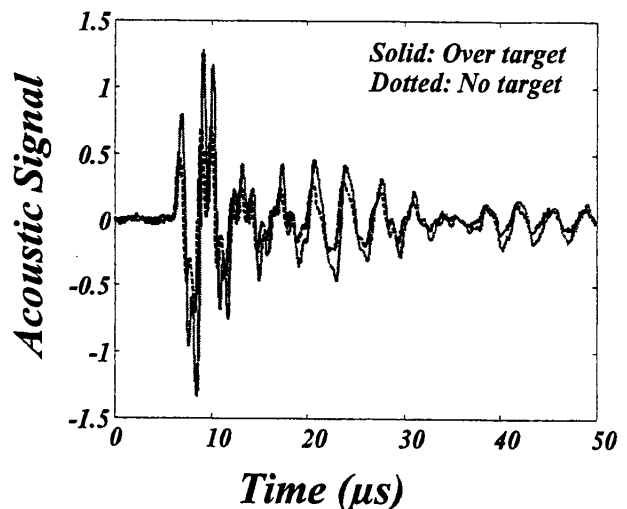


Figure 2. Unprocessed detector signal for a single laser pulse when the laser is over the target (solid) and when it is not over the target (dotted).

its curved side up) with its center at $x=6$ cm. Since the initial acoustic pulse arrives at the detector at the same time for each trace, subtraction of data from the second position ($x=0.5$ cm) from each of the traces reveals the presence of the buried object as illustrated in the bottom panel.

For this technique to be successful, it is necessary that the $t=0$ point be perfectly correlated for each trace. Since the beginning of each time trace is triggered by the laser pulse, this requires the surface of the ground to be perfectly level. If the distance between the laser spot on the ground and the detector changes as the laser beam is scanned, the phase difference between the signals will cause large spurious differences. In addition, as seen in Figure 3 this subtraction does not remove the 28 kHz detector ringing.

To remove the background, including the detector response, in all cases, we have implemented a normalization procedure on the Fourier-transformed data. The justification of this processing is as follows. We assume that the received signal $y(t,x)$ is a convolution of the actual acoustic signal $s(t,x)$ and an unknown time-dependent detector response $h_d(t-t')$, with an added noise term $n(t)$, as shown below in Equation 1.

$$y(t,x) = \int s(t',x) h_d(t-t') dt' + n(t) \quad (1)$$

The Fourier transform of this equation is:

$$Y(\omega,x) = H(\omega)S(\omega,x) + N(\omega) \quad (2)$$

Since we assume that the detector response $H(\omega)$ is independent of position, we can use the data when there is no target $Y(\omega,x_o)$ to get an estimate of $H(\omega)$. Assuming that $S(\omega,x_o)$ is an impulse function, since the laser pulse width is instantaneous on the acoustic timescales, we find that $H(\omega) = Y(\omega,x_o)$. In the absence of noise, we can then recover the real acoustic signal $S(\omega,x)$, assumed to be an impulse plus one or more echo signals, from the unknown data $Y(\omega,x)$ by dividing by $Y(\omega,x_o)$.

This process gives rise to singularities at frequencies where there is little amplitude in the Fourier spectra of the reference signal ($Y(\omega,x_o) \approx 0$). To allow for these cases we add a small real constant ϵ to the Fourier spectrum that we use for a reference. This process is implemented to get an estimation of the acoustic signal $\hat{S}(\omega,x)$ as shown in Equation 3.

$$\hat{S}(\omega,x) = Y(\omega,x) F(\omega) = Y(\omega,x) \left[\frac{Y^*(\omega,x_o)}{|Y(\omega,x_o)|^2 + \epsilon} \right] \quad (3)$$

The filter function $F(\omega)$ defined in Equation 3 is a variation of a Wiener filter⁷. At frequencies where $Y(\omega,x_o)$ is large compared to ϵ , $F(\omega)$ achieves a deconvolution of $Y(\omega,x)$. At frequencies where $Y(\omega,x_o)$ is small compared to ϵ , $F(\omega)$ reduces to a matched filter.

The selection of the constant ϵ and the maximum spectral frequency included in the filtering process, f_{max} , is illustrated in Figure 4. The values in the Figure, $f_{max} = 52.3$ kHz and $\epsilon = 1.75$, were chosen empirically, based on the consideration that almost all of the signal strength is concentrated below 50 kHz. In fact, the results of the filtering process are not drastically affected by the

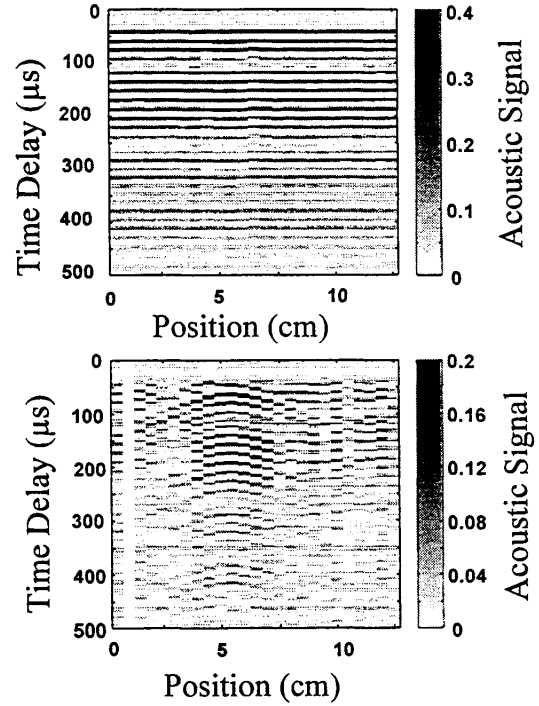


Figure 3. Example of background subtraction technique. Top: A series of time traces of the detector signal a function of time delay after the laser pulse, taken at different positions as the laser/detector are scanned over a target (vertical hockey puck). Bottom: same data with the data in the second horizontal position subtracted from each time trace.

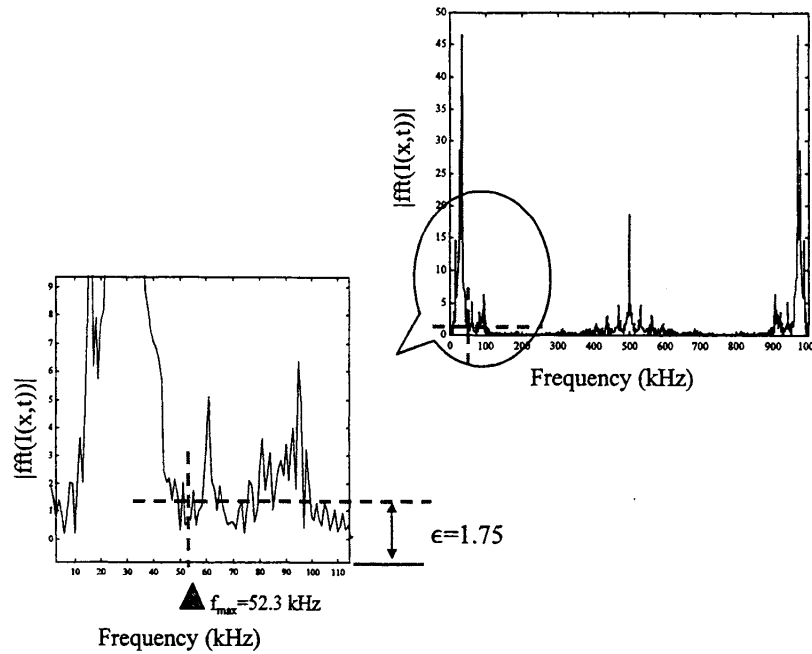


Figure 4. Fourier transform of reference laser-induced acoustic pulse received by PZT transducer, illustrating the selection of the frequency cut-off f_{max} and the constant ϵ .

choice of f_{max} and small variations in ϵ do not have a significant effect. The result of the application of the filtering process in Equation 3 to the data in Figure 3 is shown in Figure 5. The filter has removed the detector ringing and the surface of the puck is clearly indicated by the echo pulses.

The effectiveness of this Fourier filtering technique under conditions where the ground surface is not level is shown in Figure 6 where the top surface of a hockey puck that was deliberately placed under the ground in a tilted position. In this Figure, the strong return from the ground surface is shown as well as the weaker return from the puck surface. The variation in the time delay of the surface return with laser position indicates that the ground surface is uneven in this experiment, so an intensity subtraction as performed in Figure 3 would be impossible. Nevertheless, the Fourier filtered signal effectively images the puck

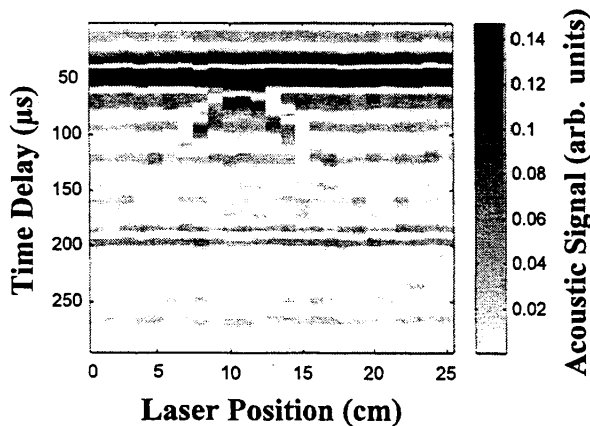


Figure 5. Data of Figure 3 after processing with Fourier filter described in the text. The target is a vertical hockey puck buried 6mm under the surface. The top curved surface is effectively imaged.

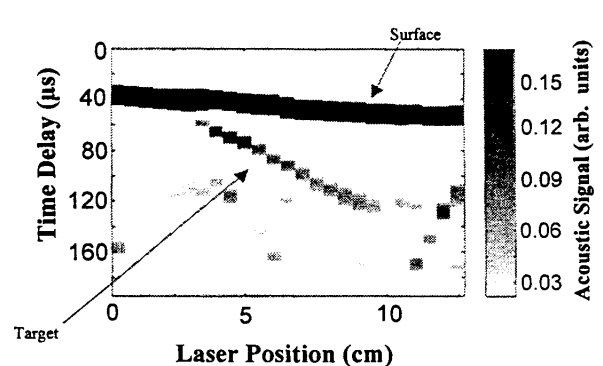


Figure 6. Fourier-filtered acoustic data for tilted hockey puck under uneven ground surface. The high intensity return near 40 μs is the ground surface and the return from the puck surface is visible from 50-130 μs .

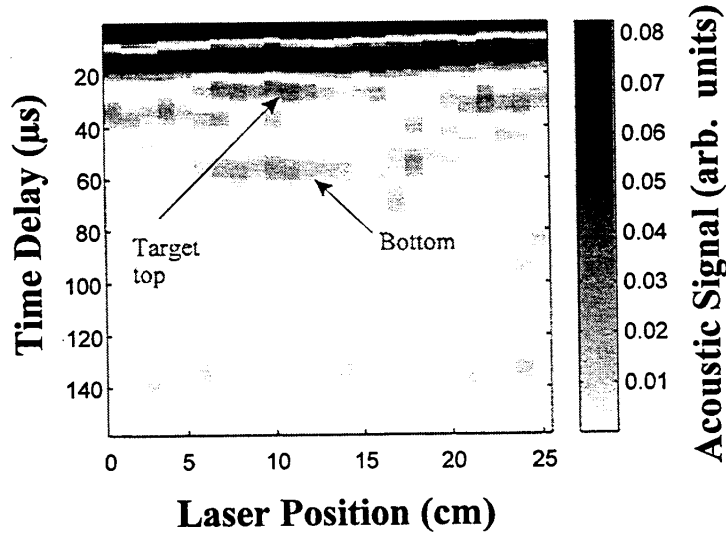


Figure 7. Fourier-filtered acoustic data for hockey puck buried horizontally under surface of dry sand. Echoes from the top and the bottom of the puck are visible.

surface underground. Using a velocity $v_s=300$ m/s that we have measured for the speed of sound in dry sand, we can convert the time difference between the surface return and the return from the puck into a distance under the surface: $d=v_s \Delta t/2$. (The factor of 2 here corresponds to the round trip of the sound from the surface to the puck and then back to the surface.) The data indicate the surface of the puck slopes from about 1mm under the surface to 1.2cm under the surface, in good agreement with the actual puck placement.

Figure 7 shows the Fourier-filtered return from a hockey puck buried horizontally with its top surface 4mm under the surface of the sand. In this Figure, not only is the top surface of the puck well imaged, but the return from the bottom of the puck is also

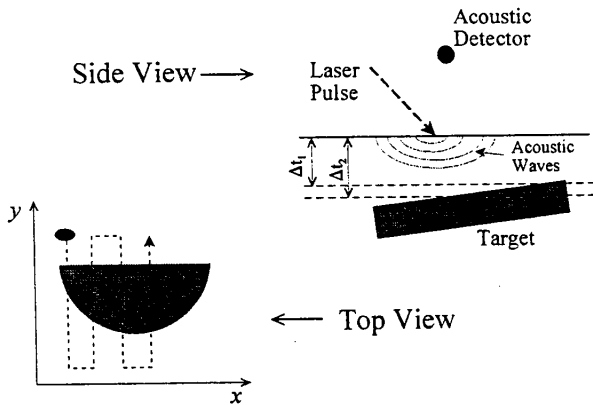


Figure 8. Experimental configuration for laser-induced two-dimensional scanning. The detector and laser are scanned in registration in the x- and y-directions. The time delay Δt identifies the depth of the object.

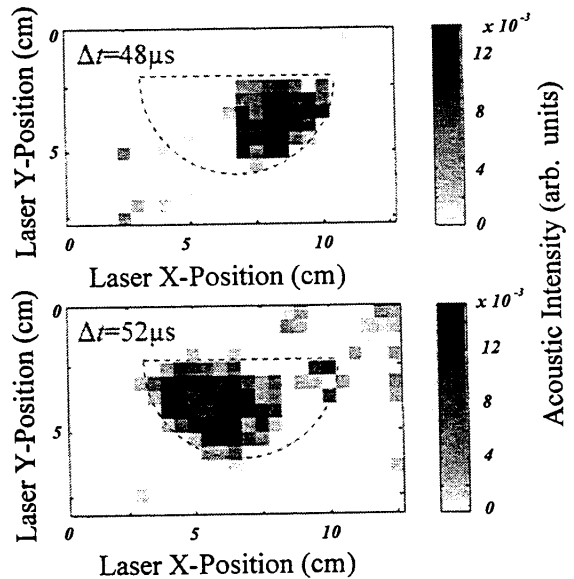


Figure 9. Results of two-dimensional scanning of underground target. The return signal strength is indicated at two different delay times shown, as a function of x- and y-position. The position of the target, a half-hockey puck, is indicated by the dotted lines. The shift of the signal from the left half of the target the right half at longer Δt is due to a slight tilt of the puck.

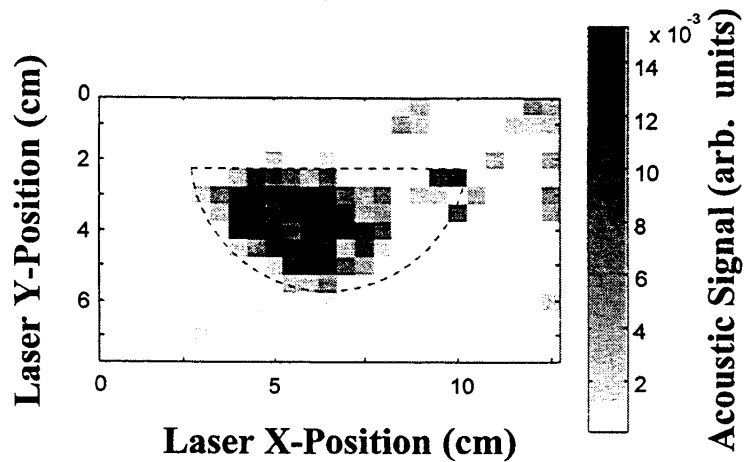


Figure 10. Sum of acoustic return from $\Delta t=45$ to $60\mu s$ along with shape of actual buried object, showing imaging of object shape in horizontal plane.

apparent near a time delay of $60\mu s$. The time delay between the return from the top surface and bottom surface of the 2.5cm thick puck is consistent with a velocity of sound in the hard rubber of 1600 m/s, within the range expected for this material. The ability illustrated in Figure 7 to image both surfaces of an object may be a powerful tool for object identification.

Three-dimensional imaging of object shape can be accomplished by combining a two-dimensional horizontal scan with the information on depth gained by the time delay measurements. This is illustrated in Figure 8. The laser spot is moved across the surface over the position of the underground object in registration with the suspended acoustic receiver. When the spot is over an underground object, an acoustic return will be detected at a time delay characteristic of the depth of the surface of the object. The results of applying this technique to a buried test target consisting of a hockey puck cut in half are shown in Figure 9. The actual position of the target is indicated by the dotted lines on the figure, and the acoustic return at two different delay times, $\Delta t=48$ and $52\mu s$ are shown in the upper and lower part of the figure. The shape of the target is well indicated and the approximately 1 mm tilt of the surface results in the right half of the surface being imaged at the earlier time delay and the left half of the surface selected at the later time delay. By summing up the returns over all time delays, the composite object shape is determined in Figure 10.

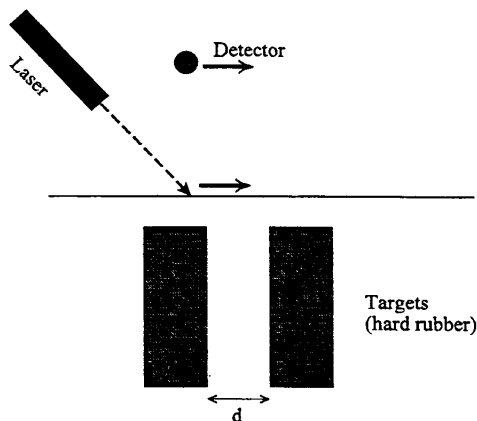


Figure 11. Experimental configuration for determining horizontal resolution of scanning laser-induced acoustic detection. The targets are half-hockey pucks set vertically in sand with flat edge up.

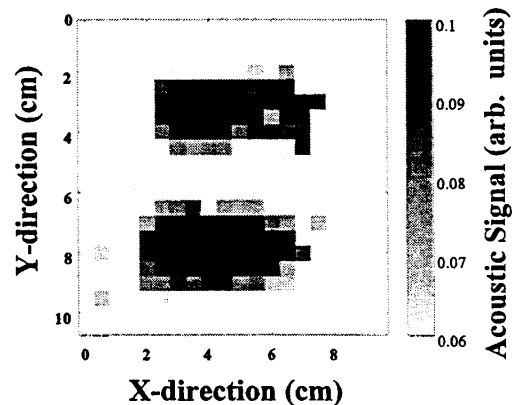


Figure 12. Sum of Fourier-filtered acoustic returns for all time delays vs x- and y-position for two targets as in Figure 11 with a separation $d=2$ cm.

To determine the horizontal resolution limit of the scanning technique, we used as a target two half hockey pucks buried vertically so their horizontal cross sections were rectangular as shown in Figure 11. We then varied the separation between the two pucks to see when we could not distinguish them as two separate objects. An x-y scan across the two half pucks when they were separated by 2cm is shown in Figure 12, and the acoustic signal vs. time delay and horizontal position for a single scan perpendicular to the two pucks is shown in Figure 13. It is clear that the two objects are well resolved and imaged at a separation of $d=2\text{cm}$, and even at a separation of $d=0.5\text{ cm}$ it is apparent that there are two objects. This resolution might be improved by focusing down the size of the laser spot which was about 1 cm in diameter for these experiments.

4. CONCLUSION

We have shown that sound pulses induced by a pulsed CO_2 laser incident on the surface of dry sand and detected with an acoustic transducer suspended over the surface can be used to determine the shape of buried objects. The ringing of the acoustic transducer was removed by a Fourier filtering technique similar to a Wiener filter. The technique is able to image the vertical dimensions of the underground object with a resolution of better than 2mm and the horizontal resolution is better than 0.5 cm. This ability to do three-dimensional acoustic imaging of underground objects has great potential for reducing the false alarm rate in de-mining operations by discriminating target from clutter on the basis of shape. The technique does not require any ground contact and the laser acoustic source can be easily redirected by scanning mirrors. A promising extension of the technique would be to use a laser Doppler vibrometer to measure the vibration of the ground surface in conjunction with the laser source, so that both source and receiver could be remotely located and scanned either together or independently. Experiments along this line are underway.

5. ACKNOWLEDGEMENT

This work was supported by the Army Research Office under the Multidisciplinary University Research Initiative De-Mining Grant DAAG55-97-1-0013

6. REFERENCES

1. W. P. Arnott and J. M. Sabatier, "Laser-doppler vibrometer measurements of acoustic to seismic coupling," *Applied Acoustics* **30**, p. 279, 1990.
2. W. R. Scott, Jr, C. Schroeder, and J. S. Martin, "An acousto-electromagnetic sensor for locating land mines," *Detection and Remediation Technologies for Mines and Minelike Targets III, Proceedings of the SPIE Vol. 3392*, pp. 176-186, 1998.
3. D. M. Donskoy, "Non-linear vibro-acoustic techniques for landmine detection," *Detection of Remediation Technologies for Mines and Minelike Targets III, Proceedings of SPIE Vol. 3392*, pp. 211-217, 1998.
4. S. W. McKnight, C. A. DiMarzio, W. Li, D. O. Hogenboom, G. O. Sauermann, "Laser-induced acoustic detection of buried objects," *Detection of Remediation Technologies for Mines and Minelike Targets III, Proceedings of SPIE Vol. 3392*, pp. 841-847, 1998.

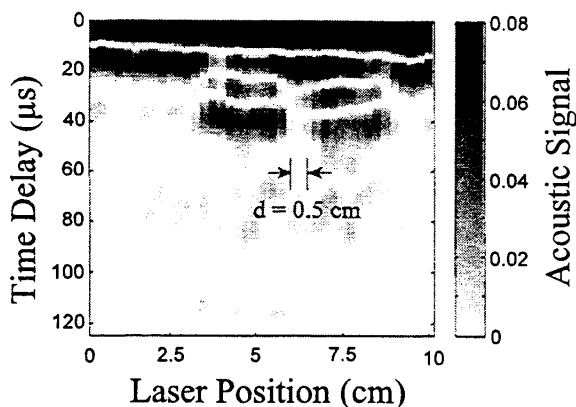


Figure 13. Fourier-filtered acoustic return as a function of time delay and laser position as the laser is scanned across the center of the two targets separated by $d=0.5\text{ cm}$ as shown in Figure 11. The two objects are resolved in the data.

5. C. Liu and S. R. Nagel, "Sound in sand," *Physical Review Letters* **68**, p. 2301, 1992.
6. Value for various plastic and rubber materials are found in the *Handbook of Chemistry and Physics*, The Chemical Rubber Co., Cleveland.
7. N. Wiener, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Wiley, New York, 1949.

Laser-induced acoustic imaging of buried land mines: experiment and modeling

S. W. McKnight^a, J. Stott^a, C. A. DiMarzio^a, R. Cleveland^b, and R. Roy^b

^aECE Department, Northeastern University, Boston, MA 02115

^bAME Department, Boston University, Boston, MA 02215

ABSTRACT

The use for subsurface buried object detection of high-frequency (15-30 kHz) acoustic waves generated by CO₂ laser pulses incident on the surface of dry sand has been demonstrated previously. In this work, field tests of the technique have demonstrated imaging of landmine simulants buried 2.5 cm below the surface in an outdoor test track. Acoustic finite-difference time-domain calculations have given insight into the observed acoustic lineshapes and verified that the over-estimate of the target dimensions in the outdoor field trials may be related to the lower frequency detector used in these measurements. The models also suggest that a large increase in detected signal may potentially be gained by the use of a Laser Doppler Vibrometer interfacial velocity detector in the place of the present airborne microphone.

Keywords: Laser-acoustic, underground imaging, land mines

INTRODUCTION

Acoustic sensing is a promising technique for the detection of low-metal-content buried land mines due to the large impedance mismatch between the porous soil and the solid non-metallic mine case. As in most mine detection strategies, however, the key is reducing false alarms by distinguishing the buried mines from clutter objects (rocks, roots, etc.). Most of the work in acoustic detection has focused on relatively low acoustic frequencies (<1kHz) where resonant oscillations of the soil above the compliant mine cover plate have been observed in both frequency-domain detection of ground motion above the insonified soil¹ and in the time-domain response to surface wave pulses.² In addition, acoustic non-linearities in the mine casing have been exploited to distinguish mines from other underground objects.³

In contrast to these low-frequency techniques, we have reported previously on a technique to probe buried objects with high-frequency ($f > 20$ kHz) acoustic waves generated by directing a pulsed CO₂ laser at the ground surface.⁴⁻⁶ At these frequencies acoustic waves are attenuated within a fraction of a meter in soil, but with the laser source it is possible to create an acoustic wave in close proximity to the buried mine, even at a safe stand-off distance. We have demonstrated that it is possible to detect the echoes from a buried object with a non-contact microphone in the air above the laser source. Since the acoustic wavelength in soil at these frequencies is about a centimeter, it is possible to probe the shape of the buried object by scanning the laser spot. In experiments in a laboratory sand box, we have demonstrated shape resolution in the horizontal plane of 0.5cm, with depth resolution of a few millimeters.⁵ Shape and depth resolution of this order could provide a powerful discriminator for clutter objects, drastically reducing the false-alarm rate.

In this work we report on the application of our technique to image mine simulants in an outdoor field trial. We also report on computational simulations of the laser-induced acoustic effect which have enabled us to identify the source of image features and make predictions of the sensitivity limits of the technique.

EXPERIMENTAL RESULTS

CO₂ radiation is efficiently absorbed in the first few microns of soil. With a sufficiently energetic pulsed source, the rapid heating and thermal expansion caused by the absorption creates a broad-band acoustic pulse. Our laser source is a pulsed laser manufactured by LSI, Inc. with a 100mJ pulse energy, 100ns pulse length, and 20 Hz repetition rate for a modest 5W average

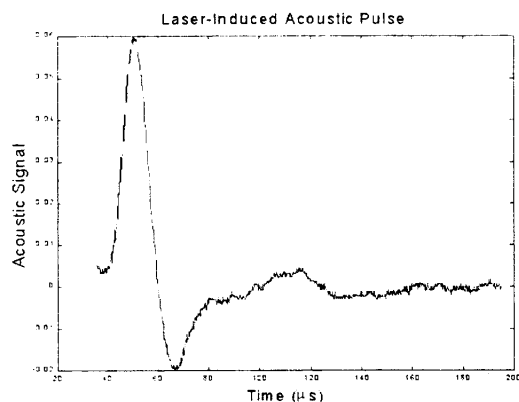


Figure 1. Laser-generated acoustic pulse measured by wideband hydrophone located a few millimeters under the surface of dry sand.

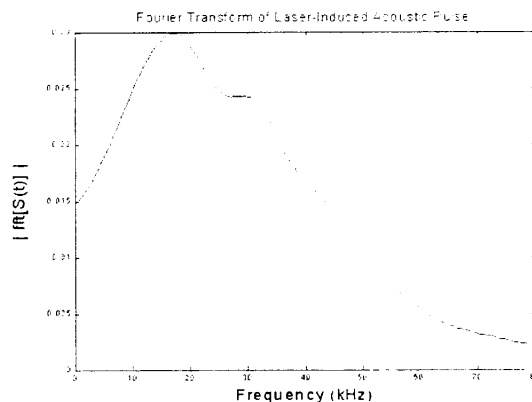


Figure 2. Amplitude of Fourier transform of laser-generated acoustic pulse in Figure 1, indicated bandwidth of laser-generated acoustic source.

power. It creates an acoustic “snap” which is audible at in air a distance of several meters from the approximately 1cm-diameter laser spot. The acoustic pulse as detected with a wide-band B&K, Inc. 100kHz hydrophone 8 mm under the surface of dry sand is shown in Figure 1. The magnitude of the Fourier transform of this pulse, shown in Figure 2, indicates a peak energy intensity between 15 and 25 kHz with significant amplitude from DC to 60 kHz. Note that the data in Figures 1 and 2 have not been corrected for the frequency-dependent attenuation in the propagation path. The wide-band nature of this source contrasts sharply with the low-frequency sources employed in other acoustic detection schemes.

In our previously reported imaging results, we used a narrow-band 30kHz PZT detector suspended in the air over the laser spot to detect the sound reflected from buried objects. This detector had the advantage of removing the omnipresent low-frequency acoustic noise from our signal, but required a Fourier-domain filtering technique to deconvolve the echo pulse from the detector ringing. It also exploited the bandwidth of our acoustic source and permitted high-resolution shape- and depth-imaging of buried non-metallic objects.

To test this technique under more realistic field conditions, we took our laser source to the Northeastern University Dedham test track facility, where nine mine simulants are buried along with several clutter objects in a 2.3m x 21m track filled with screened loam and spanned by a concrete and metal track that carries an instrument cart. A photograph of the laser source on the instrument cart is shown in Figure 3. The acoustic detector for this experiment was an Radio Shack 33-1073A unidirectional acoustic microphone with a nearly flat audio response below 15kHz. This relatively low frequency detector proved to be overly sensitive to low-frequency noise and degraded the resolution of the detector as shown below.

The target we concentrated on was an M-14 antipersonnel land mine simulant with the charge removed and replaced with silicone filler, buried with its top plate 2.5cm below ground surface. This mine is small in size (56mm diameter x

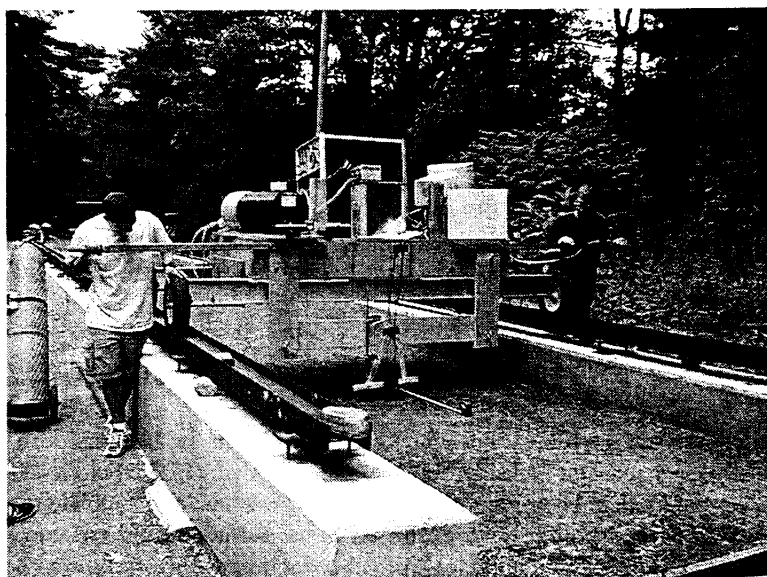


Figure 3. Field test of laser acoustic detection at the Northeastern University Dedham test track. The laser, power supply, and data acquisition computer are mounted on the cart, and the microphone detector is suspended just above the ground over the focused laser spot.

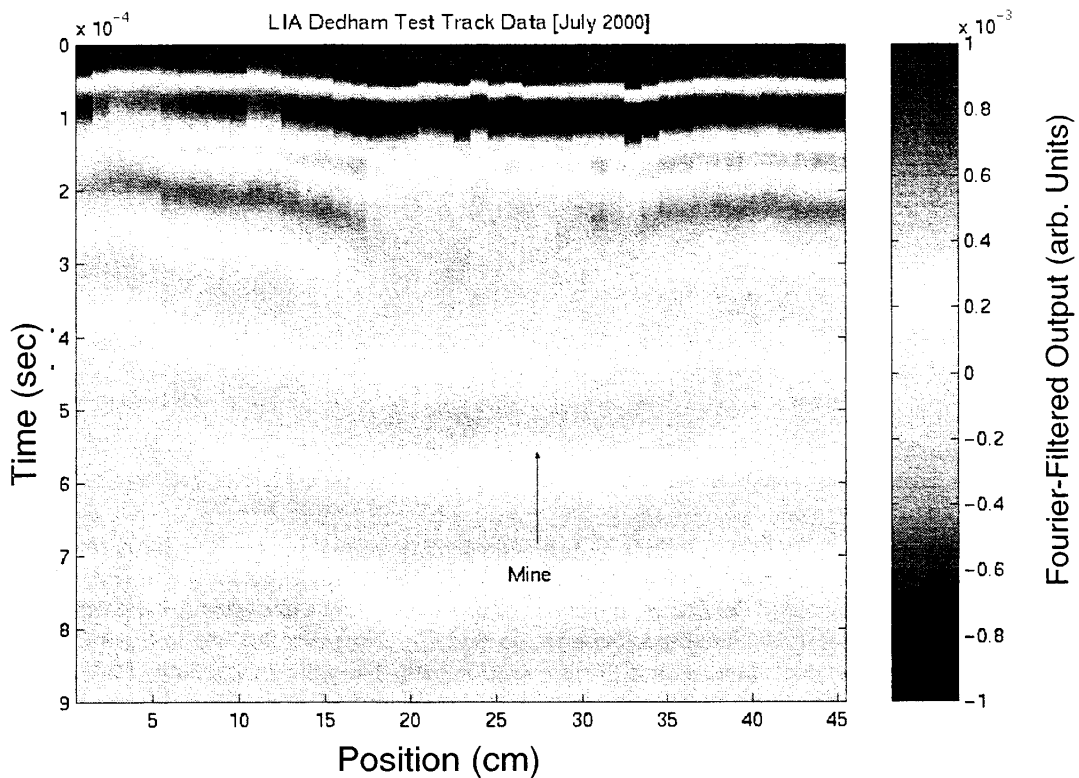


Figure 4. Processed acoustic signal as a function of position as the laser source is moved across a buried M-14 land mine simulant. The detector was an acoustic microphone with bandwidth 0-15kHz and the mine simulant was located near the 25 cm point.

40mm high) and has almost no metallic content. It is considered very difficult to detect by convention means. We moved the laser cart over the centerline of the mine, stopping to take a series of eight shots every 1.2 cm. The laser was focused to a spot about 0.5cm in diameter and the microphone was suspended about 4cm over the laser spot. We processed the microphone data first by normalizing the peak amplitude of the signal to correct for different acoustic intensity created when the laser pulse struck different parts of the soil surface (embedded pebbles, etc.). Then we applied the Weiner Fourier filter as described in Reference 6. The data is shown in Figure 4. The Weiner filter was not as successful in removing detector characteristics as in our experiments in the lab with the PZT detector, and the uncanceled detector ringing is the dominant feature of the processed data. Nevertheless, there is a clear anomaly at the position of the M-14. The size of the anomalous region exceeds the 5.6cm diameter of the mine by a factor of two, but is still considerably less than the region of soil disturbed during mine emplacement (~20 cm diameter). We will demonstrate in the next section that the most likely cause of this is the reduced resolution caused by the lower frequency detector.

ACOUSTIC MODELING

To understand the line shapes and acoustic effects in the laser-induced acoustic detection technique, we have applied a two-dimensional acoustic Finite-Difference-Time-Domain (FDTD) computation to model the problem. The soil was modeled as a linear and lossless effective media—no attempt was made to simulate the porous nature of the medium. The laser source was assumed to be a point source excitation at the surface of the soil and we did not model the surface roughness or the noise environment. The calculation was second-order in space and time on a two-dimensional Cartesian grid with a march in time. The density and sound speed were taken from empirical data, and an absorbing boundary was placed at the edge of the domain. The parameters for the calculation are indicated in Figure 5, which also indicates the alternative positions of the acoustic point source and three modeled detectors: a microphone in the air above the source, a hydrophone in the soil under the pulse, and a surface motion sensor such as a laser Doppler vibrometer. The goals of this calculation were: 1) to understand the time-domain acoustic

Numerical Set-Up

Source: 1 cycle, 100 kHz, Wavelength = 2.55 mm (sand); 3.4 mm (air)

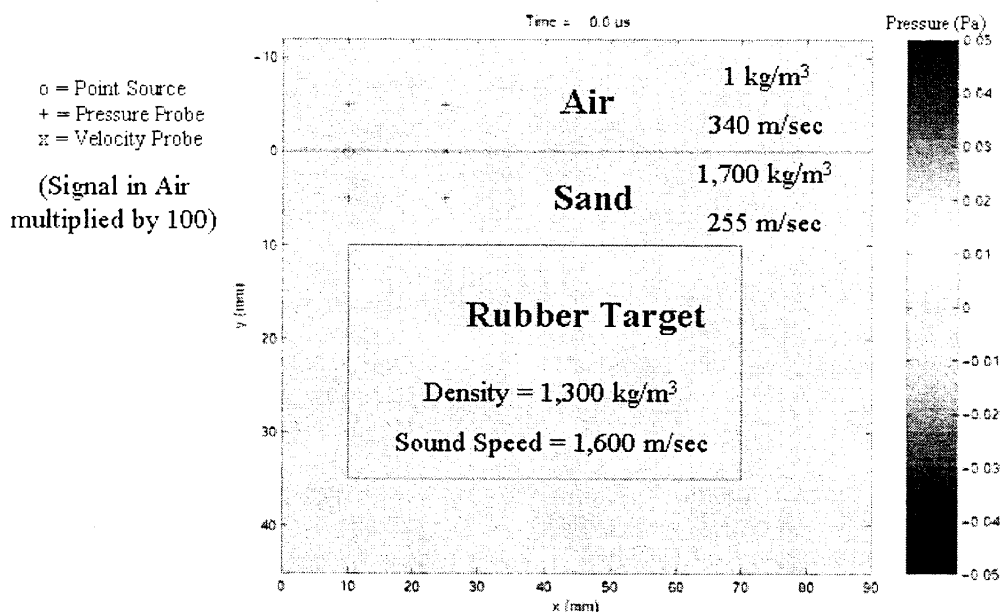


Figure 5. Parameters for acoustic modeling of laboratory experiment in dry sand. The position of the laser source, airborne microphone, buried hydrophone, and an interfacial velocity sensor (LDV) are indicated.

waveform we measure in the experiment, 2) to compare the effects of different detection bandwidths, and 3) to predict relative detection performance from different detector types, airborne or buried pressure sensors and interfacial velocity sensors. The sensitivities of the three receivers were selected to mimic the following commercial sensors: a Radio Shack Model 330-1052 Electret Condenser Microphone, with a nominal sensitivity = 2.5 nV/ Pa (airborne microphone), a B&K 8103 Hydrophone with a nominal sensitivity = 12.6 pV/ Pa (buried sensor), and a Polytec Model OVD-02 Laser Doppler Vibrometer with a nominal sensitivity = 0.2 V/(mm/sec) (interfacial velocity sensor). The frequency dependencies of the detectors are not considered in the calculation, and they were all taken as point receivers.

With the target modeled as a 25mm x 60mm hard rubber disk to simulate the laboratory hockey puck targets, we calculated the response at the airborne sensor as the acoustic source was moved from a horizontal offset of -20mm from the target edge to +5mm from the target edge (over the target). The airborne receiver is maintained at a constant offset of +15mm from the source. (The target edge is taken as the zero point of the distance scale.) The data for several positions of the source and receiver as they approach and overlap the target position are shown in Figure 6.

The direct reception from the source to the receiver is the first feature in each trace. This is reduced from the signal in the ground by a factor of over 100 due to the impedance mismatch at the air-ground interface. The small feature following the direct pulse that appears in every trace is an artifact due to incompletely absorbing boundary conditions. When the source is -10mm away from the target edge (and the receiver is at +5mm over the target), a second feature appears in the traces at a later time. This feature grows in size as the source approaches closer to the target. By observing the time evolution of the pressure waves in two-dimensional movies, the source of this second feature becomes apparent: it is a second bounce from the top surface of the target by the acoustic pulse which reflects from the top of the target, then off the air-ground interface, reflects off the target again, and transmits across the interface to the receiver. While this signal will certainly be reduced from the model calculation by the actual acoustic attenuation of the soil (which is not modeled), we have observed this second signal in our data. Figure 7 shows previously

Microphone Response in Air (Probe-3)

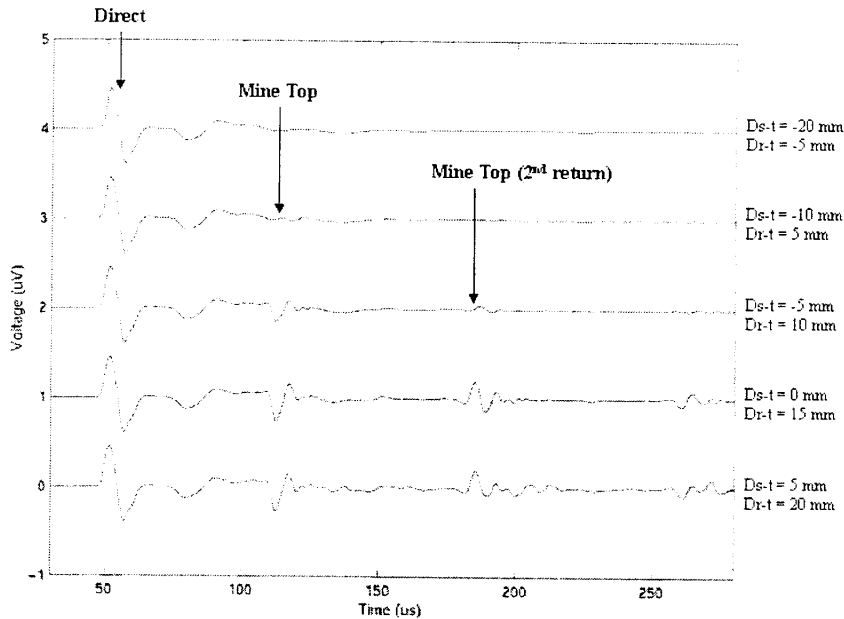


Figure 6. Predicted signal from the acoustic model as measured at an airborne microphone when the laser source is moved closer to the edge of the mine simulant. D_{s-t} is the horizontal distance from the laser source to the edge of the target. D_{r-t} is the horizontal distance from the microphone receiver to the edge of the target.

published data showing the time signal as a function of position as the laser is moved over a rubber hockey puck buried 4mm below the surface. The direct signal between 0 and 20 μ s, the return from the top of the target near 30 μ s, and the second bounce at 60 μ s (previously misidentified as a reflection from the bottom of the puck) are all clearly visible.

A second question that can be addressed by the model calculations is to what extent the apparent increased size of the target in Figure 4 can be attributed to the lower resolution caused by the lower frequency of the microphone detector (<15kHz as opposed to ~30kHz for the PZT detector used in Figure 7). We modeled this by observing the response at an airborne detector as the laser point-source is moved across the top of a 56mm dia x 40 mm target buried 25mm below the surface, as in the acoustic field tests. The positions of the source and receiver with respect to the target are as in Figure 6. The acoustic source is taken to be a single cycle at 30kHz, a signal which has significant intensity across the frequency spectrum from DC to above 50 kHz. The signal received at the detector position is Fourier transformed and the magnitude of the Fourier component at three frequencies, 7, 14, and 28kHz are compared with the actual target extent in Figure 8. The decrease in target resolution is apparent in the figure, and the signal extent at 7kHz, in the midband of the microphone used in the experiment, is approximately twice the true target extent. This suggests that bandwidth considerations may well explain the larger extent of the image in the experiment of Figure 4, and emphasizes the resolution advantage of utilizing the entire bandwidth of the laser-acoustic source.

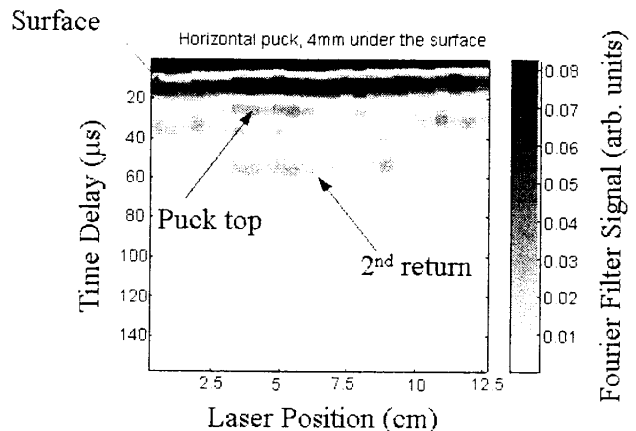


Figure 7. Laboratory data image of target (hockey puck) buried under 4mm of dry sand, showing the second return predicted by the acoustic model in Figure 6.

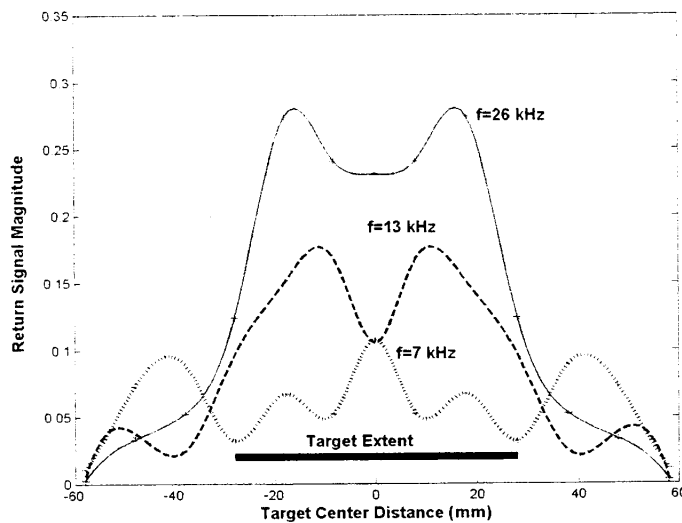


Figure 8. Predicted acoustic signal vs. laser position compared with actual target extent for different receiver frequencies, showing effect of detector frequency on target resolution.

A major loss mechanism for laser-acoustic detection is the impedance mismatch at the air-ground interface which causes a reflection of about 99% of the signal reflected off the target. While the laser source produces acoustic signal directly in the ground, thereby bypassing this loss in transmission, the received signal has to be transmitted to the airborne detector above. The solution suggests itself to use a Laser Doppler Vibrometer (LDV) to detect the motion of the free surface of the ground instead

Comparison of Microphone, Hydrophone & LDV Signals

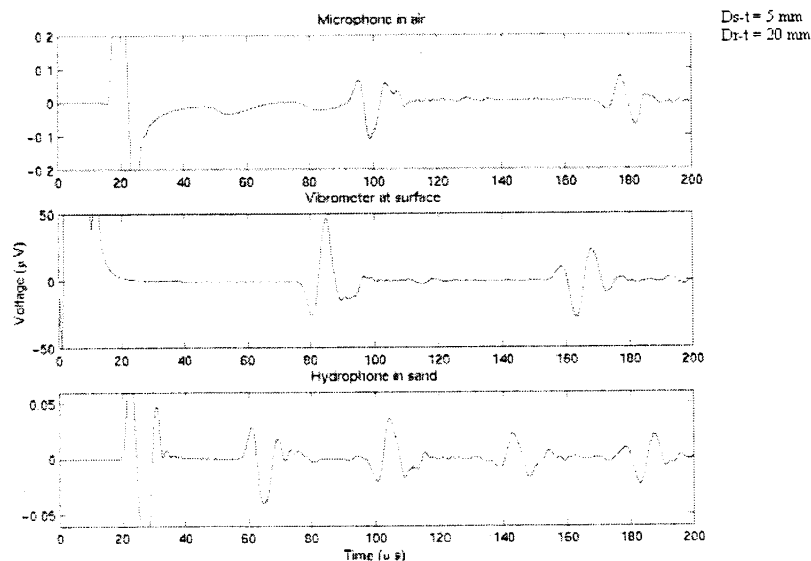


Figure 9. Predicted signal received at detector positions as shown in Figure 5 for airborne microphone sensor (top), buried, hydrophone (bottom), and LDV interfacial velocity sensor (middle). Note that the voltage scale for the LDV is approximately 100 times that for the microphone. (The extra returns in the hydrophone trace result from the detection of both the outgoing and returning wave at the buried sensor location.)

of the weak transmitted pressure signal. This configuration could not only result in increased sensitivity, but would permit look-ahead, stand-off detection which is very desirable from a counter-mine operations point of view.

Using the configuration of Figure 5, and the detector sensitivities as given above, we modeled the received detector voltages as a function of time when the laser excitation was directly over the target. No attempt was made to estimate the different noise floors of the detectors. Nevertheless, as shown in Figure 9, the predicted detector voltage output from the LDV is a factor of 100 greater than that predicted at either the airborne microphone or the buried hydrophone. The potential signal-to-noise gain of this detector configuration would be an enormous step toward making laser-acoustic detection a reality. Tests of LDV detection with laser-acoustic excitation are underway to confirm the advantage suggested by the model.

CONCLUSION

Laser-acoustic imaging of underground objects has been shown in the laboratory to allow detailed imaging of the shapes of buried objects. The shape and depth resolution could contribute to a drastic reduction in the false-alarm rate for underground clutter. We have demonstrated the practicality of the technique in tests using realistic antipersonnel mine simulants at an outdoor test track. The images obtained with laser-acoustic excitation and a low-frequency (<15kHz) microphone detector indicate anomalies associated with the position of the mine simulant, but the size and depth of the simulant are not as clearly defined as in the laboratory measurements in dry sand with a higher frequency (30 kHz) detector.

Acoustic modeling has been performed using a two-dimensional FDTD code, assuming linear and lossless propagation in a single-phase effective media. The results of the simulation accurately predict lineshape features observed in the data, and indicate that the excessive size of the image in the field trials at the test track may be a result of the lower frequency of the detector, compared with the detector used in the laboratory experiments. The acoustic model predicts a factor of 100 gain in the signal voltage may be obtained by using a LDV interfacial velocity sensor to avoid the acoustic loss at the air-ground interface.

ACKNOWLEDGMENTS

This work has been supported by the Army Research Office under the Multidisciplinary University Research Initiative Grant DAAG55-97-1-0013. Other support from the Engineering Research Centers Program of the National Science Foundation under award number EEC-9986821 is gratefully acknowledged. Computational resources were provided by the Scientific Computing and Visualisation group at Boston University.

REFERENCES

1. J. M. Sabatier and N. Xiang, "Laser-Doppler-based acoustic-to-seismic detection of buried mines," *Proc. of the SPIE, Detection and Remediation Technologies for Mines and Minelike Targets IV*, Vol. 3710, 215 (April 1999).
2. C. T. Schroder and W. R. Scott, Jr., *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 38, 1505 (2000).
3. D. M. Donskoy, "Detection and discrimination of nonmetallic land mines," *Proc. of the SPIE, Detection and Remediation Technologies for Mines and Minelike Targets IV*, Vol. 3710, 215 (April 1999).
4. S. W. McKnight, W. Li, and C. DiMarzio, "Imaging of buried objects by laser-induced acoustic detection," *Proc. of the SPIE, Detection and Remediation Technologies for Mines and Minelike Targets IV*, Vol. 3710, 231 (April 1999).
5. S. W. McKnight, C. A. DiMarzio, W. Li, R. A. Roy, "Laser-induced acoustic generation for buried object detection," *Proc. of the SPIE, Detection and Remediation Technologies for Mines and Minelike Targets V*, Vol. 4038, 734 (April 2000).
6. S. W. McKnight, C. A. DiMarzio, W. Li, and J. Stott, "Laser-induced acoustic imaging of buried objects," *Journal of Subsurface Sensing Technologies and Applications*, Vol. 2, No. 2, pp. 119-126, April 2001.

Distinguishing shape details of buried non-metallic mine-like objects with GPR

Carey M. Rappaport, Shuang Wu, Misha Kilmer, and Eric Miller

Center for Electromagnetics Research, Northeastern University
235 Forsyth Building, Boston, MA 02115

Abstract

The finite difference frequency domain is used to study the scattering of buried non-metallic mine-like targets to determine the feasibility of identifying mines from shape features. It is shown that for constant cross-sectional target area -- approximately 100 cm^2 -- the scattered fields of targets with roughly the same height-to-width aspect ratio at 500 MHz are virtually indistinguishable regardless of burial depth. A comparison of the field obtained for mine-like targets of different aspect ratios, but with constant area, buried at a depth of 5 cm, shows marked differences, as does scattered field for GPR frequencies above 700 MHz.

The conclusion of this study is that while low GPR sensing frequencies may help to detect shallow anomalies, they do not supply any useful information about the shape details -- particularly the edges -- of the buried non-metallic mine-like targets.

Keywords:

GPR, mine detection, dielectric target imaging.

1. Introduction and Computational Electromagnetic Background

For many underground detection problems, it is important to know what target features are characteristic and most easily detectable. For metallic targets in free space, edges provide strong scattering centers and are often used for scattering signature analysis. For buried non-metallic objects, however, edges do not provide as strong a scattered signal.

To determine the importance of edges -- and shape in general -- as discriminating features, the scattered signal from two-dimensional objects buried in a lossy, dispersive soil medium can be simulated using finite difference computational methods. Both the finite difference time domain (FDTD) and the finite difference frequency domain (FDFD) algorithms discretize the scatterer and its surrounding space, and hence can model any target shape, soil inhomogeneities, and rough surface boundaries [1,2].

The FDFD method is preferable to the time domain algorithm in the present analysis, because it computes the scattering response at a single frequency, giving the field distribution throughout the problem space. Since knowledge of the frequency dependence of the scattered signal is an important sensing parameter, and since there is no need to iterate until transients have died out, FDFD is the method of choice. An additional advantage of FDFD is that there is no need to rely on special methods to handle frequency-dependent soil media in the time domain [3-5].

Contrary to common perception, the FDFD method is relatively fast. For a 100 by 100 point grid, terminated on each side by a Perfectly Matched Layer (PML) absorbing boundary condition [6,7], the entire evaluation on a 233MHz Pentium Pro running the Matlab 5.0 sparse matrix solver is about one minute. Using a preconditioned QMR iterative solution method [8], the CPU time for large two dimensional grids of N unknown field values grows as $N \ln N$.

One important requirement for efficient FDFD code is the PML material absorber ABC. Since the PML is composed of just layers of propagation media with particular values of electric and magnetic conductivity, the sparse, symmetric structure of the simultaneous equation matrix is unaffected by the ABC.

Care must be used to tune the PML to the air/soil interface. With the usual PML boundary, it is assumed that this ABC terminates a region of free space. The electric and corresponding magnetic conductivity of the PML sub-layers build up from zero to the maximum value at its termination. For the PML termination to a uniform conductive scattering space, the conductivity of the first sublayer must be slightly greater than that of the scattering space. Subsequent sublayer permittivity and conductivity increase according to the anisotropic space mapping principle[9], with corresponding magnetic conductivity increasing to keep the impedance of each PML sublayer constant.

To compute the scattered fields due to plane wave incidence on a target in a lossy half-space, the incident field in the half-space is first determined analytically, using the standard transmission coefficient, and then this field across the support of the target is used as the excitation. That is, the source-free Helmholtz equation for total field:

$$[\nabla^2 + k^2(x,y)] E_z^{tot} = 0 \quad (1)$$

is rewritten in terms of material variations from the constant soil background, $k(x,y) = k_0 + k_{target}$, and incident (in the soil background) and scattered electric field, $E_z^{tot} = E_z^{inc} + E_z^{scat}$. Since E_z^{inc} solves the Helmholtz equation with background k_0 , Equation (1) becomes:

$$\begin{aligned} [\nabla^2 + k^2(x,y)] E_z^{scat} &= - [\nabla^2 + k^2(x,y)] E_z^{inc} \quad (2) \\ &= - k_{target}^2 E_z^{inc} \end{aligned}$$

Note that this formulation is exact, unlike the Born approximation. Approximation is unnecessary since FDFD calculates field values across the grid for all types of materials in any shape or form.

2. Target Shape Scattering Characteristics

The effects of scattering by corners of a dielectric target, typical of buried mine-like objects, can be studied by examining the scattered fields throughout space, and in particular above the surface of a lossy dispersive soil model half-space. Figure 1 shows the geometries of two representative scattering cases. In both cases a target with material characteristics of TNT (dielectric constant, $\epsilon' = 2.9$, loss tangent $\tan \delta = 0.001$) is embedded in a half space of material with electrical characteristics of dry sand ($\epsilon' = 2.5$, loss tangent $\tan \delta = 0.01$) [10]. The targets have the same cross-sectional area of 100 cm^2 , and are buried at a nominal depth of 5 cm. In the first case the target is circular, obviously with no corners, while the second target case has a square cross-section.

Geometry of buried non-metallic mines

Dry Sand: $\epsilon_s = 2.5$, $\tan \delta_s = 0.01$

TNT mine: $\epsilon_m = 2.9$, $\tan \delta_m = 0.001$

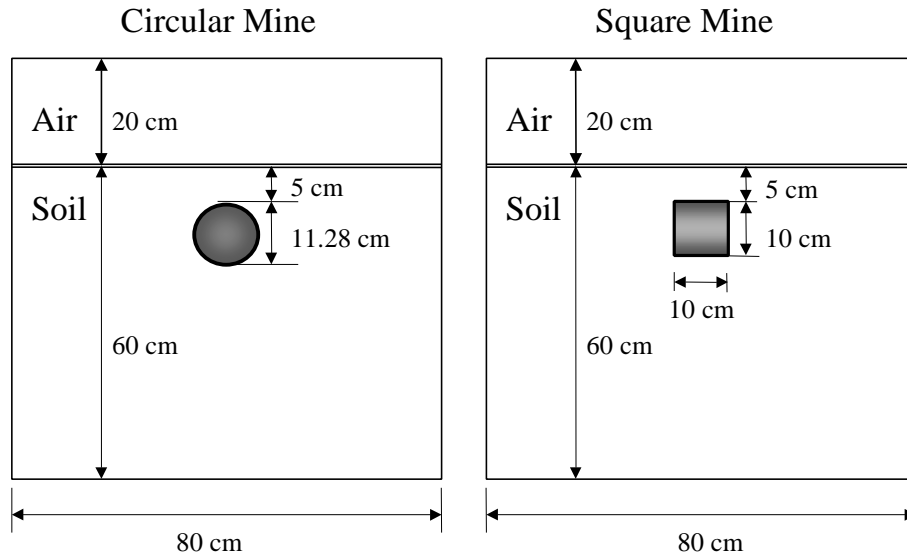


Figure 1: Nominal geometry of buried targets, 5 cm below ground surface. Soil assumed to be dry sand, targets modeled with electrical characteristics of TNT.

A normally incident time harmonic plane wave excites each target case. The scattered fields, as determined by the FDFD discretization of Equation (2), are presented in the top row of Figure 2. The excitation frequency is 500MHz, corresponding to a wavelength of about 40 cm in dry sand. The targets are roughly one-quarter of one wavelength in extent, so it is not surprising that the resolving power of the scattered field is limited.

Scattered Field - Aspect Ratio Dependence

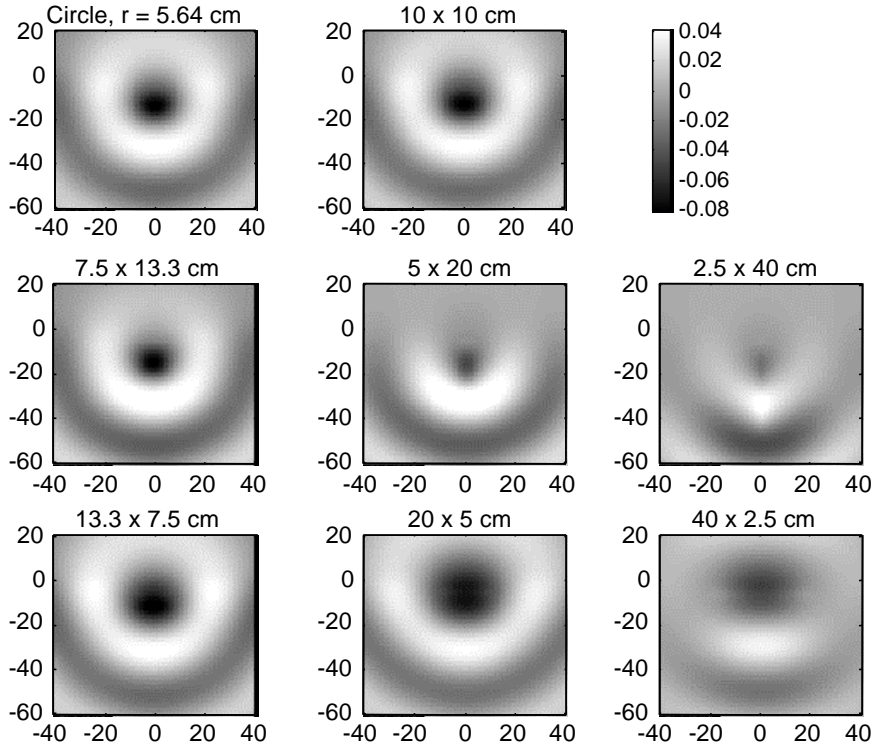


Figure 2: Real parts of scattered fields from mine-like targets with circular (top left), square (top right), and rectangular cross-sections: progressively taller (middle row), and progressively wider (bottom row). All shapes have the same cross-section, 100 cm^2 .

Surface Field - Cross Section Area Dependence

Dry Sand, $\epsilon = 2.5$, $\tan\delta = 0.01$
frequency = 500 MHz, $d = 5 \text{ cm}$

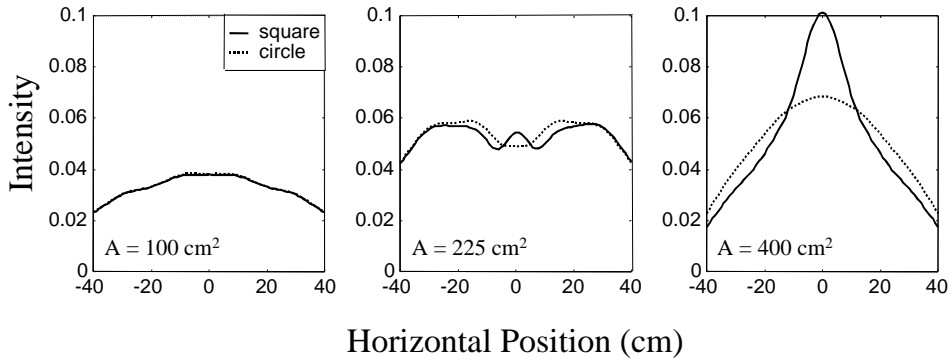


Figure 3: Magnitude of scattered field on the ground surface for circular and square target shapes, each with 100 (left), 225 (middle), and 400 cm^2 (right) cross-sectional area.

While the corners do not appear to be a detectable scattering feature at 500 MHz, the width-to-height aspect ratio does affect the scattered fields. The second row in Figure 2 shows the scattered fields from rectangles that are taller than wide with aspect ratios: 9/16, 1/4, and 1/16; while the third row shows wider rectangles, with aspect ratios: 16/9, 4, and 16. Despite all targets having the same area, the taller rectangles have an upward shattered field with roughly semi-circular contours with greater curvature than that of wider targets.

Figure 3 shows the dependence of scattered field for circular and square targets of various sizes. As expected, both the magnitude of the scattered signal and the differences between target shapes become more apparent with increasing size. In addition, the horizontal variation of scattered field is greater for the larger targets.

Several additional target shapes are presented in Figure 4, including a square tilted 45 degrees, an asterisk, and an asymmetric, irregular shape. The real part of the scattered fields for these shapes are given in Figure 5 for 500 MHz incident field. Note again that it is very difficult to ascertain any shape information from these scattered fields.

Different Buried Test Target Shapes

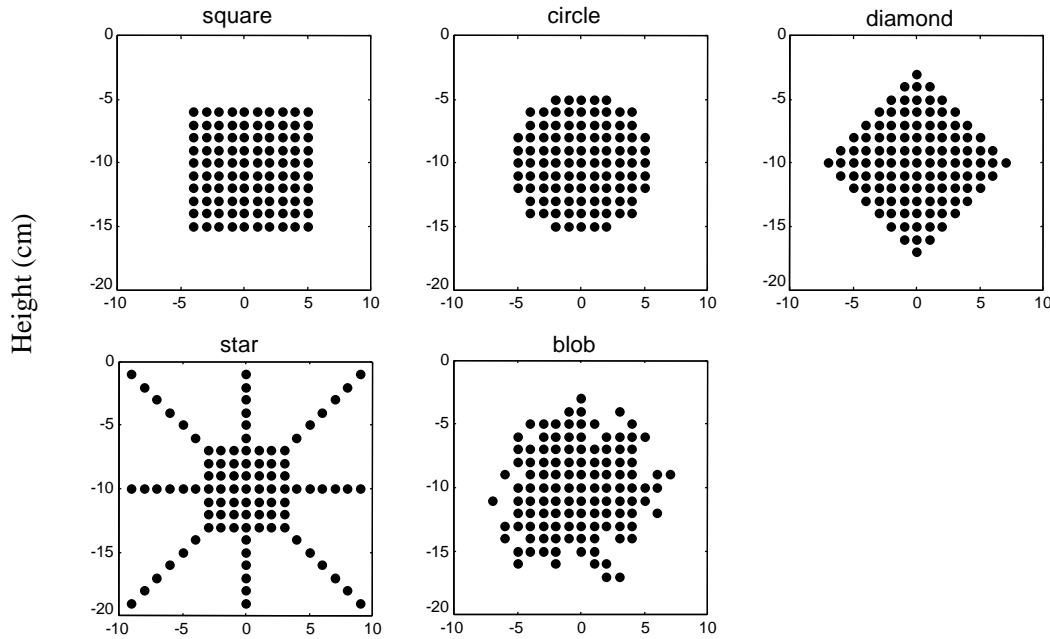


Figure 4: Various target shapes, each with 100 cm² cross-sectional area.

Figure 6 shows the scattered fields for the same target shapes of Figure 4, but for 1 GHz incident field. At this higher frequency, the differences in the scattered fields are visible. In particular, the scattered signal due to the finned structure in the lower right is weaker under ground, and that of the diamond target in the upper right is weaker above the ground.

Scattered Field - Real Part

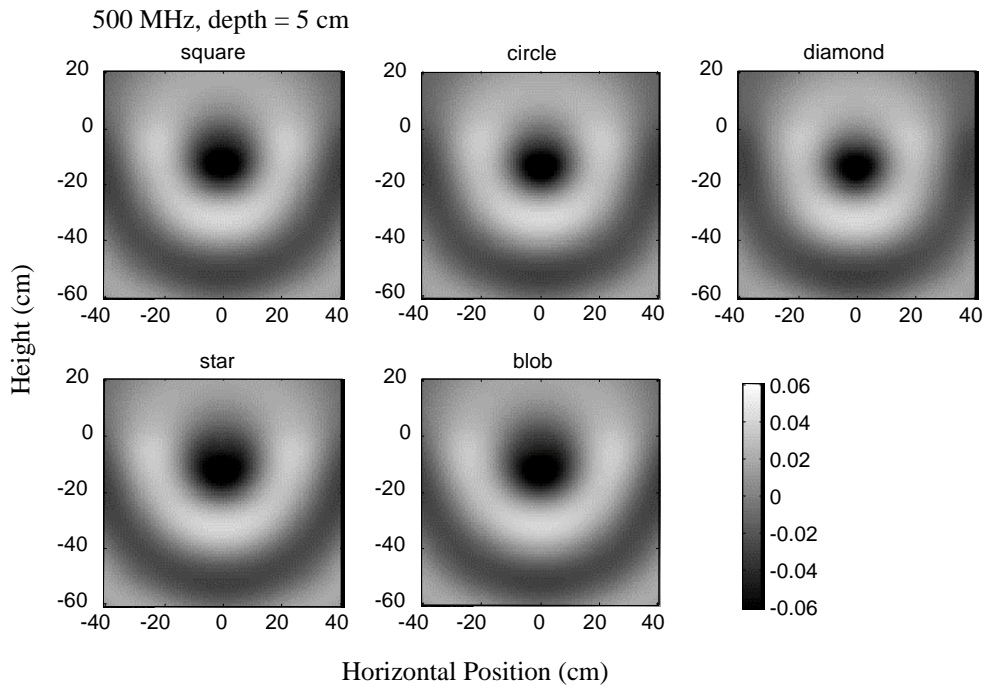


Figure 5: Real parts of fields scattered from the target shapes of Fig. 4 for 500 MHz excitation.

Scattered Field - Real Part

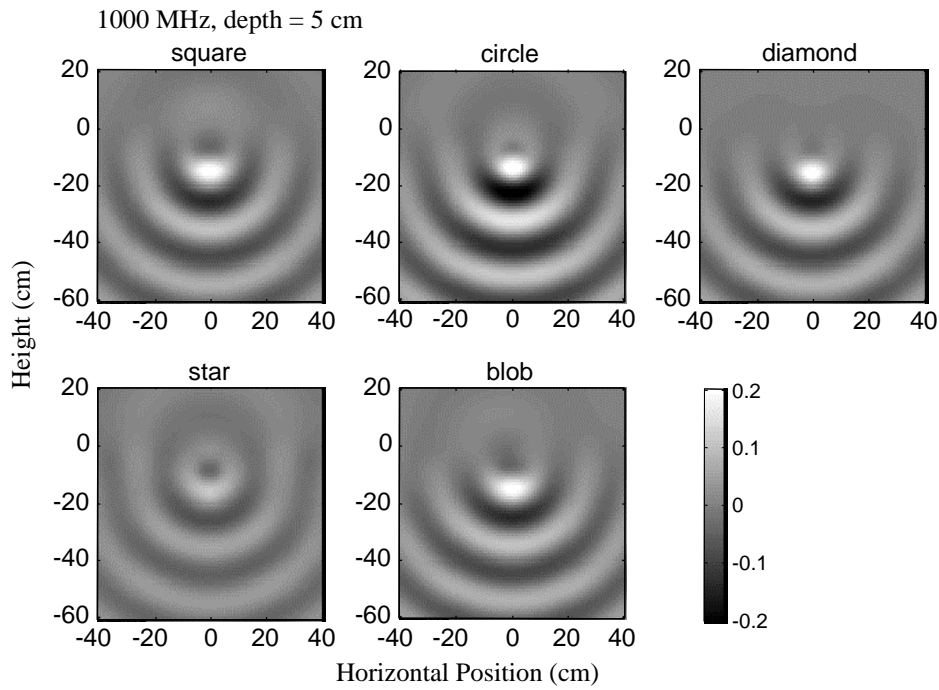


Figure 6: Real parts of fields scattered from the target shapes of Fig. 4 for 1 GHz excitation.

The surface scattered fields for the two excitation frequencies are shown in Figure 7. These would be the signals measured by radar near the surface of the ground across the target from 40 cm to one side to 40 cm to the other. Although slight differences are visible for 500MHz scattering, these would be difficult to detect in practice. However, for 1 GHz, there are quite visible differences due to shape. For example, a tilted square has almost an order of magnitude lower scattered signal than the level square.

Surface Field - Magnitude

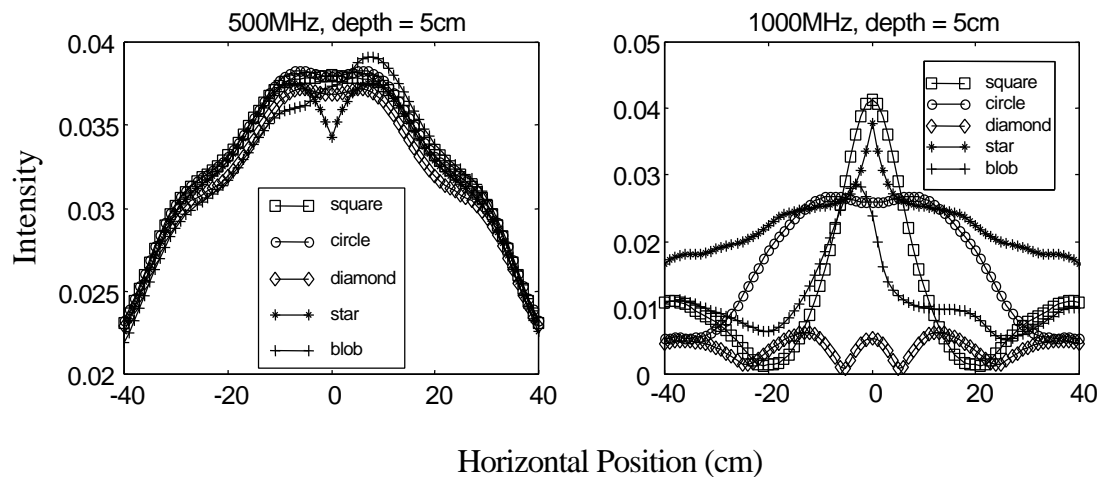


Figure 7: Magnitude of scattered field at the ground surface for the target shapes of Fig. 4: 500 MHz excitation (left) and 1 GHz (right).

Clearly, between 500 MHz and 1 GHz, the appearance of shape features becomes apparent. Figure 8 shows this effect along with burial depth dependence for the 100 cm² cross-section circle and square. At 750 MHz there are several slight differences between the scattered signals for each target depth: 1 cm, 5 cm, and 10 cm below the surface. However, these differences are not nearly as prominent as those for 1 GHz incident waves.

Finally, Figure 9 shows that the angle of incidence of the exciting wave does not help to distinguish target shape. Once again, 100 cm² cross-sectional targets are illuminated with 500 MHz plane waves. While the pattern of scattered field is indeed different for incidence angles of 30 or 45 degrees, there is no noticeable difference between scattering from the circle as oppose to the square target.

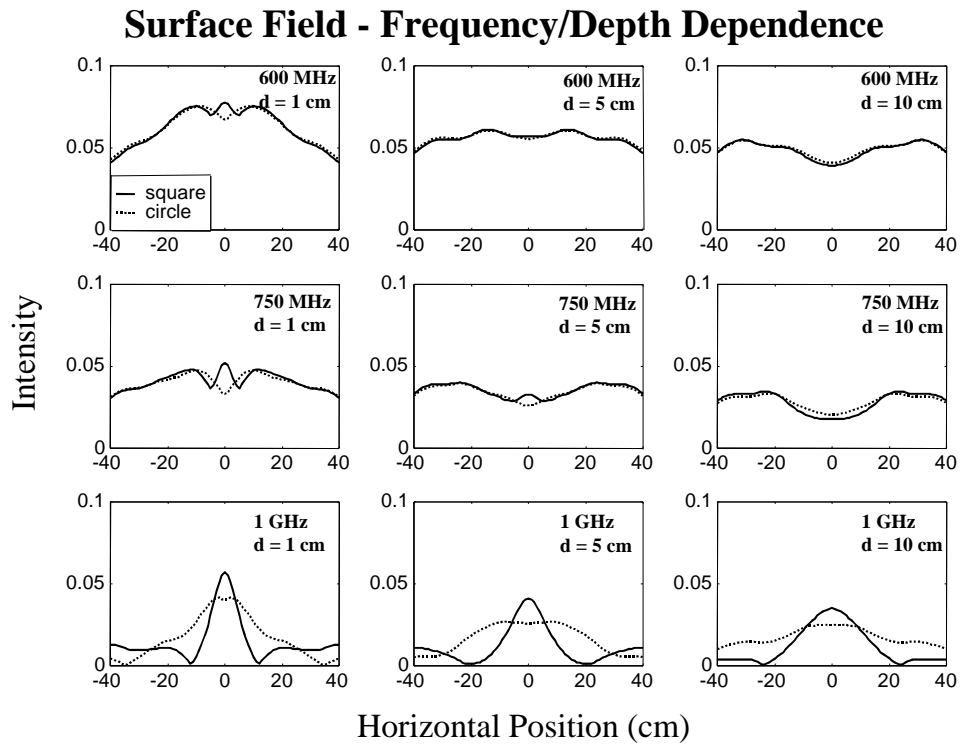


Figure 8: Surface field magnitude at the ground surface for circular and square targets for 600 MHz (top row), 750 MHz (middle row), and 1 GHz (bottom row); and burial depths 1 cm (left), 5 cm (middle), and 10 cm (right).

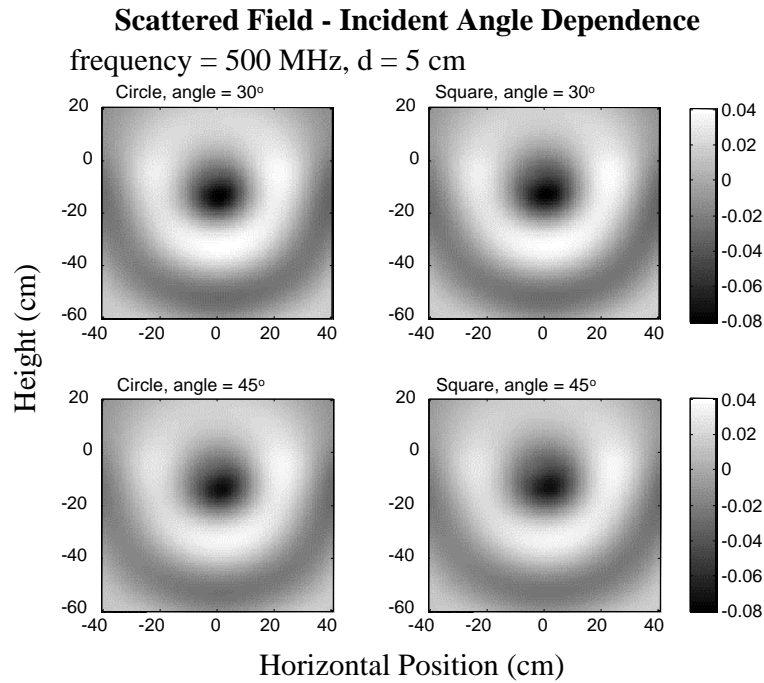


Figure 9: Real part of scattered field from circular (left) and square (right) targets at 500 MHz for inclined incidence angle excitation.

3. Conclusions

The scattered fields from various two-dimensional non-metallic mine-like target shapes have been calculated and compared. Although the target models are relatively simple, analysis of these fields indicates that for typical antipersonnel mine sizes of 10 cm, there is very little shape-distinguishing feature information available when the incident wave is at frequency 600 MHz or below. Aspect ratio appears to be the dominant discriminator of target shape, while corners, rotational symmetry, tilting, depth of burial, even concavity of target boundary do not seem to contribute much information to the scattered signal. At 1 GHz, shape features of 10 cm targets are quite evident, but a wide aperture of multiple sensors across the ground surface are needed to measure the variations of scattered field.

Clearly, for non-metallic antipersonnel mine detection, frequencies above 600 MHz must be employed if there is to be any hope of distinguishing mines in terms of the shape of target anomalies in soil.

4. Acknowledgements

The authors gratefully acknowledge the support of The Army Research Office MURI grant DAAG55-97-1-0013.

5. References

- [1] K. Yee, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media", *IEEE Transactions on Antennas and Propagation*, Vol. AP-14, pp. 302-307, 1966.
- [2] K. Kunz and Luebbers, R., *The Finite Difference Time Domain Method for Electromagnetics*, CRC Press, 1993.
- [3] C. Rappaport and W. Weedon "A General Method for FDTD Modeling of Wave Propagation in Arbitrary Frequency- Dispersive Media", *IEEE Transactions on Antennas and Propagation*, pp. 401-410, March 1997.
- [4] Rappaport, C., Wu, S., and Winton, S. "FDTD Wave Propagation Modeling in Dispersive Soil Using a Single Pole Conductivity Model", accepted for publication in *IEEE Transactions on Magnetics*, May 1999.
- [5] R. Luebbers, F. Hunsberger, K. Kunz, R. Standler, and M. Schneider, "A Frequency-Dependent Finite Difference Time Domain Formulation for Dispersive Materials", *IEEE Transactions on Electromagnetic Compatability*, pp. 222-227, Vol. 32, No. 3, March 1990.

- [6] C. Rappaport, M. Kilmer, and E. Miller, "Accuracy Considerations in Using the PML ABC with FDFD Helmholtz Equation Computation", accepted for publication in *International Journal of Numerical Modeling*, 1999.
- [7] E. Marengo and C. Rappaport and E. Miller, "Optimum PML ABC Conductivity Profile in FDFD", accepted for publication in *IEEE Transactions on Magnetics*, 1998.
- [8] M. Kilmer, E. Miller, and C. Rappaport, "Preconditioners for structured matrices arising in subsurface object detection", *SIAM Journal on Scientific Computation*, in review.
- [9] C. Rappaport, "Interpreting and improving the PML absorbing boundary condition using anisotropic lossy mapping of space", *IEEE Transactions on Magnetics*, Vol. 32, pp. 968-974, 1996.
- [10] A. von Hippel, *Dielectric Materials and Applications*, Wiley, New York: 1953.

COMPARISON OF A RECURSIVE T MATRIX METHOD AND THE FDTD METHOD FOR SCATTERING PROBLEMS IN LOSSY DISPERSIVE SOIL*

Scott C. Winton, Adnan Şahin, Carey M. Rappaport and Eric Miller

Center for Electromagnetics Research

Northeastern University

Boston, MA 02115

Abstract A comparison between two forward solving methods, recursive T-matrix and FDTD, is presented. The strengths and weaknesses of both methods are discussed. The recursive T-matrix method is a fast solver that is well suited to solutions in homogeneous media where the scatterers are bodies of rotation. The FDTD method is best suited to complicated and realistic problems involving inhomogeneities, rough surface interfaces and irregular shaped scatterers. Numeric results from both methods are presented.

I. INTRODUCTION

Interest in land mine remediation has helped to initiate development of techniques to accurately simulate electromagnetic waves propagating in a lossy dispersive medium, such as soil. Development of different techniques has indicated that the 'best' method to be used may depend greatly on the specific problem to be solved. With this in mind we discuss two such methods, the Recursive T Matrix method and the Finite Difference Time Domain method in order to compare and contrast the two methods to illustrate their respective strengths and weaknesses and therefore give an indication as to which types of problems each is best suited.

We begin with a brief description of each method, including a general overview as well as identification of the parameters required for comparison. This will be followed by a comparison of results from several numeric experiments. Then a detailed discussion is given which cites specific difference between the methods and how changing the problem to be solve effects each method.

II. THE RECURSIVE T MATRIX METHOD

For a single scatterer, the T-matrix method involves finding the coefficients to describe the scattered field as an expansion of basis functions(e.g. Bessel functions for 2-D problems). The elements of the T-matrix are found by using a Greens function and applying Poincaré-Huygens principle. For multiple scatterer, the T-matrices from individual scatterers are incorporated into a single T-matrix through a recurrence relation. This single T-matrix will include all multiple scattering effects. For a detailed analysis of single and recursive T-matrix algorithms, the interested reader is referred to [1-3].

* This work was supported by MURI Grant DAAG55-97-1-0013

Unlike finite difference techniques, the T-matrix methods do not require an absorbing boundary condition and substitutes the discretization of space with harmonic expansions of the fields thereby reducing the number of unknowns for a wide range of problems. Chew and co-workers have pioneered development of a number of fast, recursive T-matrix algorithms for determining the scattered fields in a variety of scenarios [4-6]. These methods basically function by tessellating electrically large objects into small sub-scatterers whose individual T matrices can be well represented using low-order harmonic expansions. A recursive formula then is used to aggregate the effects of all the sub-scatterers to compute the fields.

In [7], Şahin and Miller have extended the results of recursive T-matrix techniques to near field scattering problem using high-order harmonic expansion while keeping the computational cost at reasonable levels. This is an important extension since ground penetrating radar problems often require looking at near field solutions. The shortcoming of these recursive T-matrix methods is that they require a homogeneous background, since the Green's function needs to be expanded in cylindrical (or spherical in 3-D) harmonics in closed form. There has been some work on T-matrix techniques with inhomogeneous medium, but a computationally effective, and accurate technique is required for inhomogeneous background.

In order to use the T-matrix method for wide band excitation, the technique must be run several times at different frequencies. This may be seen as a disadvantage, but it allows dispersion to be dealt with in a straightforward manner, since electrical parameters may be defined for each frequency.

Before continuing, we must identify the parameters of the T-matrix method that will be important in the subsequent discussions. If Bessel functions (of spherical Bessel functions) are to be used as the basis functions, then clearly cylinders (or spheres) are the elemental type of scatterer. The previously mentioned tessellation of large or irregular shaped objects would be into small cylinders. For the single scatterer matrix the infinite series of basis functions will be truncated to M harmonics. It can be shown that for accurate representation of the fields due to this single scatterer, M must be increased as the radius of the scatterer is increased. In order to relate the T-matrices of the single scatterers for the multiple scatterer case, there must be a common reference point called the scattering origin, which is different from the global origin of the problem geometry. In [7] it is stated that a circle around the scattering origin must include all scatterers and none of the observation points. Furthermore, for accurate results, it is required that N harmonics are used to relate the individual T-matrices together. It can be shown that N must be increased as the distance from the scatterers to the scattering origin is increased.

III. THE FINITE DIFFERENCE TIME DOMAIN METHOD

The Finite Difference Time Domain method addresses the forward solving problem by discretizing the entire problem space and time stepping through spatial updates based on nearest neighbors. The use and development of the FDTD method has seen exponential growth over the last decade [8]. One of the primary reasons for this growth is that the technique is basically quite simple and so readily adaptable to a wide variety of problems.

Unlike the T-Matrix method, inhomogeneity is trivially modeled in FDTD since all that is required is to define the electrical parameters at each grid point. Scatterers are defined in the same way. For ground penetration radar applications, an air-ground interface is easily incorporated.

The use of the time domain allows wide band excitations, such as gaussian pulses or Ricker wavelets to be easily implemented. However, the inclusion of dispersion is problematic since multiplication in the frequency domain

becomes convolution in the time domain. Considerable effort has been expended on this problem with impressive success [9-11]. The obvious disadvantage is that to include these techniques greatly increases the complexity of the algorithm.

An important consideration with FDTD is the requirement of an Absorbing Boundary Condition(ABC) to eliminate reflections that would otherwise be caused by the abrupt termination of the discretized problem space. Great advancement in ABCs have been made over the last few years. Most notably the introduction of Berenger's Perfectly Matched Layer [12], which has helped to greatly increase the usefulness of the FDTD method. Unfortunately, as the complexity of the medium is increased, the problem of finding an effective ABC becomes more difficult. Once again, considerable work has been done on this problem and has led to the introduction of ABCs for lossy and dispersive material. However, in general, the performance of these ABCs tends to fall well short of their dielectric counterparts.

IV. NUMERICAL EXPERIMENTS

In order to facilitate a comparison between the two methods and to help illustrate the relative advantages of the methods. Four numerical experiments have been performed. In each case, a Transverse Magnetic continuous wave (CW) uniform plane wave at 1Ghz is incident on the scatterer(s) in a homogeneous medium typical of soil with dielectric constant, $\epsilon_r = 6.0$ and a conductivity $\sigma = .05S/m$.

For the FDTD method, it will be shown that the two critical parameters are the spatial increment Δ and the Courant number R . The Courant number is the ratio of the distance moved by the wave in one time increment to the spatial increment, i.e. $R = v\Delta t/\Delta$. Stability analysis reveals that for a simulation on a cubical grid to remain stable, $R \leq 1/\sqrt{d}$, where d is the dimension of the simulation.

Figure 1 shows the geometry for the first experiment. A single non-conducting dielectric scatterer is 15cm away from the array of receivers. The radius r of the scatterer is 3.75cm. For the FDTD, $R = 0.5$ and $\Delta = 5mm$ which translates to a sampling of about 25 points per wavelength in the soil and about 38 points per wavelength in the dielectric scatterer, which should be more than adequate. However, using $\Delta = 5mm$ means that there are only about 7 points along the radius of the scatterer. If Δ were any larger, the outer surface of the scatterer would not be smooth, affecting the scattered field. For the T-matrix, $M = 20$. Figure 2 shows the magnitude of the solutions obtained from both the T-matrix method and the FDTD method.

Figure 3 shows the geometry of the second experiment. Here the dielectric scatterer has been replaced with a perfect electric conductor $\sigma \rightarrow \infty$. As with the dielectric scatterer, the radius $r = 3.75cm$. As with the previous experiment, $M = 20$ for the T-matrix solution. Figure 4 shows both the T-matrix and FDTD solutions for this geometry.

For these first two experiments, the agreement between the two methods is excellent. Clearly both methods are capable of producing accurate results for these simple cases. For the T-matrix method, no recursion is required to find the scattered field. And since the scatterer is a cylinder, no tessellation is required.

Figure 5 shows the geometry for the third experiment. Here there are two PEC scatterers of radius $r = 30cm$. The separation is 1m. Because the radius of these scatterers is considerably larger than the previous case the spatial increment of the FDTD method has been increased to no more than 2cm. For the T-matrix, $M = 25$ and $N = 110$.

Figure 6 shows the T-matrix solution as well as two FDTD solutions.

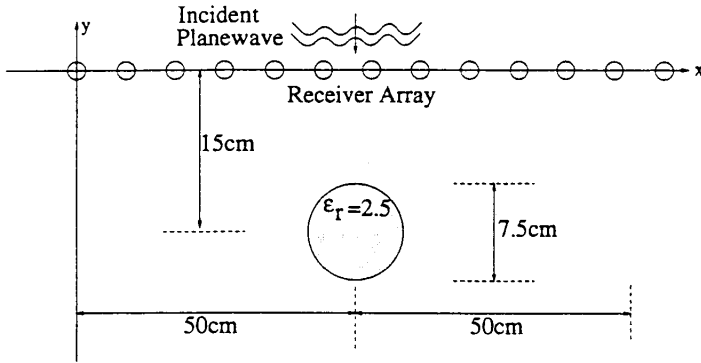


Figure 1: Geometry of experiment #1.

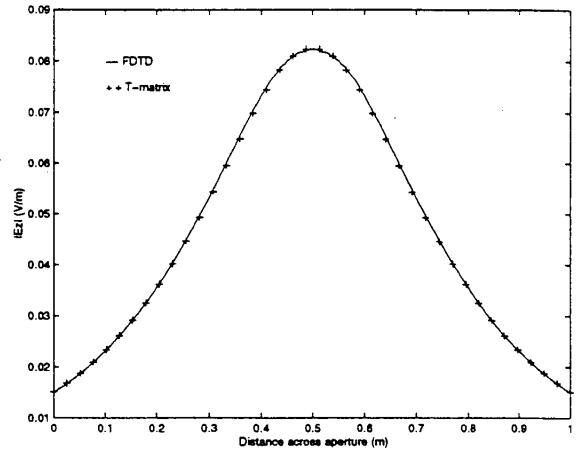


Figure 2: Magnitude of scattered electric field due to single dielectric scatterer of T-matrix and FDTD

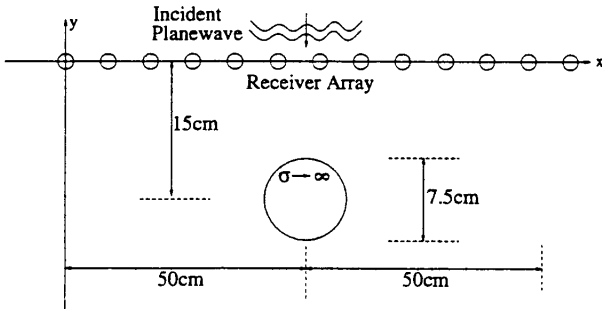


Figure 3: Geometry of experiment #2.

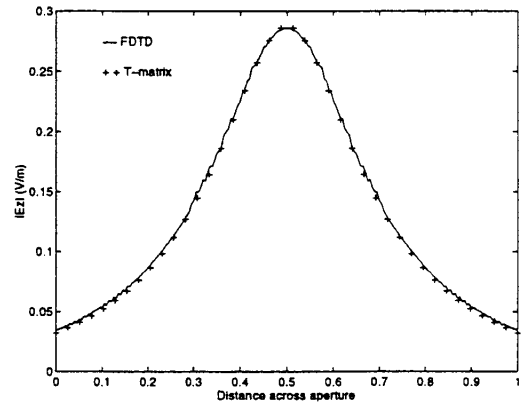


Figure 4: Magnitude of scattered electric field due to single PEC scatterer of T-matrix and FDTD

Figure 6 shows that the FDTD solution with $\Delta = 2\text{cm}$ is noticeably different than that of the T-matrix solution. At 1GHz , $\Delta = 2\text{cm}$ represents a sampling of only 6 points per wavelength, which is clearly inadequate. By decreasing Δ to 1cm , the sampling becomes about 12 points per wavelength and the solutions achieved good agreement.

Figure 7 shows the geometry for the fourth and final comparison. Here there are two dielectric scatterers with $r = 3.75\text{cm}$, the separation is 25cm . For the T-matrix solution, $M = 12$ and $N = 120$. Figure 8 shows the T-matrix solution and the FDTD solutions for this case.

Figure 8 shows that although there are three different levels of spatial resolution, there is still a discrepancy, mostly to the left and right of the scatterers, between the two methods. Since there is no analytic solution available, it is difficult to determine which method is inaccurate. Resolution of this discrepancy is the focus of continuing

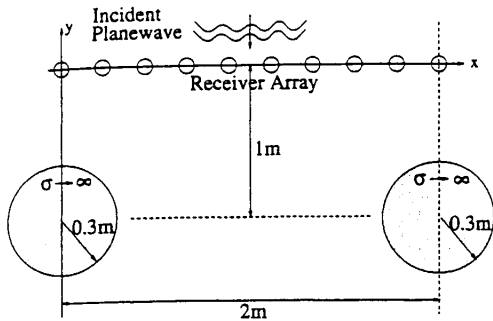


Figure 5: Geometry of experiment #3.

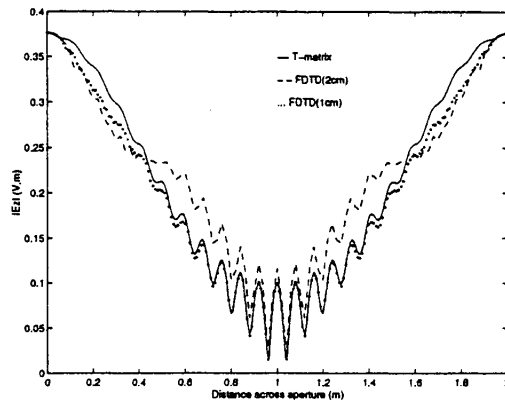


Figure 6: Magnitude of scattered electric field due to two PEC scatterers of T-matrix and FDTD

investigation. However, there is good agreement in the region between the scatterers, where most of the interaction between the scatterers takes place.

IV. COMPARISON

As part of the comparison between the two methods, the issue of 'speed' should be addressed. A comparison of execution speed is, unfortunately, not easily accomplished. If FDTD were a matrix problem, like T-matrix, we could just compare the number of computer program flops. However, the FDTD code used in these experiments was written in Fortran and executed on a PC using the Linux operating system. As it turns out, a high resolution comparison is not required because for every case considered here, the T-matrix method has a shorter execution time. For the cases of high spatial resolution, i.e. $\Delta \ll 1$, the execution time for the T-matrix method is significantly shorter than that of the FDTD method. To include some approximate numbers, all of the T-matrix solutions may be obtained using Matlab and a PC in less than 10 minutes. Comparable speed may be obtained with FDTD using the largest Δ possible and a suitable ABC. For example, for the geometry of Figure 7, with $\Delta = 5mm$ and using an 8-layer Perfectly Matched Layer [12] ABC, the simulation can be completed in less than 8 minutes. However, in order to decrease the programming complexity, the grid can be enlarged and the ABC omitted. This was done for each of the experiments. For the geometry of Figure 7 with $\Delta = 5mm$, the run time was about 20 minutes, for $\Delta = 2.5mm$ the run time was about 70 minutes and for $\Delta = 1.25mm$ the run time was over 7 hours. These numbers are clearly not representative of all FDTD or T-matrix solutions, since specific problems, requirements, hardware and software can vary widely. It can, however, be said with confidence, that for case where there is homogeneity and simple scatterers, the T-matrix method will be faster, and in many cases, considerably faster.

The effect that changes in the problem have on the execution time should not be understated and it is worthwhile to examine these effects for both methods. Changing the medium from homogeneous to inhomogeneous requires minimal effort for the FDTD method. However, since the T-matrix method for inhomogeneous material is still under development. Changing to an inhomogeneous material is quite difficult. Changes in the size of the problem can effect both methods in different, sometimes opposite ways. For example, increasing the size of scatterers may actually decrease the computational cost of the FDTD method, since Δ may be larger to achieve the same number of points per wavelength. However, an increase in the size of the scatterers requires an increase in the number of harmonics M for the single scatterer matrices. In other words, FDTD is better suited to electrically large curved scatterers while T-matrix is better suited to electrically small curved scatterers. If the scatterers are held at the same size

but moved apart, this would probably most adversely effect the FDTD method since more spatial points would be necessary. This not only effects the size of the grid, but the amount of time required to propagate the waves between the scatterers. This amount of time may be decreased by increasing the temporal increment. However, the temporal increment cannot be increased without bound since the Courant condition would not be satisfied and the simulation would become unstable. For T-matrix, if a suitable scattering origin can be found, moving the scatterers further apart simply requires increasing the number of harmonics N which are required to relate the matrices together. Adding dispersion to the problem will increase the execution for both methods, but the relative increase between the methods will require further investigation. For the T-matrix method, the solution must be set up and solved for each desired frequency. This will require multiple runs. For FDTD, dispersion can be accomplished in one simulation, but incurs a computational cost in additional difference equations and increased storage space.

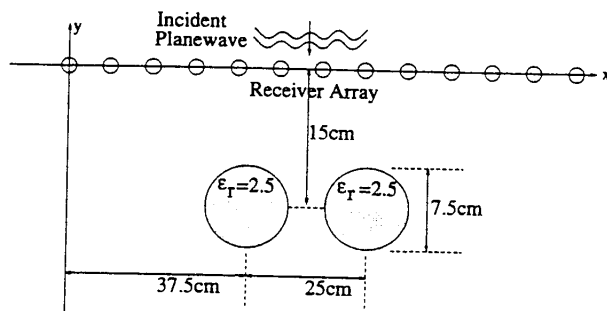


Figure 7: Geometry of experiment #3.

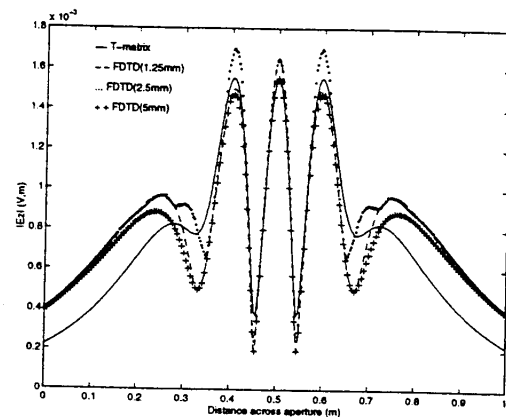


Figure 8: Magnitude of scattered electric field due to two dielectric scatterers of T-matrix and FDTD

Although the T-matrix solution for these problems may be obtained more quickly once the problem is set up, the actual creation of the program to solve the problem numerically is more involved. The FDTD method requires less understanding of the physics of the problem and fewer programming skills.

Changes in the problem can also effect the complexity of the program for each of the two methods. Changes in the size of the problem are very easily dealt with in FDTD, usually only requiring changing a few variable. For T-matrix, if the size of the problem is changed, it may require locating a new scattering origin. Changes in the shape of the scatterer are also relatively easy for FDTD. The scatterers can be determined by changing the electrical parameters point by point. For the T-matrix, changes in the shape of the scatterers can have *drastic* effects of the complexity of the problem. If a scatterer is tessellated, the final T-matrix must incorporate the T-matrices of all of the smaller scatterers. Furthermore, care must be taken to ensure that the small scatterers make an accurate representation of the larger scatterer. For example, if many small cylinders are to make up a large irregular shaped scatterer, there will be gaps in between the cylinders that represent errors. Special algorithms must be incorporated to minimize these gaps. Also tessellation is not really appropriate for highly lossy metallic scatterers since the fields inside the scatterer are zero (or very small). If a metallic scatterer is irregularly shaped, schemes must be employed to only incorporate the effect of the outside edge of scatterer into the recursive T-matrix. Adding dispersion to the problem has the opposite effects to changing the shape of the scatterers, i.e, it is much more detrimental for FDTD computation. For the T-matrix, the electrical parameters for a given frequency are easily incorporated into the algorithm, the method is simply employed several times. Because dispersion in the time domain requires

convolution, it can be quite problematic to incorporate it into the FDTD method. With the addition of dispersion, the relative simplicity of the FDTD algorithm is lost and much greater complexity is required to insure accurate results.

In order to be useful, a forward solver must be accurate. As mentioned previously the accuracy of the T-matrix method is controlled by the number of harmonics used while for the FDTD method, the accuracy is controlled by the spatial increment. There is also a temporal increment used in FDTD, but if the Courant condition is met, the accuracy can be seen to be controlled by Δ . Ten to twenty spatial points per wavelength are usually considered adequate to obtain sufficient accuracy. However, this also puts a limit on the size of the grids that can be discretized as discussed above. Also, there is another source of error in FDTD known as numerical dispersion. Numerical dispersion arises from the fact that in FDTD, $k \neq \omega/v_p$, where k is the wave number, ω is the radial frequency and v_p is the phase velocity. Solving the dispersion relation for the discretized wave equation reveals that there is another term in the dispersion relation that is directly dependent on Δ and can be shown that numerical dispersion can be reduced by reducing Δ . Numerical dispersion is usually not a problem for short simulation, but can cause inaccuracies in lengthy simulations.

For FDTD, it was mentioned that the Courant condition, $R \leq 1/\sqrt{d}$ must be met to insure stability. It should be seen by now that the Courant number is the *critical* parameter for the method since it effects not only the stability, but the execution time and the complexity of the problem. These facts in turn limit the types of problems that can be solved with FDTD. For the T-matrix method, it was originally found that instability could arise with higher order harmonics, hence the motivation to tessellate larger objects in order to be able to use lower order monopole expansions. However, that problem has been eliminated for many problems of interest [7] and it has been shown that the use of higher order expansions in un-tessellated scatterers can actually reduce the complexity of the problem.

V. CONCLUSIONS

Through numerical experiment and the subsequent discussion, a comparison of the T-matrix and the FDTD method for electromagnetic forward solving problems has been performed, the relative strengths and weaknesses of each method has been discussed. It should be clear that the advantages of the T-matrix solution are speed with accuracy whereas the advantages of FDTD are simplicity and the ability to adapt FDTD to a wide variety of practical problems.

Obviously, these strengths should effect the decision of which method to use for a given situation. For problems with bodies of rotation in a homogeneous environment, T-matrix is a clear choose. But T-matrix may also be used as an approximation to a more realistic problem in order to perform a proof of principle solution. For the situation that requires more unusual geometries, and especially for wide band excitation or where the transient responses are desired, the FDTD is obvious chose. Both methods represent powerful tools in electromagnetic forward solving problems for use in land mine remediation.

REFERENCES

- [1] Waterman, P.C., "New Formulation of Acoustic Scattering", *J. Acoust. Soc.*, Vol. 45, No. 6, pp. 1417-1429, 1969.
- [2] Waterman, P.C., "Symmetry, Unitarity and Geometry in Electromagnetic Scattering", *Physical Review D*, Vol. 3, No. 4, pp. 825-835, Feb 1971.

- [3] Peterson, B, Strom, A., "Matrix Formulation of Acoustic Scattering from an Arbitrary Number of Scatterers", *IEEE Trans. Antenna and Propagation*, Vol. 56, No. 3, pp. 771-780, September 1974.
- [4] Chew, W.C., Friedrich, J.A. and Geiger, R., "A Multiple Scattering Solution for the Effective Permittivity of a Sphere Mixture", *IEEE Trans. Geoscience and Remote Sensing*, Vol. 28, No. 2, pp. 207-214, March 1990.
- [5] Wang, Y.M, Chew, W.C., "An Efficient Algorithm for Solution of a Scattering Problem", *Microwave and Optical Technology Letters*, Vol. 3, No. 3, pp. 102-106, March 1990.
- [6] Chew, W.C., Wang, Y.M, Gurel, L., "Recursive Algorithm for Wave-Scattering Solutions Using Windowed Addition Theorem", *Journal of Electromagnetic Waves and Applications*, Vol. 6, No. 11, pp. 1537-1560, 1992.
- [7] Şahin, A., Miller, E., "Recursive T-Matrix Methods for Scattering from Multiple Dielectric and Metallic Objects", "IEEE Trans. Antennas and Propagation", To appear in May 1998.
- [8] Shlager, K.L., Schneider, J.B., "A Selective Survey of the Finite -Difference Time-Domain Literature", *IEEE Antennas and Propagation Magazine*, Vol. 37, No. 4, pp. 39-56, 1995.
- [9] Weedon, W., Rappaport, C.M., "General Method for FDTD Modeling of Wave Propagation in Arbitrary Frequency-Dispersive Media", *IEEE Trans. Antennas and Propagation*, Vol. 45, No. 3, pp. 401-410, 1997.
- [10] Sullivan, D.M, "Z Transform Theory and the FDTD Method", *IEEE Trans. Antennas and Propagation*, Vol. 44, No. 1, pp. 28-34, 1996.
- [11] Kelly, D.F., Luebbers, R.J., "Piecewise Linear Recursive Convolution for Dispersive Media Using FDTD", *IEEE Trans. Antennas and Propagation*, Vol. 44, No. 6, pp. 792-797, 1996.
- [12] Berenger, J., "A Perfectly Matched Layer for the Absorption of Electromagnetic Waves", *Journal of Computational Physics*, Vol. 114, No. 1, pp. 185-200, October 1994.

Conductivity Profile Optimization for the PML ABC in FDFD

Carey M. Rappaport, Edwin A. Marengo, Eric L. Miller
Center for Electromagnetics Research
Northeastern University
Boston, MA 02115

The Perfectly Matched Layer (PML) proposed by Berenger (Berenger *J. Comp. Phys.*, 1994) is an extremely effective lattice termination condition in both time domain and frequency domain field computation. The ideal, continuous specification of the PML can have arbitrary loss characteristics, which can absorb without reflection waves incident from any angle. In finite difference wave modeling, however, the staggered electric and magnetic field grids prevent the specification of the same medium at a given position for both E-field and H-field, leading to numerical reflections. To minimize these reflections, the value of conductivity in each PML layer must be profiled so that it is smallest next to the free space/PML boundary, and rises for subsequent layers. The optimal specification of this profile is an open question of much interest to the computational electromagnetic community.

Unlike the in time domain case (Gedney, *ACES* 1997, and Fang and Wu, *IEEE Micro. Guided Wave Let.* 1996), the FDFD formulation of a multilayer PML can be stated explicitly and exactly. One way of doing this is to use the discretized impedance of one layer relative to that of the subsequent layer to iteratively determine the impedance “looking into” the PML at its boundary, relative to the PML termination. For normalized impedance of the i -th layer of width Δ , $\eta_i = (E_i/H_{i+\frac{1}{2}})/\eta_0$, for a plane wave incident with angle θ on a planar layer boundary normal to \hat{x} , the recursive relation is:

$$\eta_i = jk\Delta \cos^2 \theta + \frac{S_{i+\frac{1}{2}} + S_i}{2} \cos^2 \theta + \frac{1}{jk\Delta + (S_{i-\frac{1}{2}} + S_i)/2 + 1/\eta_{i+1}}$$

where $k = \omega\sqrt{\mu\epsilon}$ and $S_i = k\Delta\sigma_i/\omega\epsilon_0$ ($= \sigma_i\eta_0\Delta$ when all PML layer dielectric constants are 1). Note that this formula allows the specification of conductivity values at every half layer, doubling the number of layer parameters from the original. Also, the effects of a tuned PML termination, given by η_{max} (Rappaport, *IEEE Trans. Mag.*, 1996) can be carefully examined.

To find the best conductivity coefficients for a wide range of angles and frequencies, numerical optimization is performed to minimize a cost function based on the error between the free-space transverse wave impedance, $\cos \theta e^{jk\Delta \cos \theta/2}$, and the impedance looking into the PML, η_1 . The goal is to find a set of σ_i and a termination η_{max} that keeps the weighted error small for all frequencies of interest (0 to 75°, say) and all frequencies of interest. The solution to this problem also applies to the coefficients for the FDTD PML layer.

Advanced Land Mine Detection Using a Synthesis of Conventional Technologies

Carey M. Rappaport
Center for Electromagnetics Research
Northeastern University
Boston, MA 02115
United States

The humanitarian need for new, technologically advanced yet cost-effective approaches to demining has recently become critical. First, the proliferation of mines has increased to the extent that not only are there more than 100 million mines buried throughout the world, but new mines are being placed faster than they can be cleared. And second, newer types of mines with inexpensive plastic and non-ferrous metal casings—which are difficult to detect with conventional methods—are becoming more prevalent.

Developing the next generation of demining technology requires using multiple sensors and signal processing algorithms which take into account the way in which the sensor data is gathered rather than treating the data as generically-derived information. We at Northeastern University are leading a team to develop and optimize land mine detection based on ground-penetrating radar (GPR), infrared thermography (IRT), electromagnetic induction (EI), and high frequency acoustic (HFA) sensors. We are implementing sophisticated, physics-based mathematical models to describe the interaction of EM or acoustic waves with mines buried in realistic (electromagnetically lossy, inhomogeneous) soil and as a result are developing “smart” signal processing algorithms to identify and classify mines. These mathematical models are derived from actual soil and land mine measurements, and include detection statistics of the sensors. In addition, we are building prototype land mine detection systems using currently available industrial hardware configured in novel ways based on the physical models under development. This approach allows us to utilize cheap, off-the-shelf components and “smart” algorithms with the hope of providing developing countries with reliable and cost-effective sensor systems. Finally, we are working to integrate these components into a set of scenario-tunable systems for person-portable, vehicle-mounted, and/or airborne use that will be effective for demining operations in both hostile and peaceful areas.

To achieve these goals requires a research program based on a first-principles understanding of the technical challenges posed by the demining problem and the development of an integrated framework for addressing these issues. At its root, the demining problem is particularly challenging because it is an *inverse* problem. That is, one can often obtain information regarding the presence of mines *only* through indirect measurements using electromagnetic, chemical or thermal sensors. Most sensing systems operate by transmitting energy into the earth and measuring the scattered fields arising from the interaction of the energy with sub-surface structures. All mine-related information to be extracted from the data is encoded by the complex scattering processes underlying these non-invasive detection methods.

Non-ferrous mines are undetectable with magnetometry, while quick mobile deployment precludes particle beam and spectrometry methods. Shallow wave penetration in soil limits millimeter and visible frequency techniques, while resolution limitations create difficulties for seismic and gravity sensing. Although olfactory, chemical, and biological sensing may offer advantages for detecting older mines (and hence be useful for humanitarian purposes), they have limited sensitivity, especially in battlefield situations where pervasive explosive residues may be widely scattered across the ground. Newer mines do not give off significant chemical vapors and there is no direct means of continuously sampling soil.

By concentrating our efforts on EI, GPR, IRT, and HFA sensors, we feel that we have the best opportunity for making significant advances in demining. These technologies have been available for many years, and have individually met with limited success. Most importantly, all four technologies are easily integrated since they possess similar modalities, involve similar wave propagation, can be addressed with unified mathematical models, and can be mounted on a variety of demining platforms. We are performing experiments to determine the optimal platforms, sensors, and frequency ranges for specific demining situations and plan to exploit

the capability of multiple sensors and physically-based signal processing algorithms to generate substantial improvements over currently available land mine detection schemes.

Ultrawideband Array-Based Ground-Penetrating Radar

As with standard radar, GPR is useful for the remote sensing of submerged targets. It is an important tool for determining the locations of unexploded ordnance, land mines, underground installations, and hidden arms caches, but also offers many dual-use applications, including precisely identifying the positions of buried waste drums (for environmental clean-up), finding pipes or cables (of interest to utilities), and identifying cavities or obstacles (for construction applications). Ground penetrating radar has been available for many years, and has been fairly successful at the remote identification of buried objects. However, by making use of short pulse and shaped pulse signals, several important advantages become apparent. First, for wet, lossy soils, the wave penetration depth is severely limited for high frequency signals, while the detector resolution is limited at low frequencies; short pulse signals can be tailored to optimize the trade-off between penetration and resolution. Second, since most of the targets are in the nearfield of the radar source, the signal phase effects are unusually important. By mixing the phases of a wide bandwidth of frequencies appropriately, the signals scattered by the target can be more sharply analyzed, and the targets more accurately identified.

Soil is inhomogeneous and dispersive, with dielectric and electrical conductivity parameters depending on the soil composition, density, and excitation frequency. One research effort at Northeastern University uses measured soil data to form a simple clutter model appropriate for both time- and frequency-domain computational wave propagation simulations. In addition, we will include the effects of dielectric constant and conductivity variations as well as rock inclusions as perturbations of the propagation medium. Instead of having to guess soil characteristics or use a look-up table, this model will provide high-accuracy predictions of the penetration depth, resolving capability, and sensitivity to noise and clutter of the EM signals.

It is only through use of the entire RF frequency band and the largest possible aperture that sufficient sensing information becomes available for sophisticated inversion processing. While monostatic data from various locations can be combined for rudimentary imaging, collecting the scattered response with multiple receivers simultaneously provides the extra dimension of information necessary to detect low contrast mine targets. We are developing strategies for optimizing element positioning in one- and two-dimensional arrays, considering synthetic aperture radar (SAR) antenna optimization, and testing trade-offs between array size and platform geometry for synthesizing the best practical sensing aperture.

For the ultra-wideband radar systems we will be developing, it is essential to make use of antenna elements which are compact, efficient, inexpensive, and can faithfully radiate all of the frequency components in the generated radar signal. Geo-Centers, Inc., an industrial partner of Northeastern University, has developed a novel wideband antenna based on a transverse electromagnetic horn with rhombic taper, folded back into itself and terminated with a 50 ohm resistive load used for both transmitting and receiving. The measured performance of this transverse electromagnetic rhombus (TEMR) indicates a very uniform radiation pattern in the plane of curling flare, perpendicular to the metal plates. It is also a wide-band antenna; throughout the 200 MHz to 2.0 GHz range, the radiated signal faithfully duplicates the input signal shape *for all angles*. Thus, the TEMR appears to be close to an optimum antenna element for a time-domain radar array. This wideband antenna has been incorporated into a time domain array as part of a vehicle-mounted GPR countermine system. The detection results for this system have been very encouraging, with 100% of all metal mines and over 90% of plastic mines detected in realistic simulations.

Microwave Enhanced Infrared Thermography

In addition to using individual microwave and infrared sensing systems, we propose a hybrid sensor in which a high-power microwave transmitter is combined with an infrared camera. The transmitter introduces energy into the ground which is absorbed at different rates by buried objects and the surrounding soil. The resulting differential heating is sensed by the infrared camera thereby leading to an image of the underground object.

While careful modeling will be required to evaluate performance and optimize the transmitting antenna design, the basic soundness of the approach has been demonstrated with simulated differential sunlight heating. We are examining the selection of a microwave wavelength to penetrate to the desired mine depth making use of the dielectric contrast of the mine as well the contrast in its thermal parameters. Microwave energy can be focused and directed into the ground at the pseudo-Brewster angle to minimize reflection and thus maximize coupling. For the common variations in soil characteristics, an elliptical reflector-type antenna could be positioned so that its major axis aligns with the Brewster angle far in front of a detector. The converging rays from all parts of the elliptical reflector will arrive at the target focal point with a range of angles. Finally, by taking IR images before and after energy is delivered, difference images can be constructed which will reduce the masking.

Acoustic Phenomena

Although the acoustic mismatch between particulate soils and solid mines makes acoustic sensing a promising area of research for mine detection, previous studies have demonstrated the difficulty of coupling sufficient acoustic energy from air into the ground. Contact sensors, on the other hand, are slow and awkward for use by moving troops, and dangerous in applications where pressure must be minimized.

One alternative is laser-induced acoustic-wave generation. CO₂ laser pulses are strongly absorbed in the surface of most solids. Modulating the laser power at acoustic frequencies produces audible and trans-audible sound vibrations. The dominant frequencies have been identified with dimensional resonances in a 300 cm³ irregular sample. Applied to a mine detection system, the laser could be raster-scanned across the area under investigation with detection through a remote ground contact or rolling sensor array. Acoustic resonance when excitation is near a mine could be correlated with the mine position by knowing the laser excitation position.

Acoustic array signal processing will benefit from the concurrent work in electromagnetic signal processing. Multistatic processing of acoustic signals will closely parallel the work on multistatic EI and GPR. Algorithms for multistatic GPR and EI will be tested with simulated or experimental HFA data to measure the enhanced probability of acoustic detection from array processing.

Summary

Although there undoubtedly exist more exotic sensors which could be used for detecting land mines, we expect that clever use of mature technologies will ultimately prove to be efficient, cost effective, reliable, and with a shorter development horizon since we will utilize commercial, off-the-shelf components which already exist. The novel aspects of Northeastern University's approach are: (1) to combine multiple sensors synergistically, yielding more information than would be available to any single sensor technology operating alone, and (2) to use "smart" signal-processing algorithms derived from physics-based models which take into account the actual sensor parameters as well as material and electrical characteristics of the soil and land mines. By processing data in an intelligent manner, rather than simply treating it as a generic bit stream, we hope to utilize information that would otherwise be discarded, and which will make mine detection much more successful.

Accepted, "IEEE Trans. Image Processing"

Wavelet Domain Image Restoration With Adaptive Edge-Preserving Regularization*

Murat Belge, Misha Kilmer and Eric Miller
235 Forsyth Building
Northeastern University
360 Huntington Ave.
Boston, MA 02215
Tel: (617) 373-8386
email: elmiller@cdsp.neu.edu

August 5, 1999

Abstract

In this paper we consider a wavelet based edge-preserving regularization scheme for use in linear image restoration problems. Our efforts build on a collection of mathematical results indicating that wavelets are especially useful for representing functions that contain discontinuities (i.e. edges in two dimensions or jumps in 1-D). We interpret the resulting theory in a statistical signal processing framework and obtain a highly flexible framework for adapting the degree of regularization to the local structure of the underlying image. In particular, we are able to adapt quite easily to scale-varying and orientation-varying features in the image while simultaneously retaining the edge preservation properties of the regularizer. We demonstrate a half-quadratic algorithm for obtaining the restorations from observed data.

EDICS: 1.4 (Image restoration)

*This work was supported by an ODDR&E MURI under Air Force Office of Scientific Research contract F49620-96-1-0028, a CAREER Award from the National Science Foundation MIP-9623721, and the Army Research Office Demining MURI under Grant DAAG55-97-1-0013,

1 Introduction

In many applications recorded images represent a degraded version of the original scene. For example, the images of extraterrestrial objects observed by ground based telescopes are distorted by atmospheric turbulence [1] while motion of a camera can result in an undesired blur in a recorded image. Despite the different origins, these two cases along with others from a variety of fields, share a common structure where the exact image undergoes a “forward transformation” and is corrupted by observation noise. The source of this noise is the disturbance caused by the random fluctuations in the imaging system and the environment. The goal of image restoration is to recover the original image from these degraded measurements.

Often the forward transformation acts as a smoothing agent so that the resulting restoration problem is ill-posed in the sense that small perturbations in the data can result in large, non-physical artifacts in the recovered image [1, 2]. Such instability is typically addressed through the use of a regularization procedure which introduces *a priori* information about the original image into the restoration process. The prior information underlying the most commonly used regularization schemes is that the image is basically smooth [2]. While the regularized restorations are less sensitive to noise it is well known that the smoothness assumption impedes the accurate recovery of important features, especially edges.

In response to this problem, there has recently been considerable work in the formulation of “edge-preserving” regularization methods which result in less smoothing to areas with large intensity changes in the restored image. These methods necessarily require non-quadratic regularization functions and therefore result in nonlinear image restoration algorithms. Along these lines, Yang and Geman [3] introduced the concept of “half quadratic regularization” which addresses the nonlinear optimization problem that results from using such functions. Later, Charbonnier-Aubert-Blanc Feraud-Barlaud [4] built upon the results of this work by providing the conditions for edge preserv-

ing regularization functions. Another recent advance in this area is the Total Variation (TV) based image restoration algorithms [5]. In this approach, images are modeled as functions of bounded variation which need not be continuous. Therefore, formations of edges are encouraged and the restorations obtained by the TV based algorithms look sharper than those obtained by conventional techniques, especially if the exact image is piecewise continuous.

In this work, we consider a statistically based, wavelet-domain approach to edge-enhanced image restoration in which we employ a stochastic interpretation of the regularization process [6–8]. We note that most all of the work to date on wavelet-based, statistical regularization methods has concentrated on the use of multi-scale smoothness priors [9–12]. While Wang *et. al.* did consider issues of edge preservation in [12], their method was based on the processing of the output of an edge detector applied to the noisy data to alter the degree of regularization in a multiscale smoothness constraint. As described below and in subsequent sections, our approach is significantly different as the edge preservation is built directly into the regularization scheme itself.

Specifically, we regard the image as a realization of a random field for which the wavelet coefficients are independently distributed according to generalized Gaussian (GG) distribution laws. This model is motivated by two factors. First, recent work [6, 7, 13, 14] suggests that these models, which have heavier tails than a straight Gaussian distribution, provide accurate descriptions of the statistical distribution of wavelet coefficients in image data. Second, in addition to being a basis for $L^2(R)$, wavelets also are unconditional bases for more exotic function spaces whose members include functions with sharp discontinuities and thus serve as natural function spaces in which to analyze images [8, 15–17]. Because the norms in these Besov spaces are nothing more than weighted l_p , $0 < p$, norms of the wavelet coefficients, it is shown easily that deterministic regularization with a Besov norm constraint is equivalent to the specification of an appropriately parameterized GG wavelet prior model. From this perspective, our work can be viewed as an extension of the research

done mostly in the area of image denoising. Specifically, the wavelet domain image model of interest in this paper and the resulting nonlinear restoration algorithm are related to the large body of work originating from the *wavelet shrinkage estimators* first proposed by Donoho and Johnstone [18]. In a series of papers, Donoho and Johnstone have shown that wavelet shrinkage estimators achieve near optimal estimation performance when the unknown signal belongs to Besov spaces. Later, several authors contributed to the advancements in the area. The notion of Besov regularization has been introduced by Amato and Vuza [17] and Chambolle-DeVore-Lee-Lucier [8] and the resulting theory was interpreted in a function space setting. On the other hand, Simoncelli and Adelson [6] developed a similar denoising scheme, which they called *Bayesian wavelet coring*, by stochastically modeling the image subbands.

In this work, we make use of GG wavelet priors in a number of ways. We show that their use in an image restoration problem does in fact significantly improve the quality of edge information relative to more common smoothness priors. Inspired by the “lagged diffusivity” fixed point iteration proposed by Vogel and Oman [19] for the solution of the TV problem, we also provide an efficient algorithm for solving the non-linear optimization problem defining the restoration. By appropriately structuring the weighting pattern on the wavelet l_p norm, we demonstrate that these models provide an easy and flexible framework for adaptively determining the appropriate level of regularization as a function of the underlying structure in the image; in particular, scale-to-scale or orientation based features. This adaptation is achieved through a data-driven choice of a vector of regularization parameters. For this task, we introduce and make use of a multi-variate generalization of the L-curve method developed in [20] for choosing a single regularization parameter. We verify the performance of this restoration scheme on a variety of images, comparing the results both to smoothness constrained methods and the TV restorations.

We recognize that there are asymptotic results which state that the L-curve does not provide

consistent estimates of the regularization parameters either as the noise level goes to zero or the data length goes to infinity [21, 22]. In the non-asymptotic regime however, empirical results do point to the practical utility of this method. Moreover, as described in [20], the L-curve framework is easily adapted to handle multiple regularization parameters, a feature required for the work here.

The remainder of this paper is organized as follows. In Section 2 we give the wavelet domain formulation of the image restoration problem. In Section 3 we introduce a multiscale prior model for images and use this model in Section 4 to develop an image restoration algorithm. In Section 5 we apply the “L-hypersurface” method to the simultaneous multiple parameter selection problem posed by our image restoration algorithm. In Section 6 we demonstrate the effectiveness of our algorithm by comparing our results with existing image restoration schemes. Finally, in Section 7, conclusions and future work are discussed.

2 Regularized Image Restoration

A grey-scale image, f , can be considered as a collection of pixels obtained by digitizing a continuous scene. The image is indexed by (m, n) , $1 \leq m, n \leq 2^J$, and the intensity at the position (m, n) is denoted by $f(m, n)$. In image reconstruction and restoration problems, the objective is to estimate the image $f(m, n)$ from its degraded measurements. Mathematically, such a scenario can be adequately represented by the following linear formulation

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{u} \tag{1}$$

where the vectors \mathbf{g} , \mathbf{f} and \mathbf{u} represent, respectively, the lexicographically ordered degraded image, the original image, and the disturbance. The known square matrix \mathbf{H} represents the linear distortion. \mathbf{H} is typically ill-conditioned. This implies that the exact solution, $\mathbf{f} = \mathbf{H}^{-1}\mathbf{g}$, to (1) is extremely contaminated by noise. Rather, a unique and stable estimate \mathbf{f}^* is sought by incorporating prior information on the original image. This has the effect of replacing the origi-

nal ill-conditioned problem with a well-conditioned one whose solution approximates that of the original. Such a technique is called a regularization method [23–26].

In the Bayesian image restoration method of interest here the prior information is quantified by specifying a probability density on \mathbf{f} and combining this with the information contained in \mathbf{g} to produce an estimate of the unknown image. We assume here a linear, additive white Gaussian noise model so that the probability density for \mathbf{g} is $P(\mathbf{g}|\mathbf{f}, \sigma) = \frac{1}{(2\pi\sigma^2)^{N^2/2}} \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 \right\}$ where $N^2 = 2^{2J}$ is the number of pixels in the image and σ^2 is the noise variance. If it so happens that the probability distribution for \mathbf{f} is in the form $P(\mathbf{f}|\boldsymbol{\theta}) \propto \exp \{ -\Phi(\mathbf{f}, \boldsymbol{\theta}) \}$ then by Bayes's rule, the MAP estimate, \mathbf{f}^* , is obtained by minimizing the following log-posterior density with respect to \mathbf{f} [27, 28]

$$L(\mathbf{f}, \sigma, \boldsymbol{\theta}) = \frac{1}{2\sigma^2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \Phi(\mathbf{f}, \boldsymbol{\theta}). \quad (2)$$

The function $\Phi(\mathbf{f}, \boldsymbol{\theta})$, called the *energy* function in the context of Bayesian estimation, is the energy attributed to the image \mathbf{f} , and $\boldsymbol{\theta}$ is the vector of possibly unknown model parameters. We give low energy to the images which coincide with our prior conceptions and high energy to those which do not. Thus, if our prior belief about the image is that the original image is smooth, then the energy is a measure of the *roughness*.

2.1 Wavelet Representation of Image Restoration Problem

In this paper, we adopt a wavelet domain approach to the image restoration problem. A comprehensive introduction to the wavelet theory can be found in [14, 15, 29]. It is possible to obtain the wavelet transform of images through a separable representation. This decomposition can be implemented by 1-D filtering of rows and columns of images. In Fig. 1, we have schematically illustrated a 1-level wavelet decomposition of an image $f(n, m)$ with $f_J^{(0)}(n, m)$ denoting the finest scale scaling coefficients. The 1-level wavelet decomposition of the image $f_J^{(0)}(n, m)$ produces four sub-images of size $2^{J-1} \times 2^{J-1}$, $f_{J-1}^{(k)}$, $k = 0, \dots, 3$. $f_{J-1}^{(0)}$ represents the scaling coefficients at scale

$J-1$ and $f_{J-1}^{(k)}$, $k = 1, \dots, 3$ are the wavelet coefficients at scale $J-1$ corresponding to the vertical, horizontal and diagonal orientations in the image plane. Multi-level wavelet decompositions of the image $f(n, m)$ can be obtained by applying the 1-level wavelet decomposition scheme, outlined above, recursively to the scaling coefficients $f_{J-1}^{(0)}(n, m)$. For an l -level wavelet decomposition, $j_0 = J - l$ denotes the lowest resolution at which the image is represented. We will use $\mathbf{f}_j^{(k)}$ to denote the vector of wavelet (scaling) coefficients obtained by lexicographically ordering the elements of the 2-D array $f_j^{(k)}(m, n)$ and $\hat{\mathbf{f}}$ to denote a lexicographically ordered version of all wavelet coefficients $\hat{f}(n, m)$.

With the conventions above, we can represent the problem in (1) in the wavelet domain as

$$\begin{aligned}\mathcal{W}\mathbf{g} &= (\mathcal{W}\mathbf{H}\mathcal{W}^T)\mathcal{W}\mathbf{f} + \mathcal{W}\mathbf{u} \\ \hat{\mathbf{g}} &= \hat{\mathbf{H}}\hat{\mathbf{f}} + \hat{\mathbf{u}},\end{aligned}\tag{3}$$

where \mathcal{W} is the 2-D wavelet transform matrix, $\hat{\mathbf{g}}$, $\hat{\mathbf{f}}$ and $\hat{\mathbf{u}}$ are the vectors holding the scaling and wavelet coefficients of the data, the original image, and the disturbance, $\hat{\mathbf{H}}$ is the wavelet domain representation of our linear degradation operator \mathbf{H} , and $\mathcal{W}^T\mathcal{W} = \mathbf{I}$ follows from the orthogonality of the wavelet transform. Note that since the wavelet transform is orthonormal \hat{u} is again Gaussian with zero mean and variance σ^2 .

3 A Multiscale Image Model

A key component of our image restoration algorithm is the use of a multiscale stochastic prior model for f . To motivate the particular choice of prior model used here, consider the wavelet coefficients of a typical image at a particular resolution. Wavelet coefficients are obtained by differentiation-like operations. Since the spatial structure of many images typically consists of smooth areas dispersed with occasional edges, the distribution of wavelet coefficients should be sharply peaked around zero, due to the contribution of smooth areas, and have broad tails repre-

senting the contribution of the edges [6].

Following the work in [6, 7] on image coding and denoising, we model the distribution of wavelet coefficients of images by a Generalized-Gaussian (GG) density [13, 14]

$$P\left(f_j^{(k)}(m, n) | p, \kappa_j^{(k)}\right) \propto \exp\left\{-\frac{1}{p} \left|\frac{f_j^{(k)}(m, n)}{\kappa_j^{(k)}}\right|^p\right\}, \quad (4)$$

where $0 \leq p \leq 2$ is a parameter which determines the tail behavior of the density function and $\kappa_j^{(k)}$ is a *scale parameter* similar to the standard deviation of a Gaussian density. We will refer to the zero mean density in (4) as $GG(0, \kappa_j^{(k)}, p)$. For $p = 1$ we have the Laplacian density and for $p = 2$ we have the familiar Gaussian density. The tails of the GG distribution becomes increasingly heavier as p approaches zero. We assume that the mean of the image is subtracted from the image and that the scaling coefficients $f_{j_0}^{(0)}(m, n)$, are i.i.d. $GG(0, \kappa_{j_0}^{(0)}, p)$.

The specification of one κ parameter for each scale and orientation results in an image model far too complex to be of use in a restoration procedure. Nonetheless, the structure of the model in (4) coupled with the specification of the problem in the wavelet domain does suggest a variety of simplifications which are of use for the restoration problem. In this work, we consider the following three models:

1. **Model 1:** The scaling coefficients $f_{j_0}^{(0)}(m, n)$, are i. i. d. with $GG(0, \kappa_{j_0}^{(0)}, p)$ and the wavelet coefficients are i. i. d. with exponentially decreasing variances, i.e. $f_j^{(k)}(m, n) \sim GG(0, \kappa 2^{-\alpha(j-j_0)}, p)$, $i = 1, 2, 3$, $j_0 \leq j \leq J - 1$ with j_0 the coarsest scale, κ the scale parameter corresponding to j_0 and $\alpha \geq 0$. The rationale behind this model is that it is equivalent to a deterministic modeling of the image as a member of a Besov space [15].
2. **Model 2:** The scaling coefficients $f_{j_0}^{(0)}(m, n)$, are i. i. d. with $GG(0, \kappa_{j_0}^{(0)}, p)$ and the wavelet coefficients at a particular scale are i. i. d. with $GG(0, \kappa_j, p)$, $j = j_0, \dots, J - 1$. This model is useful in cases where the variance of the wavelet coefficients at different scales cannot be

well-approximated by a simple exponential law.

3. **Model 3:** The scaling coefficients $f_{j_0}^{(0)}(m, n)$, are i. i. d. with $GG(0, \kappa_{j_0}^{(0)}, p)$ and the wavelet coefficients at different orientations (horizontal, vertical or diagonal) are distributed with $GG(0, \kappa^{(k)} 2^{-\alpha(j-j_0)}, p)$, $i = 1, 2, 3$, $j_0 \leq j \leq J - 1$. Such a model is most suitable for images with significantly different characteristics in different orientations as might arise in geophysical restoration problems involving layered structures.

We make several observations regarding these models. First, they are indeed of low dimensionality. In addition to the α and p parameters, Model 1 is characterized by two κ coefficients: one for the coarsest scale scaling coefficients and one multiplying the exponential for the wavelet coefficients. There are a total of $J - j_0 + 1$ κ 's for Model 2 and four κ values required to characterize Model 3. In subsequent sections, we shall see that the number of regularization parameters to be determined in the restoration algorithm is equal to the number of κ 's characterizing the prior model being used. Moreover, an appropriate on-line choice of the model parameters provides a mechanism for adapting the level of regularization in an image to the underlying scale-to-scale structure (Models 1 and 2) or to orientation-dependent structure (Model 3). While, the above three models certainly do not represent an exhaustive enumeration of all possible multiscale regularization approaches, as seen in Section 6, they do provide a strong indication as to the utility of this type of modeling technique for image restoration.

Finally, we comment on the estimation of the hyper-parameters, p , α , and $\kappa_j^{(k)}$. In a restoration algorithm, these parameters could be estimated from the data by assigning appropriate priors to each and maximizing the resulting log-posterior function with respect to the hyper-parameters and the image. However, such an approach presents many computational difficulties and unnecessarily complicates the problem. Instead, for the remainder of this paper we choose to simplify the problem

by fixing the p and α *a priori*. Generally, the performance of the regularizer is impacted to a greater extent by the on-line identification of the κ parameters [30] (or as explained in subsequent sections, quantities closely related to κ) so we concentrate our effort on identifying good choices of κ .

The issue of selecting an appropriate p is extensively discussed in Section 4.1. As for the selection of α , we propose using a fixed *a priori* choice obtained from the empirical study of a number of images. According to our findings, for most images representing natural scenes the α value which produced the best fit to the image data under the Model 2 scheme (for $p = 1$) fell between 0.6 and 1.6 with mean $\alpha \approx 1.2$. We evaluated the effects of varying the α value on a number of restoration problems and saw that the results were relatively insensitive to variations in α in the range suggested by the observations. The first example in Section 6 supports this. Note also that past research reveals similar conclusions [30] indicating that the performance of the estimator is degraded little by the error in α . Therefore, for all experiments performed we used $\alpha = 1.2$ as our fixed *a priori* choice.

4 A Multiscale Image Restoration Algorithm

The MAP estimate of the wavelet coefficients of the original image is found by maximizing the log-posterior function in (2). Substituting the prior probability density developed in Section 3 into (2), the MAP estimate of $\hat{\mathbf{f}}$ is seen to be the minimum of the following cost function with respect to $\hat{\mathbf{f}}$ (assuming for the time being that $\lambda_j^{(k)}$ is known)

$$J(\hat{\mathbf{f}}, \boldsymbol{\lambda}) = \|\hat{\mathbf{g}} - \hat{\mathbf{H}}\hat{\mathbf{f}}\|_2^2 + \lambda_{j_0}^{(0)} \|\mathbf{f}_{j_0}^{(0)}\|_p^p + \sum_{j=j_0}^{J-1} \sum_{i=1}^3 \lambda_j^{(k)} \|\mathbf{f}_j^{(k)}\|_p^p \quad (5)$$

where $\lambda_j^{(k)} = \frac{2\sigma^2}{p(\kappa_j^{(k)})^p}$ are weighting parameters and $\boldsymbol{\lambda} = [\lambda_{j_0}^{(0)}, \lambda_{j_0}^{(1)}, \dots, \lambda_{J-1}^{(3)}]^T$. The formulation in (5) easily accommodates the Model 1-3 regularization schemes discussed in Section 3 by defining the appropriate relationships for $\lambda_j^{(k)}$. For example, putting $\lambda_{j_0}^{(0)} = \lambda_1$ and $\lambda_j^{(k)} = \lambda_2 2^{\alpha(j-j_0)}$ results in the Model 1 regularization scheme while assigning a different λ_j to each scale in the wavelet

domain without regarding the orientation we obtain the Model 2 regularization scheme. Suppose that $J(\hat{\mathbf{f}}, \boldsymbol{\lambda})$ has a minimum in $\hat{\mathbf{f}}$, then at a stationary point $\hat{\mathbf{f}}^s$, the gradient of $J(\hat{\mathbf{f}}, \boldsymbol{\lambda})$ must vanish. Unfortunately, the l_p norm terms appearing in (5) is not differentiable for $p \leq 1$. Hence, we propose the following *smooth* approximation to the l_p norm, raised to the power p , as in [19]

$$\|\mathbf{x}\|_p^p \approx \sum_i \left((|x_i|^2 + \beta)^{p/2} - \beta^{p/2} \right) \quad (6)$$

where $\beta \geq 0$ is a stabilization constant and x_i denotes the i -th element of the vector \mathbf{x} . Substituting (6) into (5) and taking the gradient of the cost function we arrive at the following equation

$$\mathbf{D}^* = \text{diag} \left[\frac{\lambda_i}{(|\hat{f}_i^*|^2 + \beta)^{1-p/2}} \right]_{i=1}^{N^2} \quad (7)$$

$$\left(\hat{\mathbf{H}}^T \hat{\mathbf{H}} + \frac{p}{2} \mathbf{D}^* \right) \hat{\mathbf{f}}^* = \hat{\mathbf{H}}^T \hat{\mathbf{g}}, \quad (8)$$

where $\hat{\mathbf{f}}^*$ is the minimum of $J(\hat{\mathbf{f}}, \boldsymbol{\lambda})$ with the approximation in (6), \hat{f}_i^* is the i -th element of $\hat{\mathbf{f}}^*$ and λ_i is the associated regularization parameter. The above equation gives the first order conditions that must be satisfied by $\hat{\mathbf{f}}^*$. By direct analogy with the lagged diffusivity method of Vogel and Oman [19], we can develop a fixed point iteration to solve for $\hat{\mathbf{f}}^*$. Starting with an initial point $\hat{\mathbf{f}}^0$, we solve the following equation for $\hat{\mathbf{f}}^{k+1}$

$$\left(\hat{\mathbf{H}}^T \hat{\mathbf{H}} + \frac{p}{2} \mathbf{D}^k \right) \hat{\mathbf{f}}^{k+1} = \hat{\mathbf{H}}^T \hat{\mathbf{g}}, \quad (9)$$

where \mathbf{D}^k is obtained by replacing \hat{f}_i^* by \hat{f}_i^k in (7)¹. The iteration is terminated whenever $\frac{\|\hat{\mathbf{f}}^{k+1} - \hat{\mathbf{f}}^k\|}{\|\hat{\mathbf{f}}^k\|} < \gamma$, with γ being a small positive constant. We use 10^{-5} in our simulations. The fixed point iteration in (9) is a special case of the “half quadratic regularization” scheme introduced by Geman *et. al.* [3] and the ARTUR scheme due to Charbonnier *et. al.* [4]. Adopting the notation in [4] we define the following function

$$\phi(t) = (t^2 + \beta)^{\frac{p}{2}} - \beta^{\frac{p}{2}}. \quad (10)$$

¹The iteration index for the iterative optimization, k should not be confused here with the index (k) used to describe the orientations of the wavelet/scaling coefficients.

Then, the approximated cost function can be expressed in terms of the function $\phi(t)$. Furthermore, $\phi(t)$ satisfies the conditions (a)-(i) presented in [4](page 300, equation (12)). Roughly speaking, these conditions ensure that the $\phi(t)$ function applies less and less penalty as the magnitude of the wavelet coefficients increase (so that the large-magnitude wavelet coefficients, primarily associated with edges, are well-preserved) and that the restoration algorithm is convergent in the sense that the sequence $J(\hat{\mathbf{f}}^k, \boldsymbol{\lambda})$ is convergent and that $(\hat{\mathbf{f}}^{k+1} - \hat{\mathbf{f}}^k) \xrightarrow{k \rightarrow +\infty} \mathbf{0}$. In the special case where $\phi(t)$ is convex (which occurs if $p \geq 1$) and $\hat{\mathbf{H}}$ is full-rank, the iterates $\hat{\mathbf{f}}^k$ converge and the computed solution is the unique minimum of (5). However, when $p < 1$, $\phi(t)$ is concave and the algorithm computes a local minimum of (5) [4].

The iterative algorithm in (9) requires the solution of a very large linear matrix equation. Note that the matrix appearing on the right hand side of (9) is symmetric and positive definite. Therefore, the conjugate gradient (CG) algorithm [31] can be conveniently used to compute the solution $\hat{\mathbf{f}}^{k+1}$ in (9) at each step. In this way, the algorithm given in (9) is doubly iterative in that an outer iteration is used to update the solution $\hat{\mathbf{f}}^k$ and an inner iteration is used to solve the system of equations in (9) by the CG method. The special structure of the matrices $\hat{\mathbf{H}}$ and \mathbf{D}^k could be used to decrease the computational cost substantially. The first matrix, $\hat{\mathbf{H}}$ is merely the wavelet domain representation of our degradation operator. If the kernel is convolutional, it has been shown by Zervakis *et. al.* [32] that this matrix can be diagonalized by a special Fourier transform matrix by invoking the circulant assumption. On the other hand, the second matrix \mathbf{D} is diagonal in the wavelet domain. Therefore, the vector matrix multiplications required for the implementation of the CG algorithm can be computed in an efficient way by going back and forth between the wavelet and the Fourier transform domains. In this case, the cost of multiplying a vector with the matrix $\hat{\mathbf{H}}^T \hat{\mathbf{H}} + \frac{p}{2} \mathbf{D}^k$ is dominated by the cost of the FFT which is $O(N^2 \log N)$.

We note that the iterative algorithm in (9) can be efficient even in the case where $\hat{\mathbf{H}}$ is not convo-

lutional since the wavelet domain representation of a wide range of operators is sparse [18]. In those cases, standard techniques for sparse matrices can be used to reduce computational complexity.

4.1 Selection of Appropriate p and β values

The possibility that multiple local minima of (5) may exist for $p < 1$ presents an interesting trade-off. From a computational viewpoint, it is highly desirable to use $p \geq 1$, since in this case the cost function is convex and global convergence is guaranteed. However, based on empirical studies of the wavelet coefficients of images it has been shown that the GG model for the distribution of the wavelet coefficients usually corresponds to $p < 1$ and a typically recognized value is $p = 0.7$ [13]. Analysis of the use of GG priors in the context of image denoising has been performed by Moulin and Liu [33]. The results in [33] suggest that only modest improvement can be achieved by using $p < 1$ as compared with $p = 1$. In our experiments, we essentially arrived at the same conclusion. That is, the restored images obtained by using the best value of p (in terms of model fit) were visually almost the same as the results obtained by using $p = 1$, although slightly lower estimation errors were observed for $p < 1$. Therefore, we propose using $p = 1$ as the fixed *a priori* choice for the shape parameter of the GG distribution. Note that we do not claim that $p = 1$ is the right value for all types of images. Rather, we are saying that the estimation of p directly from the data is a complicated problem and in the absence of accurate prior information on p , $p = 1$ provides strong restoration results with guaranteed global convergence properties.

The role of the parameter β is two-fold. First it controls how close the approximation in (6) is to the original l_p norm. Using a relatively small β provides better restoration of edges in the image since a smaller β value provides better approximation to the l_p norm. Second, it essentially determines the convergence speed of the algorithm. While we do not intend to carry out a numerical analysis of the fixed point iteration in (9), the basic reason is that for $\beta = 0$, $\phi(t)$ in (10) is not differentiable at $t = 0$ and instability in the numerical computations may arise. If β

is relatively large, the algorithm is fast, and the convergence speed deteriorates as β gets smaller. Therefore, β should be set so as to achieve a compromise between the convergence speed and the edge preservation. Based on our experience on natural scenes, we found that restorations obtained for $\beta \approx 1$ were visually indistinguishable from the restorations obtained for $\beta \ll 1$. We note that a similar value is recommended for the TV algorithm [19].

5 Regularization Parameter Selection

In this paper, we use a multi-dimensional extension of the L-curve method [20], called the L-hypersurface method [34], to determine $\boldsymbol{\lambda}$ in (5). In order to describe the method thoroughly, we consider the following generalized image restoration scheme where the estimate of the original image f is obtained by minimizing the following cost function

$$J(\mathbf{f}, \boldsymbol{\lambda}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \sum_{j=1}^M \lambda_j \|\mathbf{R}_j \mathbf{f}\|_p^p \quad (11)$$

where λ_j , $j = 1, \dots, M$ are the regularization parameters and \mathbf{R}_j are the corresponding regularization operators. The cost function in (11) represents a multiply constrained least squares problem and includes many popular image restoration schemes as its special cases. Our wavelet domain image restoration algorithm is obtained if \mathbf{g} , \mathbf{H} and \mathbf{f} are in the wavelet domain and \mathbf{R}_j , $j = 1, \dots, M$ are the operators which extract desired portions of the wavelet transform of \mathbf{f} . For example, we can take \mathbf{R}_1 as the operator extracting the coarsest scale scaling coefficients and \mathbf{R}_2 as the operator extracting the wavelet coefficients for a doubly constrained, Model 1-type problem.

To extend the L-curve, we first introduce the following quantities

$$\mathbf{f}^*(\boldsymbol{\lambda}) = \arg \min_{\mathbf{f}} J(\mathbf{f}, \boldsymbol{\lambda}) \quad (12)$$

$$z(\boldsymbol{\lambda}) = \log \|\mathbf{g} - \mathbf{H}\mathbf{f}^*(\boldsymbol{\lambda})\|_2^2, \quad x_j(\boldsymbol{\lambda}) = \log \|\mathbf{R}_j \mathbf{f}^*(\boldsymbol{\lambda})\|_p^p, \quad j = 1, \dots, M \quad (13)$$

With the above definitions, the “L-hypersurface” [34] is defined as a subset of \mathcal{R}^{M+1} associated

with the map $s(\boldsymbol{\lambda}) : \mathcal{R}_+^M \rightarrow \mathcal{R}^{M+1}$, such that

$$s(\boldsymbol{\lambda}) = (x_1(\boldsymbol{\lambda}), \dots, x_M(\boldsymbol{\lambda}), z(\boldsymbol{\lambda})) \quad (14)$$

For a single constraint, the L-hypersurface reduces to the conventional L-curve which is simply a plot of the residual norm versus the norm of the restored image in a doubly logarithmic scale for a set of admissible regularization parameters. In this way, the L-curve displays the compromise between the minimization of these two quantities. It has been argued and numerically shown that the so called “corner” of the L-curve corresponds to a point where regularization and perturbation errors are approximately balanced [20].

Analogous to the one dimensional case, the L-hypersurface is a plot of the residual norm $z(\boldsymbol{\lambda})$ against the constraint norms $x_j(\boldsymbol{\lambda})$, $1 \leq j \leq M$. Intuitively, the “generalized corner” of the L-hypersurface should correspond to a point where regularization errors and perturbation errors are approximately balanced. By a generalized corner, we mean a point on the surface around which the surface is maximally warped. We can measure how much a surface is warped around a point by computing the Gaussian curvature [34]. In Fig. 2 we plot a typical L-hypersurface along with its Gaussian curvature and the error between the original and the restored images for a range of regularization parameters. The experiment for which the L-hypersurface was computed was the restoration of a 32×32 image degraded by a Gaussian blur of variance 1 pixel and corrupted by white Gaussian noise at 30dB SNR. We used our multiscale algorithm with $p = 2.0$ and Model 1 regularization scheme as shown in Fig. 3(a). Figure 2(b) shows the curvature of the L-hypersurface shown in Fig. 2(a) with λ_1 (resp. λ_2) being the regularization parameter for the scaling (resp. wavelet) coefficients. Figure 2(b)-(c) clearly indicate the usefulness of the Gaussian curvature plot in assessing the goodness of regularization parameters. It is observed that the curvature is significant along an extended maxima which is very close to a region in the error norm, $\|\mathbf{f} - \mathbf{f}^*(\boldsymbol{\lambda})\|_2^2$, plot in Fig. 2(c) where the error between the actual and the restored images is minimized. Moreover, the

curvature plot indicates that there is in fact more than one good regularization parameter for the scaling coefficients, and as long as we choose the correct value for the regularization parameter corresponding to the wavelet coefficients the restorations should have approximately the same quality. The error norm plot in Fig. 2(c) supports this point of view.

For the numerical experiments described in the following section, the regularization parameters are selected by searching over a grid of parameter values in λ space and choosing that point whose curvature is maximum. The computational complexity of this technique is clearly quite high. The major difficulty here is that the curvature of the L-hypersurface possesses many maxima/minima as seen in Figure 5(c) and therefore the use of well-known optimization techniques such as the Gauss-Newton method would fail. As the primary issue of interest here is in demonstrating that there is utility to the L-hypersurface method, we leave the considerable effort of finding a more efficient implementation to future work.

6 Experimental Results

In this section, we illustrate the performance of our proposed multiscale image restoration algorithm. All computations were carried out by using the Matlab commercial software package with double precision arithmetic. We used the routines in Donoho's Wavelab toolbox [35] for the computation of forward and inverse wavelet transforms with Daubechies' eight tap most symmetrical wavelets [29]. In all cases below, we limited the number of levels of wavelet decomposition to 3. In the first example, we used a Gaussian convolutional kernel, $h(x, y) = \frac{1}{4\sigma_x\sigma_y} \exp\{-\frac{x^2+y^2}{2\sigma_x\sigma_y}\}$, with $\sigma_x = \sigma_y = 2.0$ to blur the 256×256 Mandrill image. Zero mean white Gaussian noise was added to set the SNR to 30dB. In Fig. 4 (a)-(b) we display the original and the blurred, noisy images.

We restored the degraded Mandrill image using three regularization techniques: our proposed multiscale regularization scheme, the Constrained Least Squares (CLS) algorithm with a 2-D Lapla-

cian regularizer [2], and the TV algorithm. The CLS and the TV algorithms are special cases of the generalized image restoration scheme in (11) for $j = 1$ in which taking $p = 2$, and $\mathbf{R}_1 = \Delta$ (i.e. 2-D Laplacian) results in the CLS cost function and $p = 1$ and $\mathbf{R}_1 = \nabla$ gives the cost function corresponding to the TV algorithm. The action of the 2-D Laplacian operator Δf at the pixel (m, n) is $\Delta f(m, n) = -4f(m, n) + f(m, n + 1) + f(m + 1, n) + f(m, n - 1) + f(m - 1, n)$ and the action of the gradient operator ∇f is given by $\nabla f(m, n) = [\nabla_x f(m, n), \nabla_y f(m, n)]^T$ with $\nabla_x f(m, n) = -f(m, n) + f(m + 1, n)$ and $\nabla_y f(m, n) = -f(m, n) + f(m, n + 1)$. Both operators are implemented by circularly wrapping the image at the boundaries. The relevant regularization parameters were determined using the L-curve or the L-hypersurface method. For the TV algorithm and our algorithm we used $\beta = 1.0$. Experimental results obtained for $\beta = 10^{-10}$ indicate that smaller choices of β do not improve the visual quality of the restorations (see Fig. 4(g)-(h) and Fig. 7(e)-(f)).

In Fig. 4(c)-(h) we display the restored Mandrill images corresponding to the CLS, the TV and the multiscale algorithm. For our multiscale image restoration method we computed four restorations, displayed in Fig. 4(e)-(h), according to the Model 1 and Model 2 regularization schemes described in Section 3. Figure 4 shows that both the TV algorithm and our algorithm produce restored images visually superior to the CLS algorithm. We also observe that the images restored by our algorithm are a little sharper than the image restored by the TV algorithm and that the texture-like regions abundant in the Mandrill image (e.g. the hairs around the mouth of the Mandrill) are better recovered by our algorithm. The Root Mean Square Error (RMSE), $\sqrt{\frac{1}{N^2} \|\mathbf{f} - \mathbf{f}^*(\boldsymbol{\lambda})\|_2^2}$, between the original and restored images are listed in Table 1.

For the Model 1 restoration in Fig. 4(e) the L-hypersurface was used to determine two parameters, λ_1 and λ_2 corresponding to the coarsest scale scaling coefficients and the wavelet coefficients respectively as shown in Fig. 3(a). In this case, the curvature of the L-hypersurface is a 2-D function

of the regularization parameters as seen in Fig. 5(a). Also shown in Fig. 5(b) is a plot of RMSE as a function of these regularization parameters. Examining these plots shows that the curvature surface has a distinct extended maxima along which the norm of the error is very close to being a minimum. Thus, we see that the restoration algorithm is not overly sensitive to the scaling coefficient regularization parameter and locating the correct regularization parameter for the wavelet coefficients is more important.

In the Model 2 restoration in Fig. 4(f), each scale in the wavelet domain is assigned a different regularization parameter as seen in Fig. 3(b). Based on the L-hypersurface obtained for the Model 1 restoration in Fig. 5(a), we set the scaling coefficient regularization parameter to 10^{-5} . Figure 6(a)-(c) shows the curvature of the L-hypersurface obtained for this experiment. Since in this case the curvature is a 3-D function (one parameter for each wavelet scale), each of the 2-D plots in Fig. 6(a)-(c) is actually a slice of the curvature hypersurface with the regularization parameter corresponding to the coarsest scale being constant. Again, the maxima of the curvature of the L-hypersurface track well the minima of the RMSE surface so that we are close to the “optimal” regularization parameters. We see little difference either in terms of the error norm or in terms of visual quality between the Model 1 and Model 2 restorations in Fig. 4(e) and (f). This example verifies the primary assumption of Model 1 scheme where it was assumed that the variance of the wavelet coefficients decrease uniformly across scales according to an exponential law.

Finally, in Fig. 4(g)-(h) we display the Model 1 restorations corresponding to an idealized case where the parameters α and p were estimated directly from the original image. Clearly, this is not a realistic situation since in practice the original image is not available. Nonetheless, this example is interesting since it gives us an idea about how much improvement can be expected when using the optimal α and p values as opposed to fixed *a priori* choices $\alpha = 1.2$ and $p = 1.0$. The optimal p was estimated by using the method proposed in [14] and was found to be $p_{opt} = 0.7280$. The

exponential parameter α_{opt} was estimated by computing the slope of the line fitted to the $\log \kappa_j$ for $j = j_0, \dots, J - 1$. It was found to be $\alpha_{opt} = 0.6117$. Since p_{opt} yields a non-convex optimization task, we computed the restorations for this case in 2 stages. The first stage starts with computing the restoration for $p = 1.0$, which is unique and then the restored image for $p = 1.0$ is fed as the starting point to the restoration algorithm with $p = 0.7280$. There is no guarantee that the restored image for p_{opt} corresponds to the global minimum of the cost function, nevertheless we obtained good results with this scheme. Figure 5(c)-(d) shows the L-surface and the RMSE surface for p_{opt} , respectively. Figure 4(g)-(h) are the restorations obtained for this case. Figure 4(g)-(h) differs only in the value used for β which was set to 1.0 in (g) and 10^{-10} in (h). As already pointed out the two restorations are visually indistinguishable, though the convergence of the algorithm took significantly longer for $\beta = 10^{-10}$. Finally, comparison of Fig. 4(g)-(h) with Fig. 4(e)-(f) reveals that there is visually little difference between the restored images corresponding to p_{opt} and $p = 1.0$ cases. This example shows that using $p < 1$ does not yield a significant improvement in the performance of the multiscale algorithm.

In our second example, we first blurred the original Bridge image in Fig. 7(a) with a 9×9 uniform motion blur and added white Gaussian noise to the degraded image to set the SNR at 40dB. The blurred image obtained by this way is shown in Fig. 7(b). Having established the edge preserving utility of the TV and the proposed algorithm over the conventional CLS method, we only display the restorations obtained by the TV and the proposed algorithm in Fig. 7(c)-(f). For our multiscale algorithm, we applied the Model 1 and Model 2 regularization schemes with the L-hypersurface choice of regularization parameters. As in the previous example, we determined 2 regularization parameters corresponding to the scaling and the wavelet coefficients for Model 1 and 3 parameters corresponding to the wavelet coefficients at each scale for Model 2. In Model 2 restoration, the regularization parameter for the scaling coefficients were set to 10^{-4} from Fig. 8(a).

Figure 7(d) shows the Model 1 restoration and Fig. 7(e) and (f) show the Model 2 restorations. In Fig. 7(e)-(f) all parameters except for β are the same ($\beta = 1.0$ in (e) and $\beta = 10^{-10}$ in (f)).

Although the RMSE values in Table 1 were similar, the restored images in Fig. 7(c)-(d) exhibit vastly different visual characteristics. The TV algorithm fails to recover many of the small features in the image and produces an overly homogenized restoration resembling an “oil painting” of the original scene. The multiscale algorithm is able to reproduce finer detail thereby yielding a more visually appealing restoration. As in the previous example, we see little difference in terms of the visual quality between the Model 1 and Model 2 restorations. Note that we used the same value $\alpha = 1.2$ in both Mandrill and Bridge examples regardless of the image considered.

In our final example, we demonstrate the orientation adaptive nature of our approach. In Fig. 9 (a), we display an artificial 32×32 image which has significant structure in the horizontal direction, but little in the vertical and diagonal directions. This image was blurred by a Gaussian convolutional kernel with $\sigma_x = \sigma_y = 1$, and zero mean white Gaussian noise was added to set the SNR at 30dB. Because of the large differences between the structure in the horizontal and vertical directions, an ideal image restoration algorithm should use different regularization parameters for vertical, horizontal and diagonal directions. With this in mind, in Fig. 9(c)-(d) we display the restorations obtained using Model 1 and Model 3 schemes which require three regularization parameters, λ_1 , λ_2 and λ_3 , as displayed in Fig. 3(b)-(c), respectively. The L-hypersurface was employed to determine the required regularization parameters. For both Model 1 and Model 3 schemes we set the scaling coefficient regularization parameter to 10^{-5} . For the Model 3 restoration, the regularization parameters obtained for the vertical and diagonal orientations (in which the image is constant) were approximately two orders of magnitude larger than the regularization parameter obtained for the horizontal orientation. It is clear from Fig. 9(c)-(d) that the orientation adaptive algorithm produces a much better restoration than the scale adaptive algorithm.

7 Conclusions

In this paper, we introduced a wavelet domain multiscale image restoration algorithm for use in linear image restoration problems. Following the recent results in the area of image denoising and coding, we developed a statistical prior model for the wavelet coefficients of images. Our priors are able to capture spatial, scale and orientational characteristics of images accurately. We developed a half-quadratic algorithm to solve the nonlinear optimization problem resulting from using such priors and utilized the L-hypersurface method for choosing the required regularization parameters. Experimental results show that our algorithm can produce restorations which are visually significantly better than that of the traditional techniques and at least comparable, if not better, than that of the edge-preserving algorithms.

Acknowledgments The authors would like to thank the anonymous reviewers for their comments and suggestions which have greatly improved the quality of this paper.

References

- [1] R. L. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images*, Kluwer Academic Publishers, Boston, 1991.
- [2] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*, Prentice Hall, Englewood Cliffs, NJ, 1977.
- [3] D. Geman and C. Yang, “Nonlinear image recovery with half-quadratic regularization,” *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, July 1995.
- [4] P. Charbonnier, G. Aubert L. Blanc-Feraud, and M. Barlaud, “Stochastic relaxation, gibbs distribution, and the bayesian restoration of images,” *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, February 1997.

- [5] S. Osher L. I. Rudin and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Phys. D*, vol. 60, pp. 259–268, 1992.
- [6] R. W. Buccigrossi and E. P. Simoncelli, “Image compression via joint statistical characterization in the wavelet domain,” Tech. Rep. 414, GRASP Lab. University of Pennsylvania, May 1997.
- [7] E. P. Simoncelli and E. Adelson, “Noise removal via bayesian wavelet coring,” in *Proceedings of the 1996 IEEE Int. Conf. on Image Proc.*, Lausanne, Switzerland, September 1996, vol. 1, pp. 379–382.
- [8] A. Chambolle, R. E. DeVore, N. Lee, and B. J. Lucier, “Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage,” *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 320–335, March 1998.
- [9] Eric L. Miller and Alan S. Willsky, “A multiscale approach to sensor fusion and the solution of linear inverse problems,” *Appl. Comput. Harmon. Anal.*, vol. 2, pp. 127–147, 1995.
- [10] Eric L. Miller and Alan S. Willsky, “Multiscale, statistically-based inversion scheme for the linearized inverse scattering problem,” *IEEE Trans. Geosc. Remote Sensing*, vol. 34, no. 2, pp. 346–357, March 1996.
- [11] M. R. Banham and A. K. Katsaggelos, “Spatially adaptive wavelet-based multiscale image restoration,” *IEEE Trans. Image Process.*, vol. 5, no. 4, pp. 619–633, April 1996.
- [12] G. Wang, J. Zhang, and G. W. Pan, “Solution of inverse problems in image processing by wavelet expansions,” *IEEE Transactions Image Process.*, vol. 4, no. 5, pp. 579–593, 1995.
- [13] P. Mathieu M. Antonini, M. Barlaud and I. Daubechies, “Image coding using wavelet transform,” *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 205–220, February 1992.

- [14] S. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. July, pp. 674–693, 1989.
- [15] Y. Meyer, *Wavelets and Operators*, Cambridge Univ. Press, New York, 1992.
- [16] R. A. DeVore, B. Jawerth, and B. J. Lucier, “Image compression through wavelet transform coding,” *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 719–747, March 1992.
- [17] D. T. Vuza U. Amato, “Besov regularization, thresholding and wavelets for smoothing data,” *Numer. Funct. Anal. Optim.*, vol. 18, no. 5&6, pp. 461–493, 1997.
- [18] D. L. Donoho, “Unconditional bases are optimal bases for data compression and for statistical estimation,” *Appl. Comput. Harmon. Anal.*, vol. 1, no. 1, pp. 100–115, December 1993.
- [19] C. R. Vogel and M. E. Oman, “Fast, robust total variation-based reconstruction of noisy, blurred images,” *IEEE Trans. Image Process.*, vol. 7, no. 7, pp. 813–824, July 1998.
- [20] P. C. Hansen, “Analysis of discrete ill-posed problems by means of the l-curve,” *SIAM Rev.*, vol. 34, pp. 561–580, 1992.
- [21] C. R. Vogel, “Non-convergence of the l-curve regularization parameter selection method,” *Inverse Problems*, vol. 12, no. 4, pp. 535–547, 1996.
- [22] M. Hanke, “Limitations of the l-curve method in ill-posed problems,” *BIT*, vol. 36, no. 2, pp. 287–301, 1996.
- [23] A. Tikhonov and V. Arsenin, *Solution of Ill-Posed Problems*, Wiley, New York, 1977.
- [24] M. Bertero, T. A. Poggio, and V. Torre, “Ill-posed problems in early vision,” *Proc. IEEE.*, vol. 76, no. 8, pp. 869–889, 1988.

- [25] M. Bertero, “Regularization methods for linear inverse problems,” in *Inverse Problems*, G. Talenti, Ed., vol. 1225 of *Lecture Notes in Mathematics*, pp. 52–112. Springer-Verlag, 1986.
- [26] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, vol. 375 of *Mathematics and its applications*, Kluwer Academic Publishers, Dordrecht (Holland), 1996.
- [27] G. Archer and D. M. Titterton, “On some bayesian regularization methods for image restoration,” *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 989–995, July 1995.
- [28] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, NewYork, 1985.
- [29] I. Daubechies, *Ten Lectures on Wavelets*, SIAM Press, New York, 1992.
- [30] R. Dufour and E. L. Miller, “Statistical signal restoration with wavelet domain prior models,” to appear in *Signal Process.*, manuscript available at <http://claudius.cdsp.neu.edu/nuwiirl/publications.html>
- [31] G. H. Golub and C. E. Van Loan, *Matrix Computations, 2nd Edition*, John Hopkins University Press, Baltimore, 1989.
- [32] M. E. Zervakis, T. M. Kwon, and J-S. Yang, “Multiresolution image restoration in the wavelet domain,” *IEEE Trans. Circuits Systems-II: Analog Digit. Sig. Proc.*, vol. 42, no. 9, pp. 578–591, September 1995.
- [33] P. Moulin and J. Liu, “Analysis of multiresolution image denoising schemes using generalized-gaussian and complexity priors,” *IEEE Trans. Inform. Theory*, vol. 38, no. 4, pp. 719–747, April 1998.

- [34] M. Belge, M. E. Kilmer, and E. L. Miller, “Simultaneous multiple regularization parameter selection by means of the l-hypersurface with applications to linear inverse problems posed in the wavelet domain,” in *Proceedings of SPIE’98–Bayesian inference for inverse problems*, July 1998, vol. 3459.
- [35] J. Buckheit and D. Donoho, “Wavelab and reproducible research,” in *Wavelets and statistics*, A. Antoniadis and G. Oppenheim, Eds., pp. 55–81. Springer-Verlag, 1995.

8 Figure and Table Captions

Table 1 RMSE values for each experiment.

Figure 1 Wavelet decomposition of an image.

Figure 2 (a) L-hypersurface, (b) Gaussian curvature of the L-hypersurface in (a), (c) the norm of the difference between the actual and restored images. λ_1 regularizes the coarsest scale scaling coefficients and λ_2 is used to penalize the wavelet coefficients in a Model 1 regularization scheme.

Figure 3 (a) Model 1 (b) Model 2 (c) Model 3 regularization schemes as used in our experiments. In each Model, the required regularization parameters, λ_j , are selected by the L-hypersurface method. In Model 1 and 3 α is set to 1.2 *a priori*.

Figure 4 (a) Original Mandrill image. (b) Blurred image, 30dB SNR. (c) Restored by the CLS algorithm. (d) Restored by the TV algorithm. (e) Restored by the proposed algorithm using Model 1 regularization scheme with $p = 1.0$ and $\alpha = 1.2$ (f) Restored by the proposed algorithm using Model 2 regularization scheme with $p = 1.0$ (g) Model 2 restoration with optimal α and p . (h) Model 2 restoration with the same parameters as in (g) except that $\beta = 10^{-10}$.

Figure 5 (a) Curvature of the L-hypersurface for Model 1 with $\alpha = 1.2$ and $p = 1.0$ and (b) the corresponding RMSE surface (c) Curvature of the L-hypersurface for Model 1 with $\alpha = 0.46$ and $p = 0.7280$ (optimal parameters) and (d) the corresponding RMSE surface (e) L-curve for the CLS algorithm and (f) corresponding RMSE curve. (g) Curvature of the L-curve for the TV algorithm and (h) corresponding RMSE curve.

Figure 6 (a)-(c) Curvature of the L-hypersurface and (d)-(f) RMSE plots for the mandrill experiment.

Figure 7 (a) Original Bridge image. (b) Blurred image, 40dB SNR. (c) Restored by the TV algorithm. (d) Restored by the proposed algorithm using (d) Model 1 regularization scheme (e) Model 2 regularization scheme. (f) Model 2 restoration with the same parameters as in (e) but with $\beta = 10^{-10}$.

Figure 8 (a) Curvature of the L-hypersurface for the proposed algorithm. (b) Curvature of the L-curve for the TV algorithm. (c)-(d) Corresponding RMSE plots.

Figure 9 (a) Original image. (b) Blurred image, 30dB SNR. (c) Restored by the proposed algorithm with Model 2 (scale adaptive) regularization. (d) Restored by the proposed algorithm with Model 3 (orientation adaptive) regularization.

	CLS	TV	Model 1	Model 2	Model 3	Model 1 with best p & α
Mandrill	24.11	23.73	23.75	23.86	-	23.71
Bridge	-	19.43	20.31	20.13	-	-
Strips	-	-	-	11.96	9.17	-

Table 1:

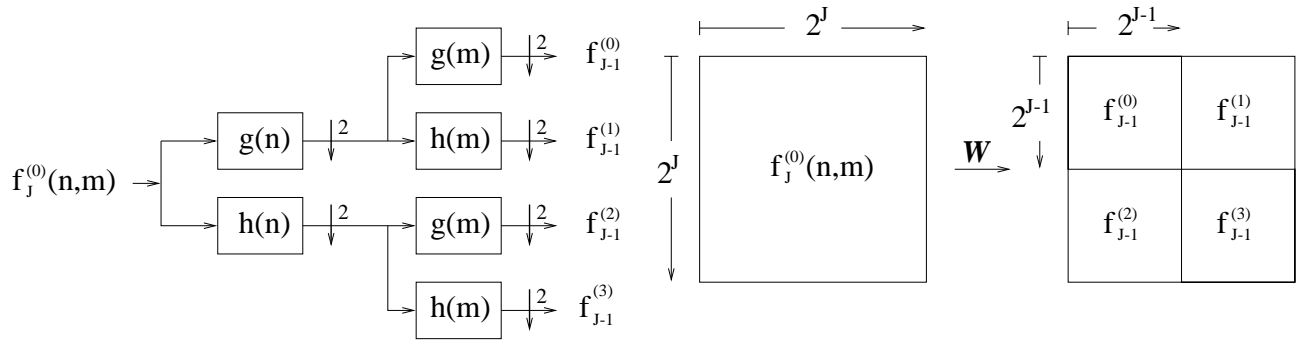


Figure 1:

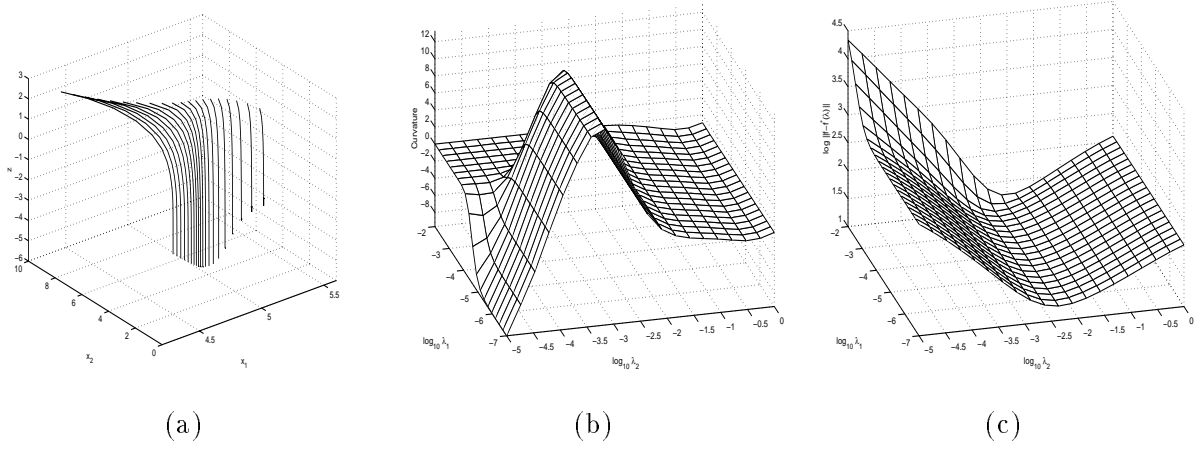


Figure 2:

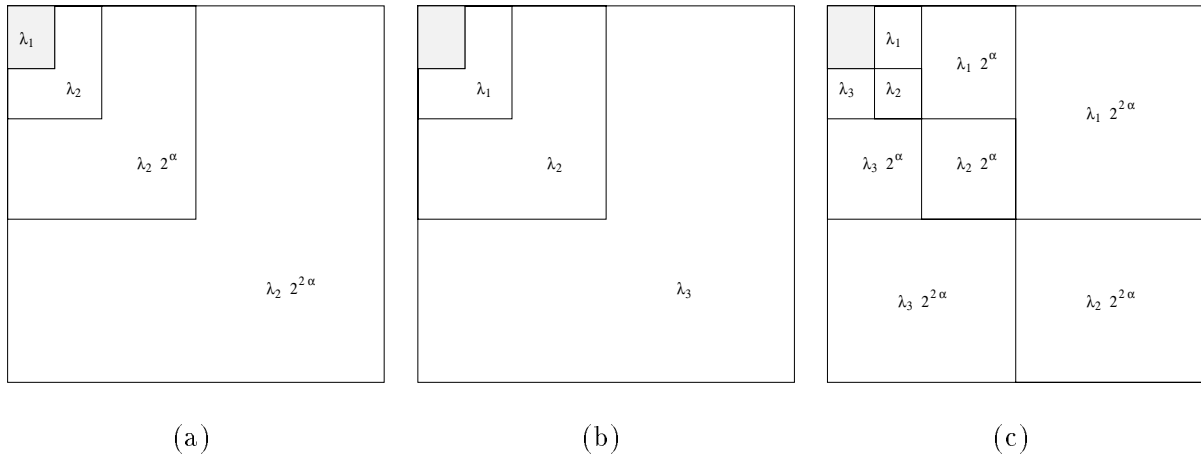
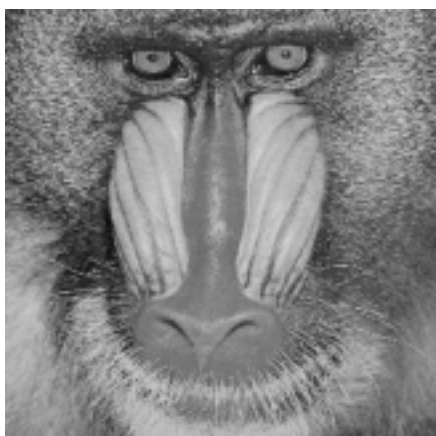


Figure 3:



(a)



(b)



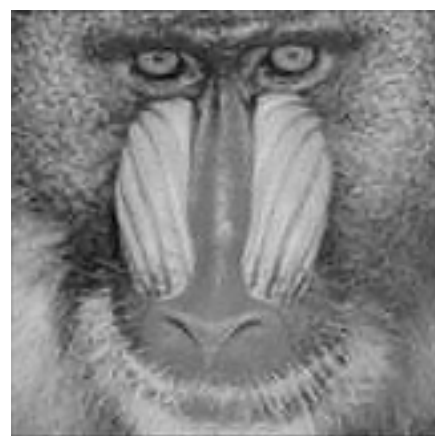
(c)



(d)



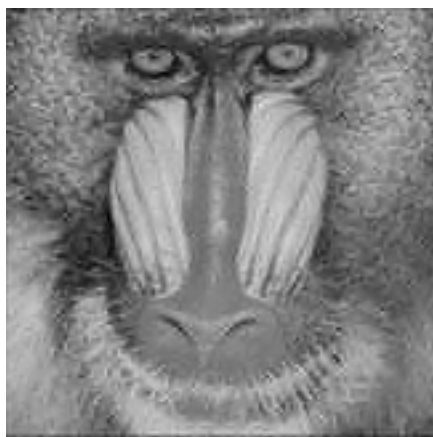
(e)



(f)

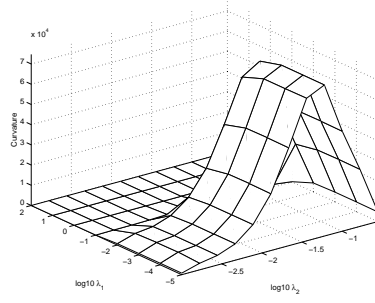


(g)

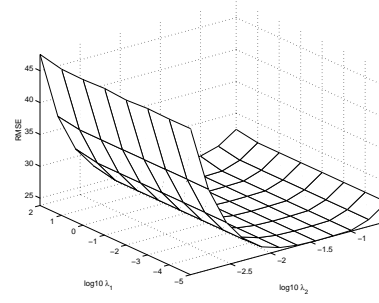


(h)

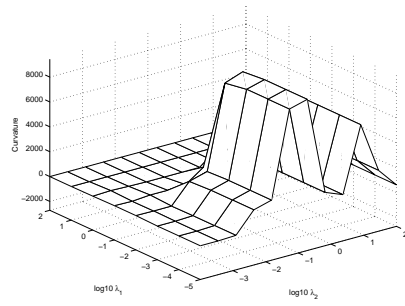
Figure 4:
28



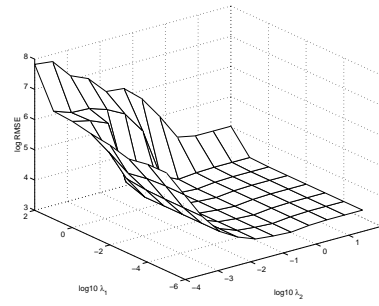
(a)



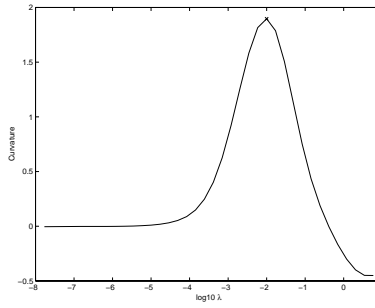
(b)



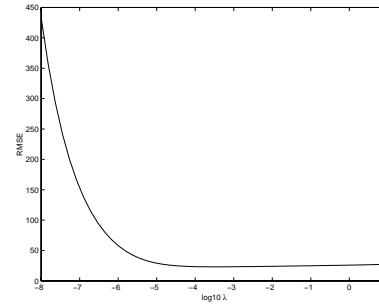
(c)



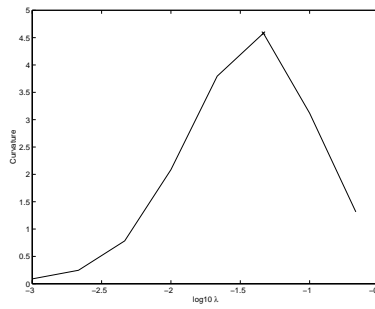
(d)



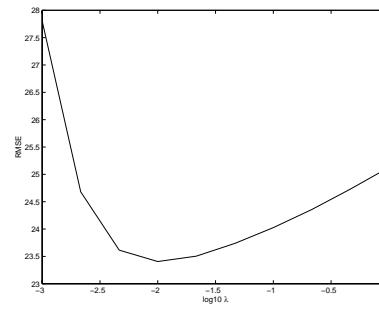
(e)



(f)

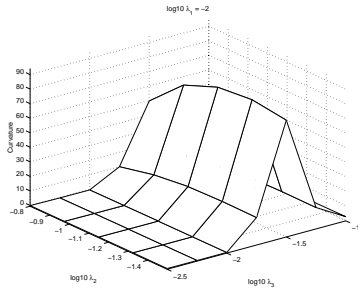


(g)

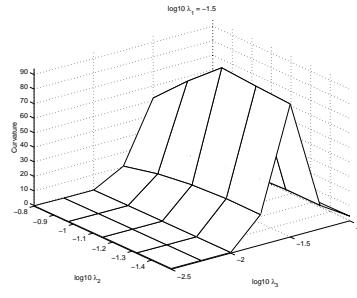


(h)

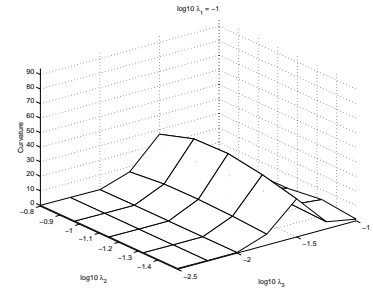
Figure 5:
29



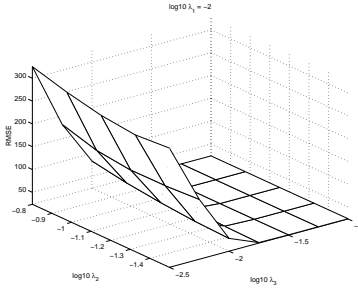
(a)



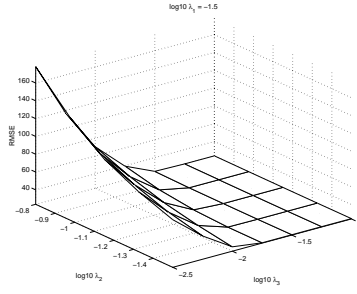
(b)



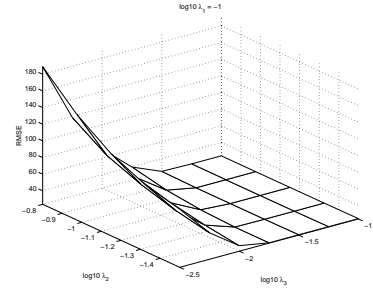
(c)



(d)



(e)



(f)

Figure 6:



(a)



(b)



(c)



(d)



(e)



(f)

Figure 7:

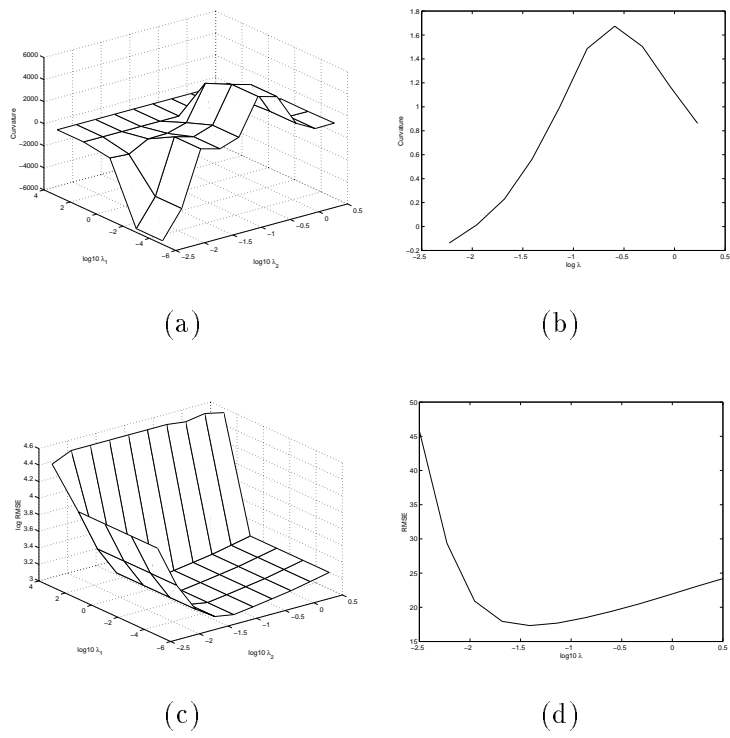


Figure 8:

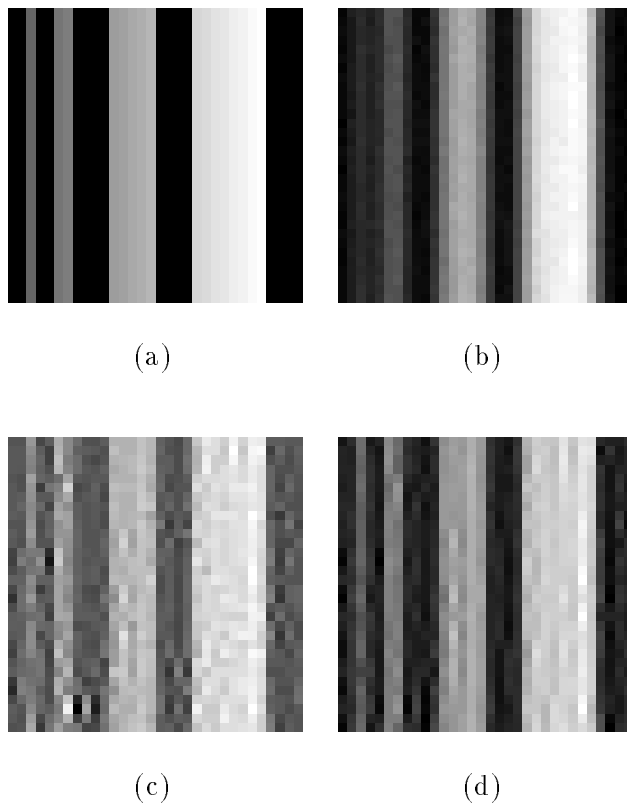


Figure 9:

FDFD Modeling of Plane Wave Interactions with Buried Objects Under Rough Surfaces



**Carey M. Rappaport and Ann W. Morgenthaler,
Northeastern University**

**Misha E. Kilmer
Tufts University**

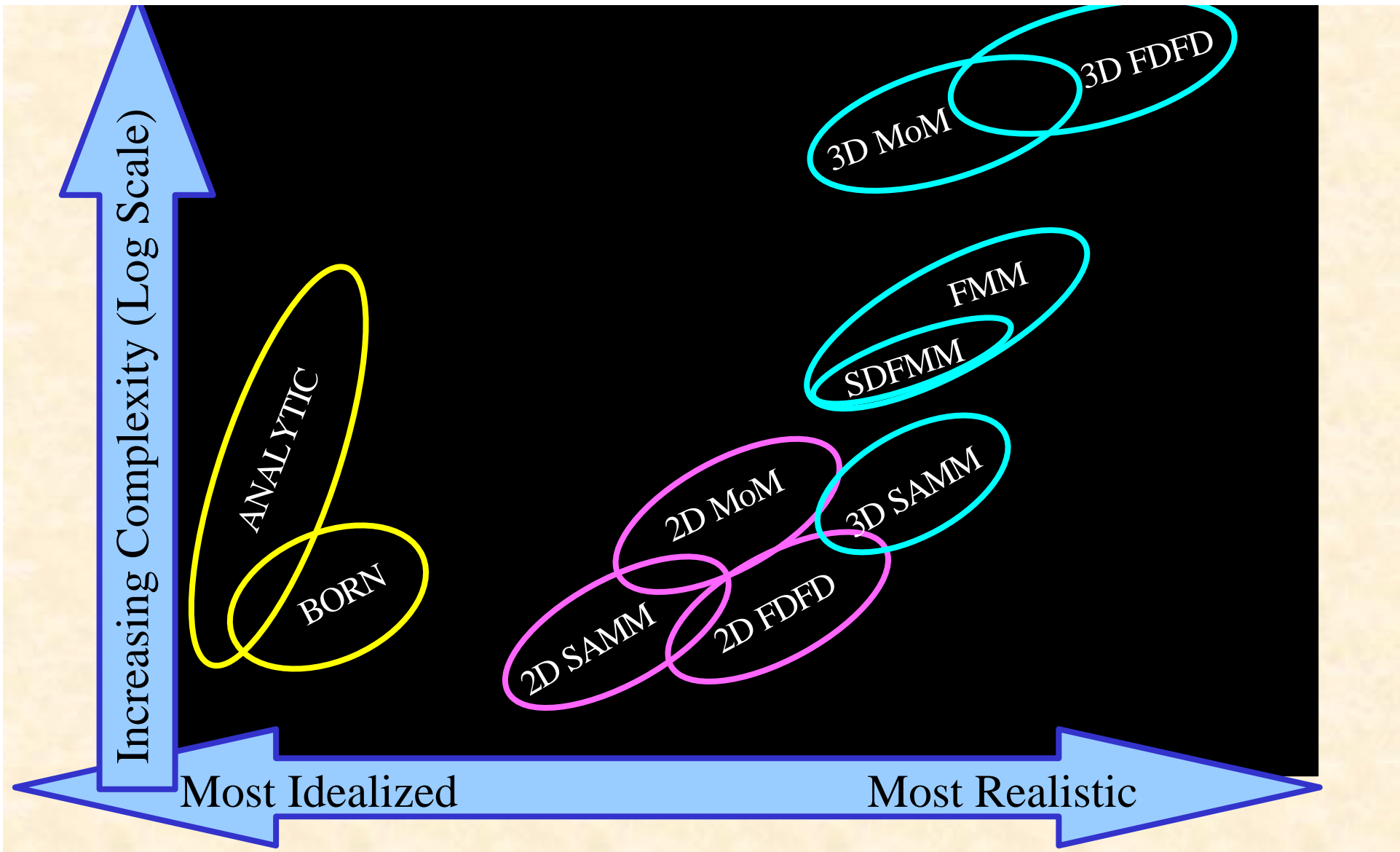
IEEE AP-S/URSI, July 11, 2001

Outline

- **Accurate modeling of subsurface wave scattering**
- **Plane wave FDFD computation with rough ground surfaces**
- **Method validation**
- **Computational examples for mine detection**

Supported by CenSSIS, the Center for Subsurface Sensing and Imaging Systems, under the ERC program of the NSF (EEC 9986821)

Frequency Domain Computational Electromagnetic Models



Numerically Modeling EM Wave Propagation in Soil to Aid Sensor Design

- Only with knowledge of the environment surrounding mines can efficient advanced sensors be developed.
- Soil environment is extremely varied and inhomogeneous.
- Propagation is quite different in different soils.
- Numerical modeling is the only viable means of testing candidate sensing concepts on the wide variety of conditions.
- Simple flexible models that capture the essential electrical characteristics are best.
- Visualization of wave propagation assists in sensor synthesis.

Finite Difference Modeling

Time Domain

- Finite Difference Time Domain (**FDTD**) is fast, accurate, easy to implement, and intuitive.
- FDTD is best for wideband (short pulse) scattering in non-uniform media.
- Special means have been developed for **Dispersive Media**, such as soil.

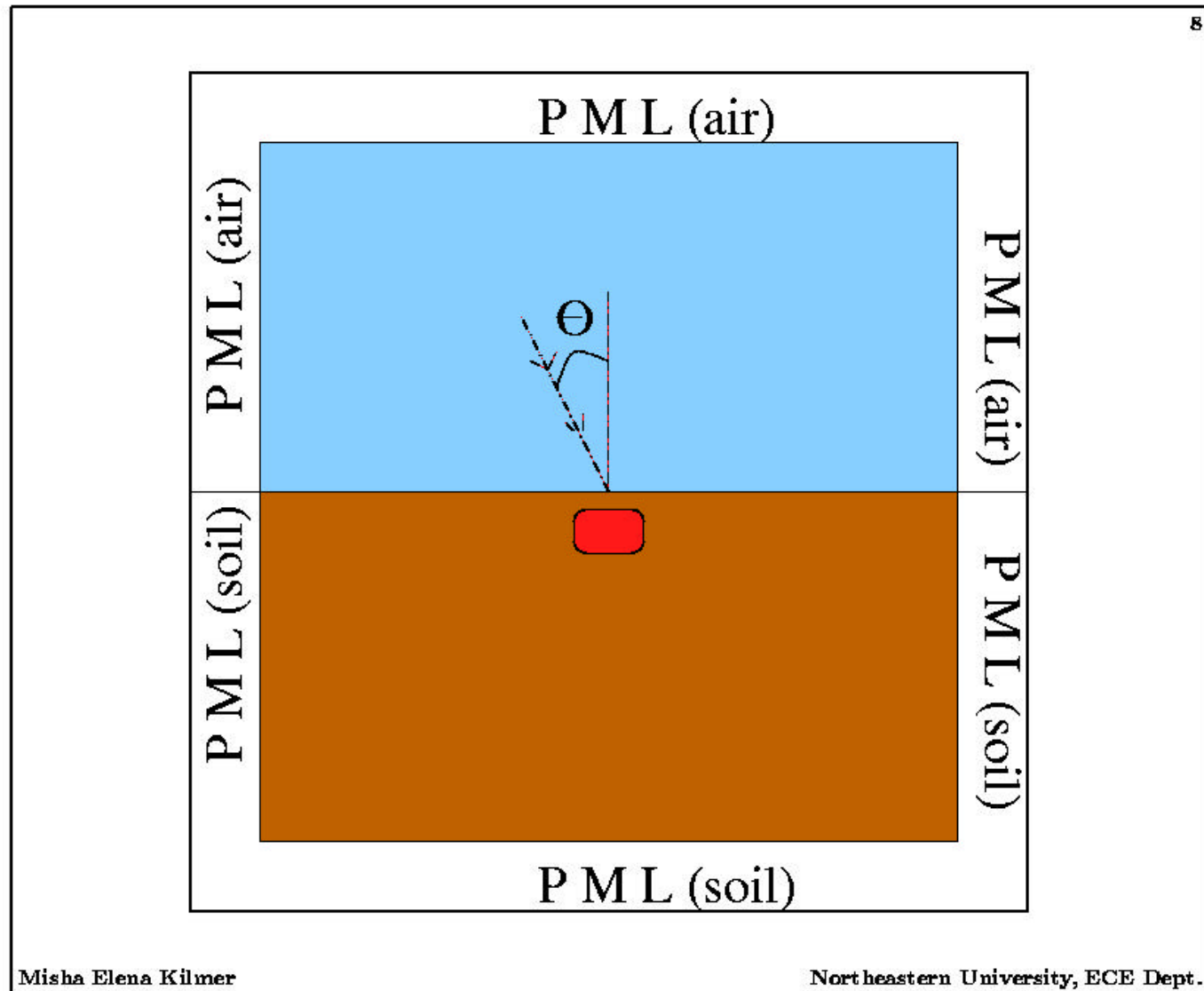
Freq. domain

- Finite Difference Frequency Domain (**FDFD**) is also good for non-uniform media.
- FDFD handles frequency dependent media easily.
- FDFD is particularly useful for signal processing forward models
- Using **Pre-Conditioning**, FDFD is less computationally expensive.

PML Absorbing Boundary Condition is perfectly suited to FDFD half-space problems

- **PML necessary for wide-angle absorption**
- **Frequency independence useful for general use**
- **PML loss profile can be added to lossy material characteristics**
- **Horizontal layering is amenable to sparse storage/computation**

Two-Dimensional Finite Difference Computational Geometry



Flop and Storage Comparison

For 10 points per λ and $N \times N$ matrix A :

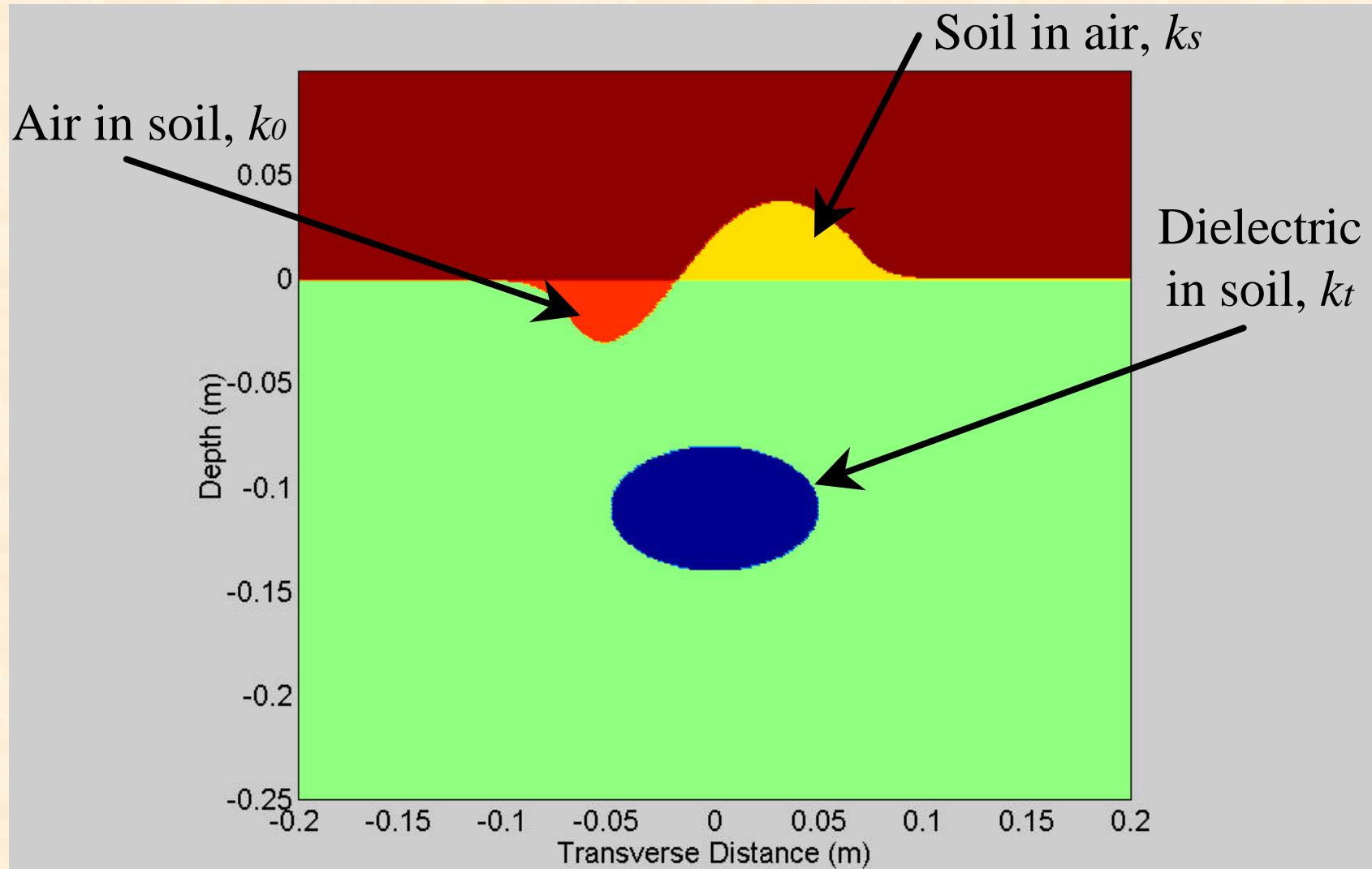
distance	N	$x = A \setminus b$		PQMR	
		flops	non-zeros	flops	non-zeros
6.2λ	6K	$1.33\text{e}+8$	$9.8\text{e}+5$	$9.13\text{e}+7$	$1.2\text{e}+5$
12.6λ	20K	$1.20\text{e}+9$	$5.8\text{e}+6$	$3.73\text{e}+8$	$3.8\text{e}+5$
25.4λ	73K	$9.72\text{e}+9$	$3.9\text{e}+7$	$1.79\text{e}+9$	$1.3\text{e}+6$
51λ	278K	$*1.5\text{e}+11$	$*2.9\text{e}+8$	$8.6\text{e}+9$	$5.0\text{e}+6$
102.2λ	1080K	$*3.5\text{e}+12$	$*2.24\text{e}+9$	$2.6843\text{e}+10$	$2.0\text{e}+7$

2 Orders of Magnitude Speed-Up Using
Pre-Conditioner for Big Problems

Dealing with rough surfaces with plane wave FDFD computation

- Analytically determine half-space reflection and transmission for nominal flat ground
- Consider buried objects as scattering sources in soil background
- Consider depressions and protrusions as perturbations to flat ground surface
 - ❖ Protrusions above nominal ground = soil scatterers in air background
 - ❖ Depressions = air void scatterers in soil background

Rough surface as perturbation to flat half space



Scattering formulation of surface and subsurface perturbations

Helmholtz Equation with varying background:

$$[\nabla^2 + k^2(x, y)](E_s + E_{unp}) = 0$$

Separate scattered field:

$$[\nabla^2 + k^2(x, y)]E_s = -[\nabla^2 + k_{unp}^2]E_{unp} - (k^2(x, y) - k_{unp}^2)E_{unp}$$

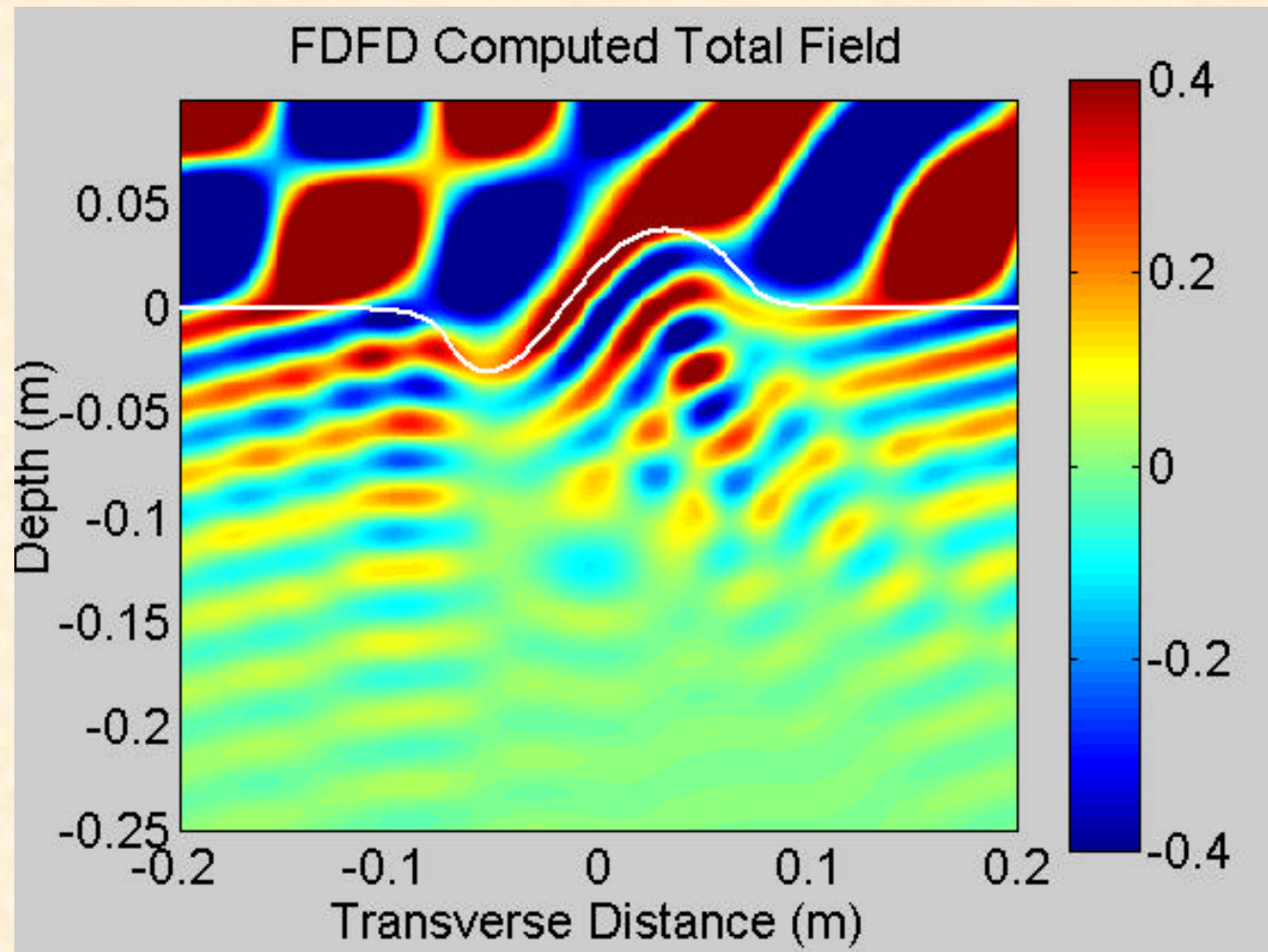
Cancel unperturbed field:

$$[\nabla^2 + k^2(x, y)]E_s = -k_s^2 O(x, y)E_{unp}$$

where the object function is given by:

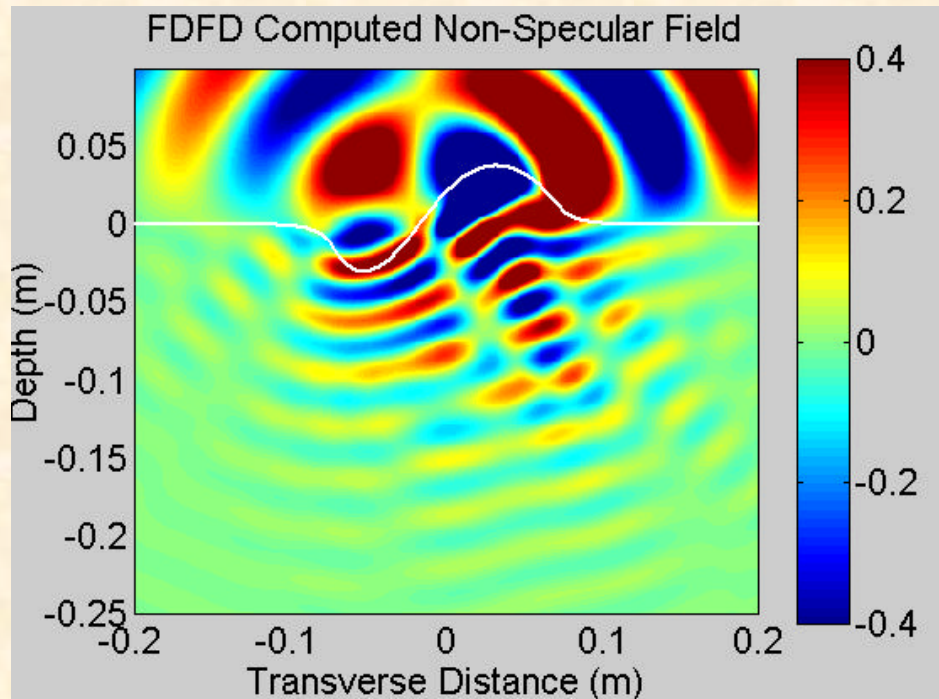
$$O(x, y) = \begin{cases} (k_t / k_s)^2 - 1, & \text{Over the buried object} \\ (k_0 / k_s)^2 - 1, & \text{Over the depressions} \\ 1 - (k_0 / k_s)^2, & \text{Over the protrusions} \end{cases}$$

3 GHz 45 deg. incident plane wave on rough surface, with buried TNT ellipse

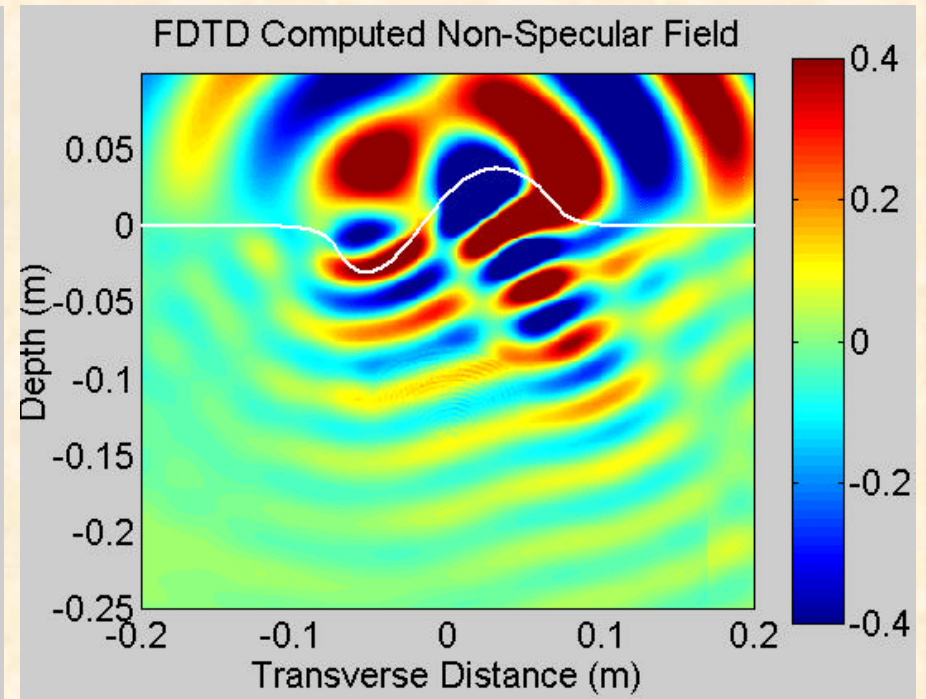


Validation: FDFD vs. FDTD run at 3 GHz for 4000 time steps with PML ABC

45 deg. incident plane wave on rough surface, dispersive soil
with buried TNT ellipse with specular wave removed

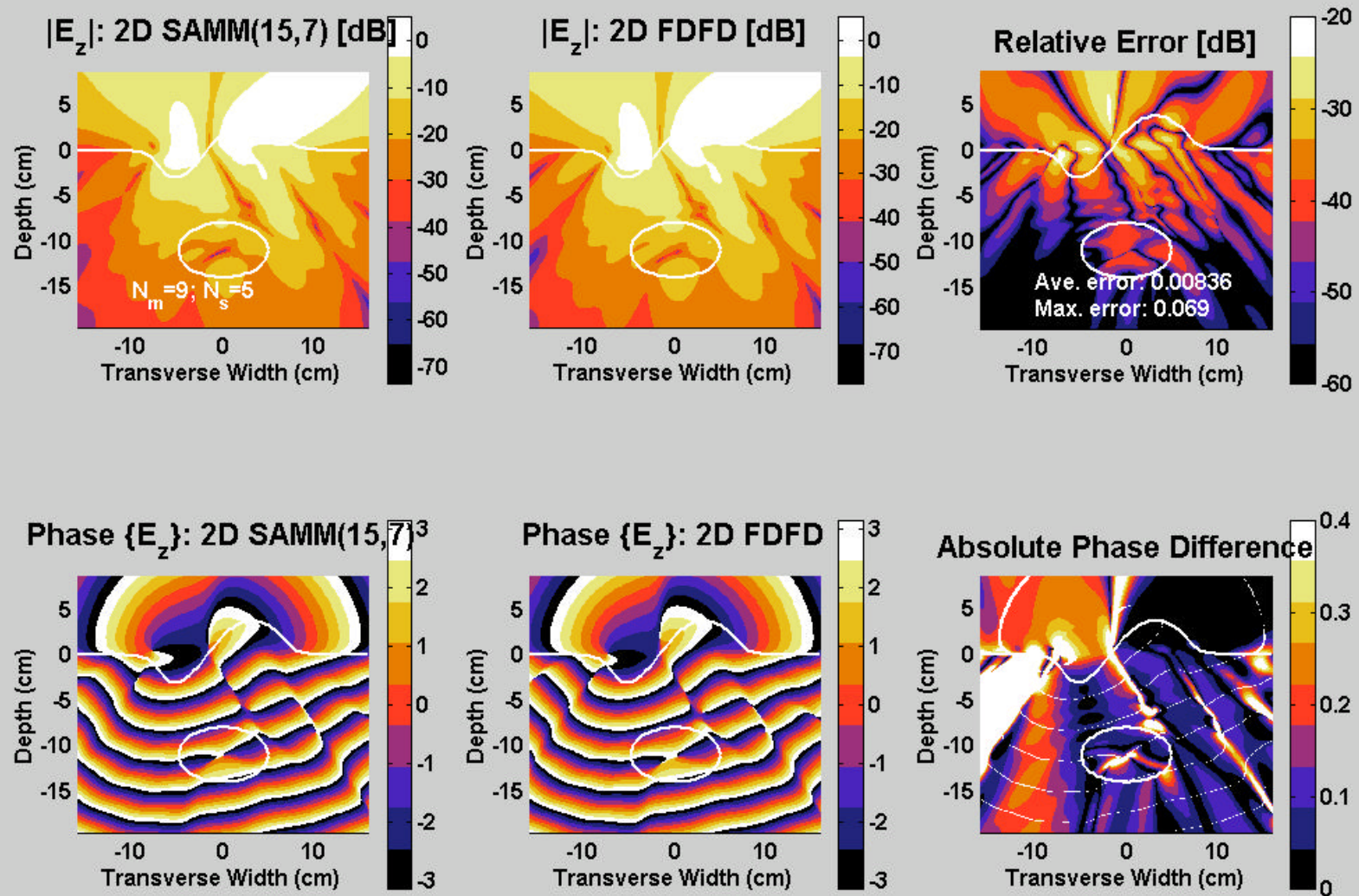


~10 min. on 850 MHz Pentium with 512 MB



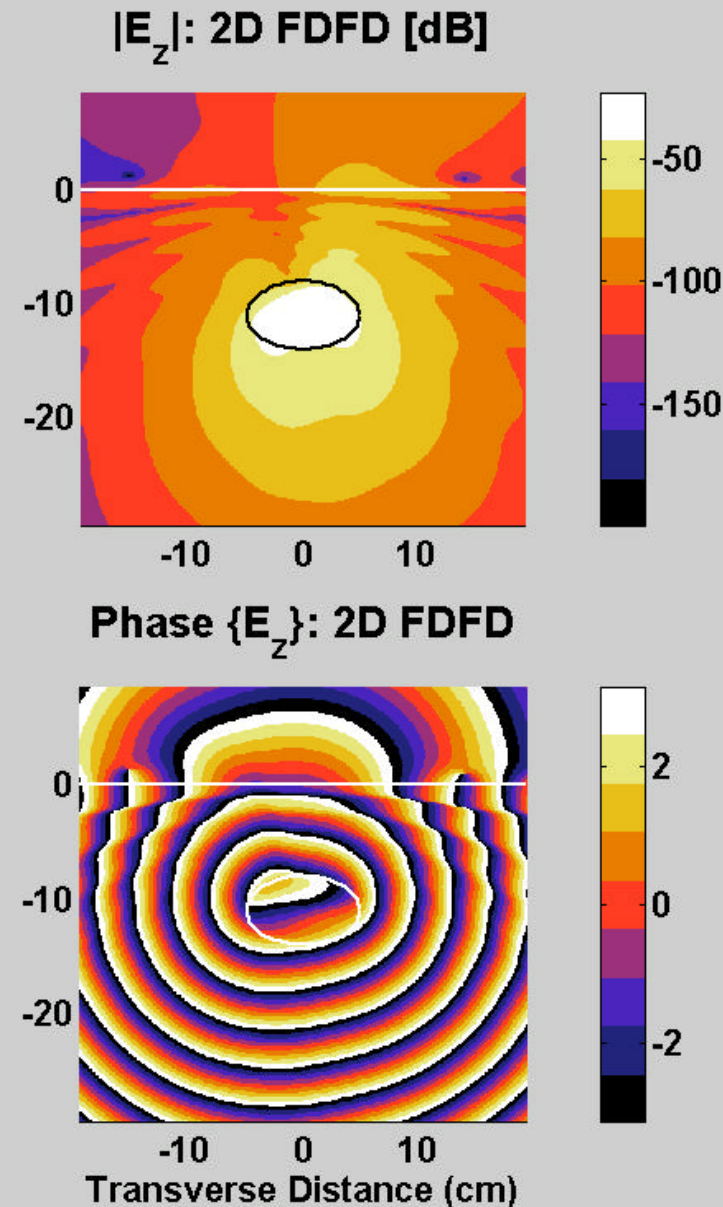
~45 min. on 850 MHz Pentium with 512 MB

Validation: FDFD vs. Semi-Analytic Mode Matching (frequency domain) method



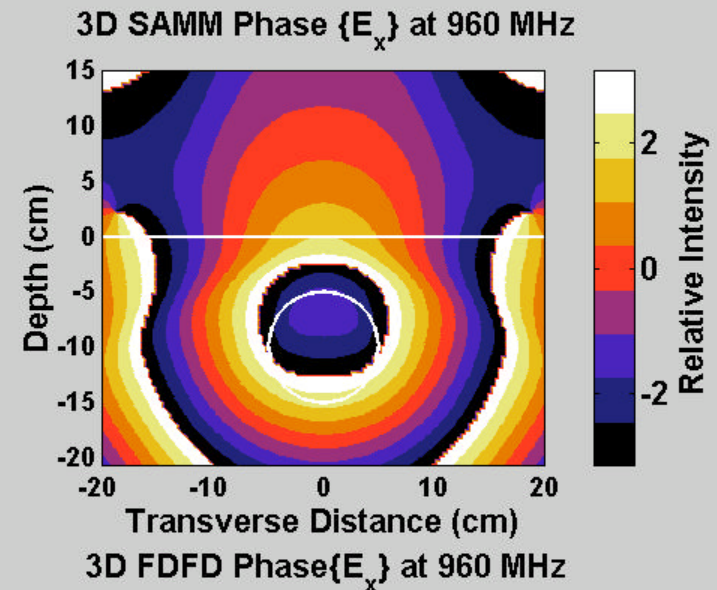
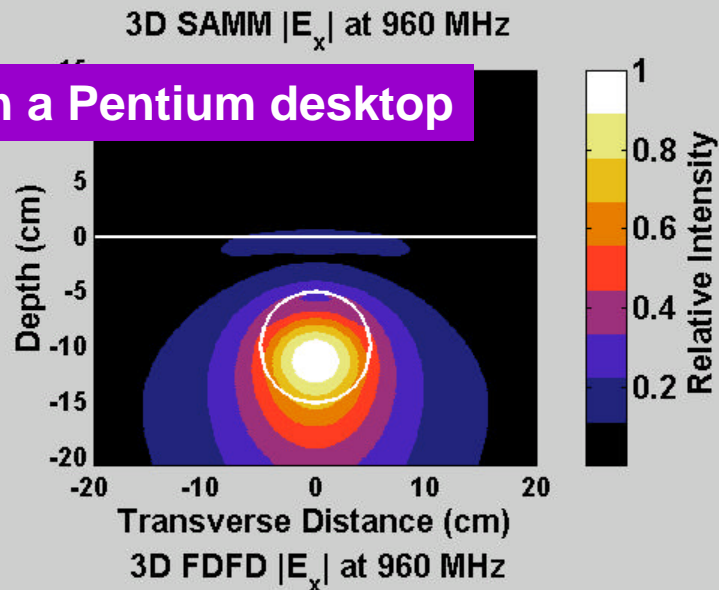
Flat surface simplifies scattered wave pattern

45 deg. incident plane wave on flat surface, dispersive soil with buried TNT ellipse with specular wave removed

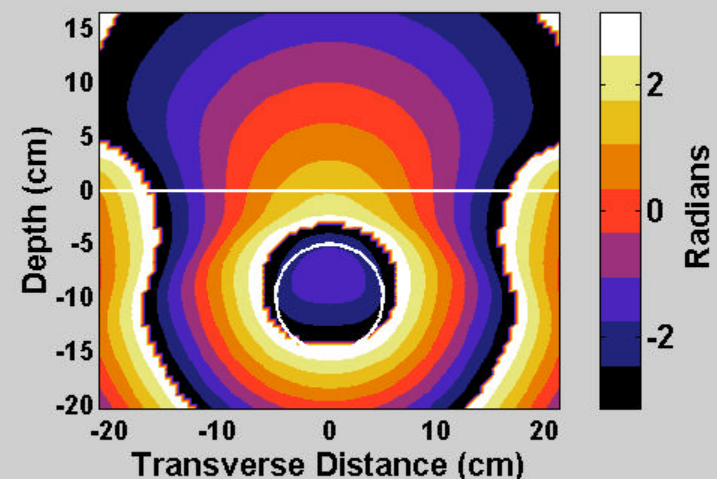
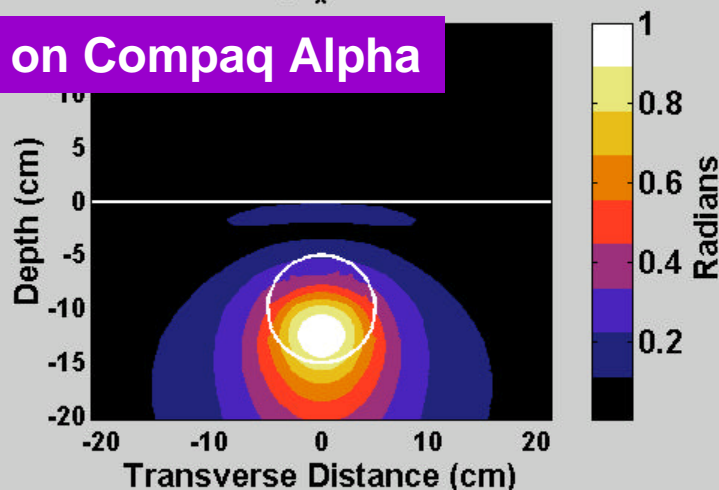


3D FDFD v. 3D Semi-Analytic: Bosnian Soil

40 min. on a Pentium desktop



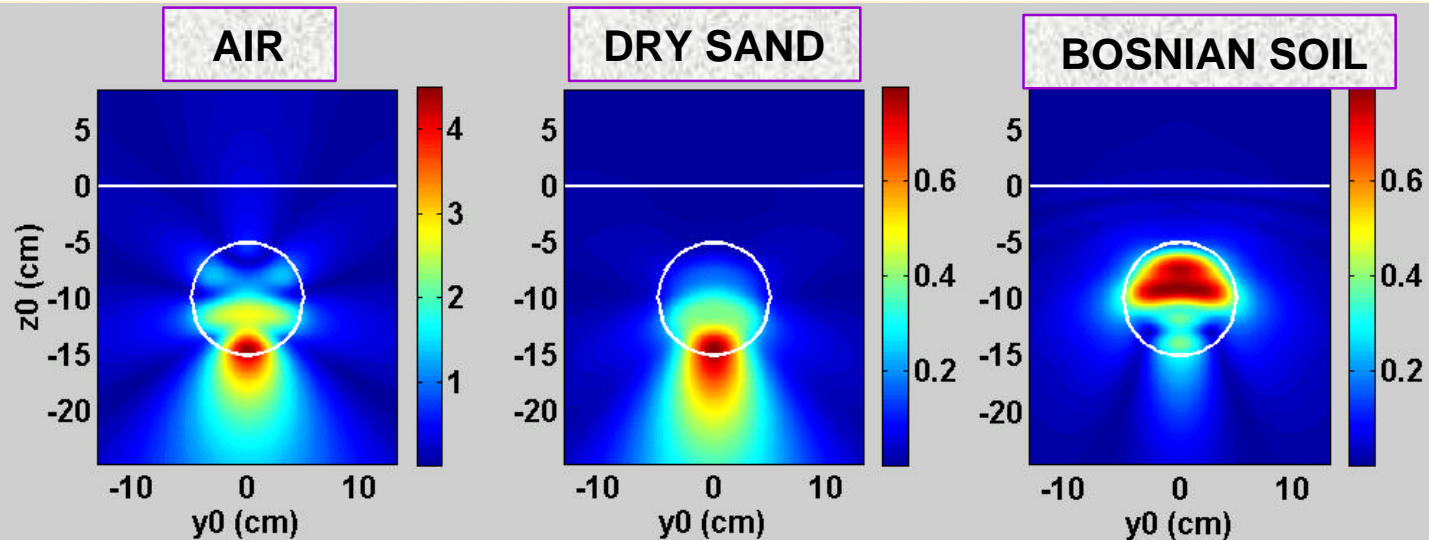
75 hours on Compaq Alpha



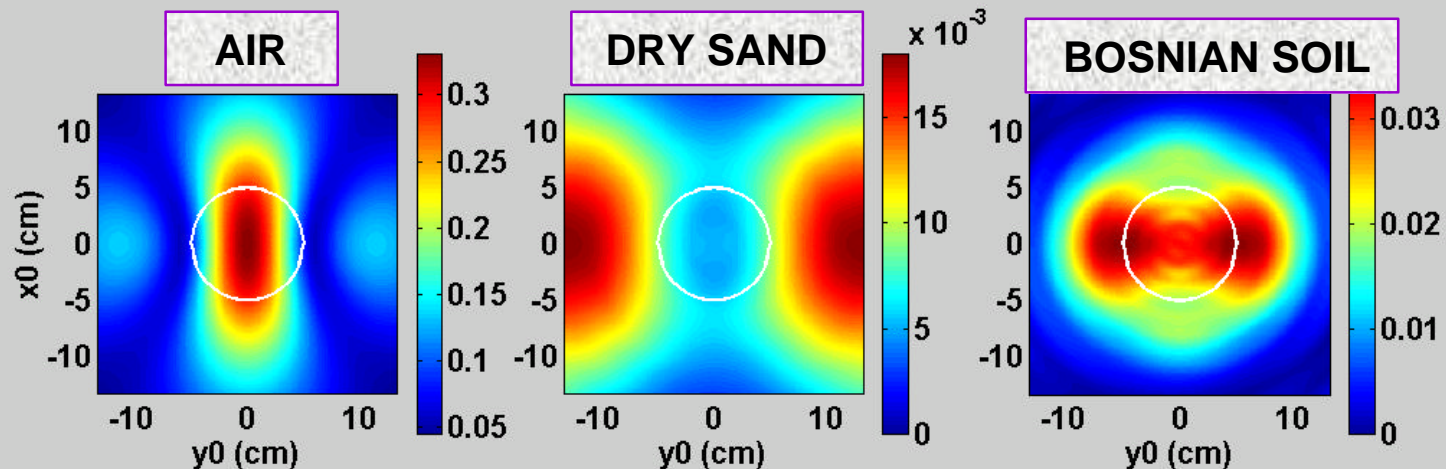
$\epsilon_g = (9.19 + 1.27 i) \epsilon_0$; normal incidence; TNT mine

2D Scattering from Circular Cylinder in Different Soils at 3 GHz

Depth



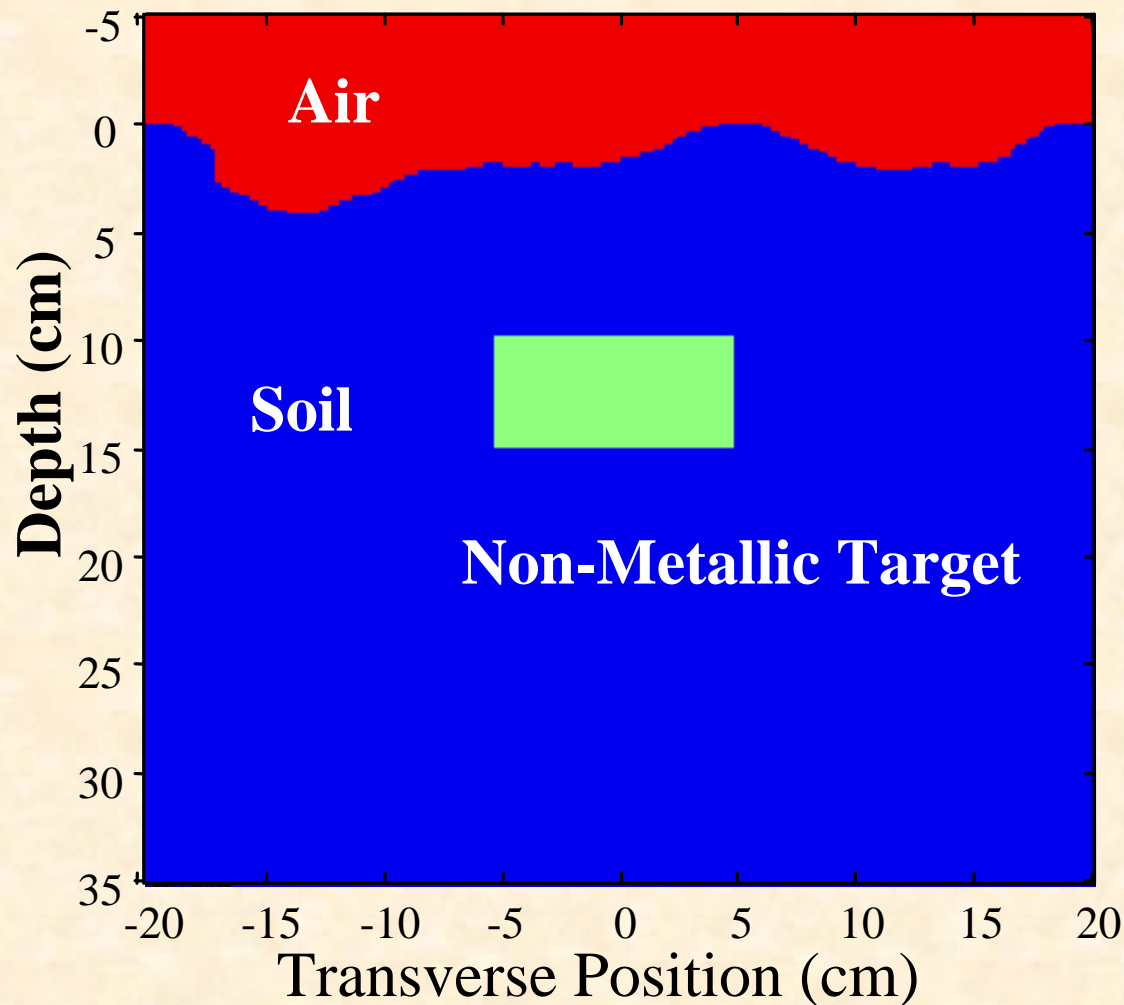
Surface



TNT mine, $f=3\text{GHz}$, $R=5\text{cm}$, $d=5\text{cm}$, normally-incident plane wave

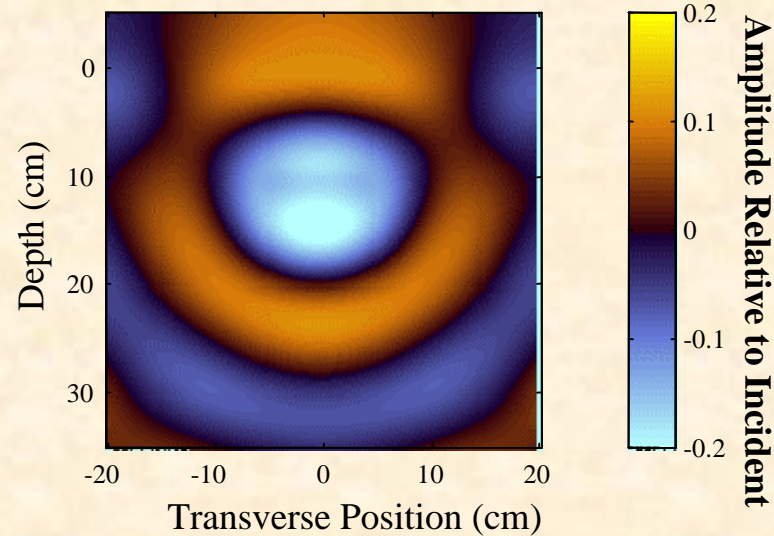
Surface Scattering Clutter Increases with Frequency.

Example: 4 GPR Frequencies, PRCL 10% moisture, 1.4 g/cc density

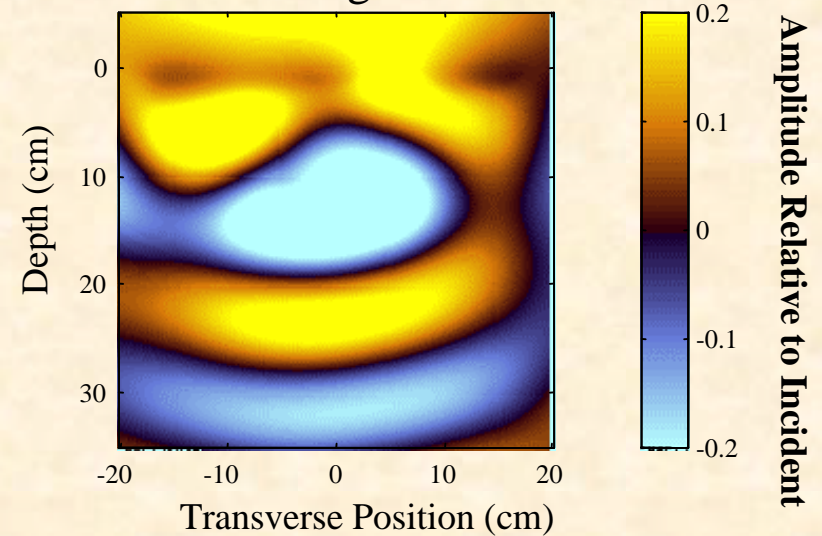


480 MHz

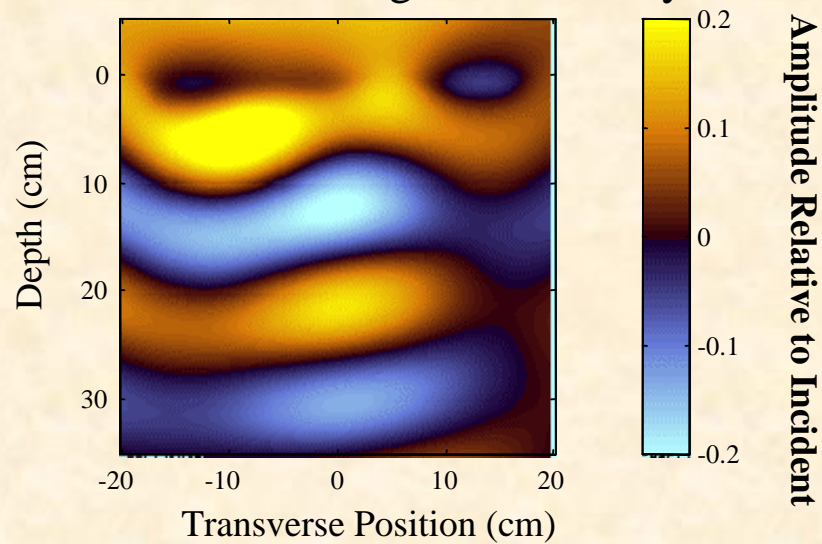
Mine scattered field: smooth surface



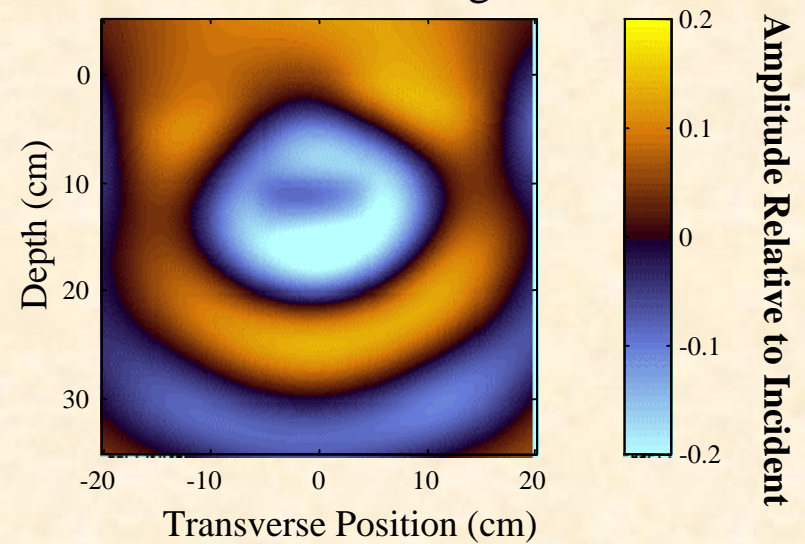
Scattered field: rough surface with mine



Scattered field: rough surface only

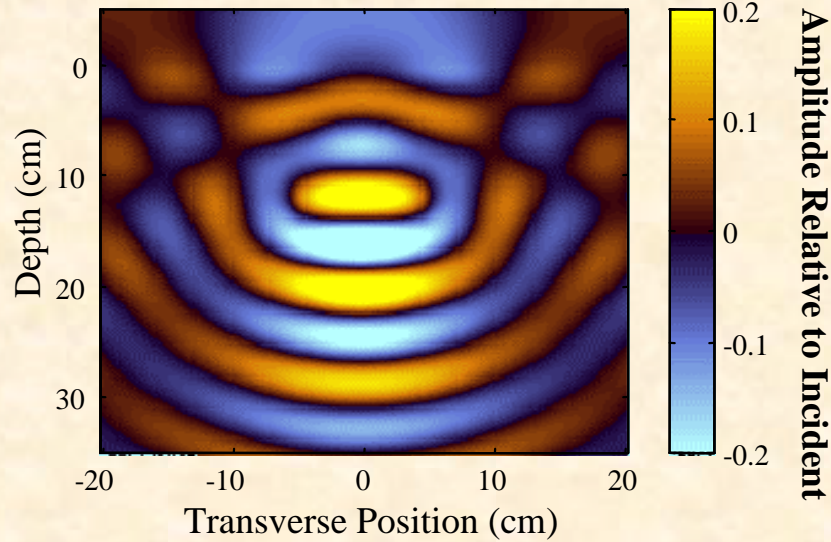


Mine scattered field: rough surface

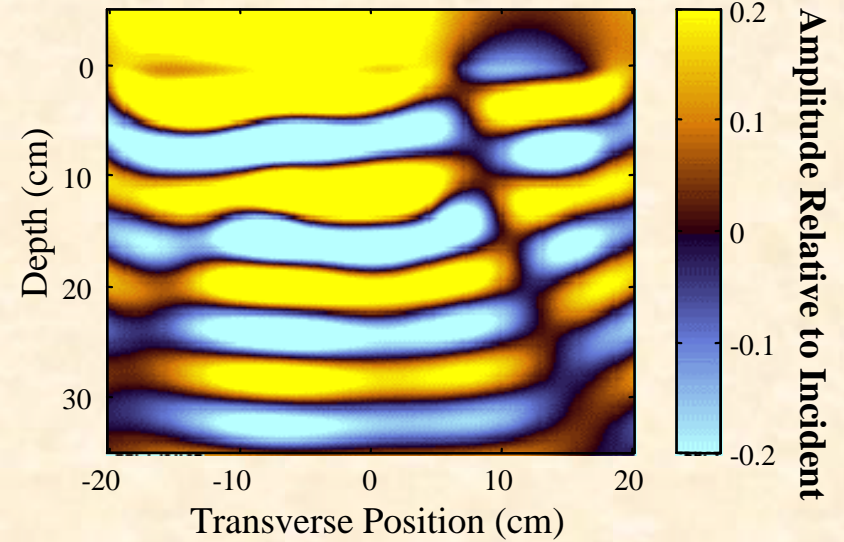


960 MHz

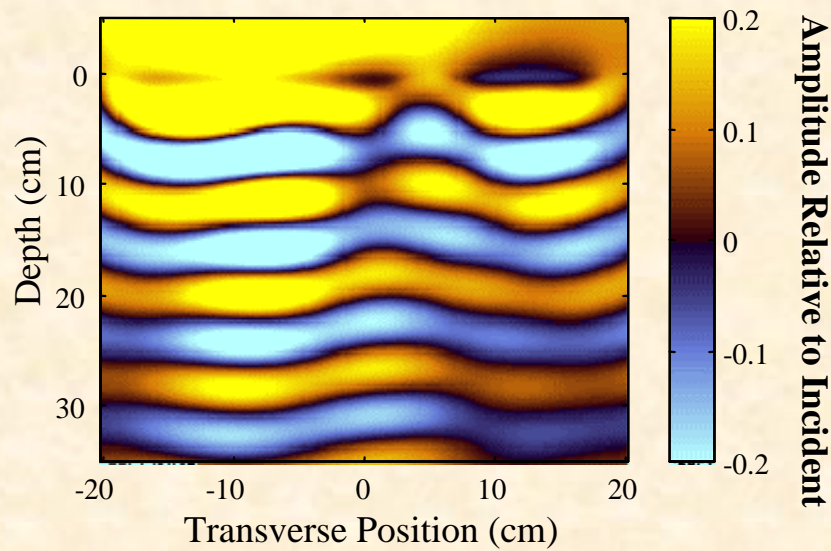
Mine scattered field: smooth surface



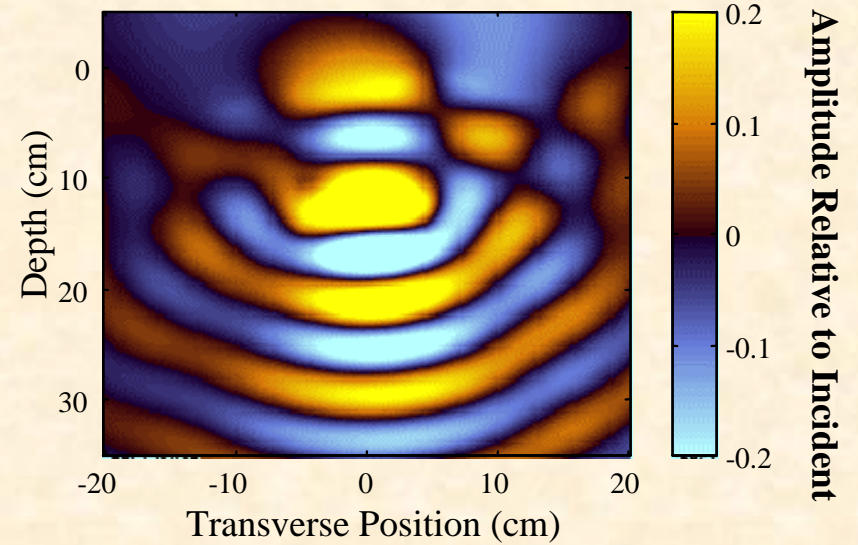
Scattered field: rough surface with mine



Scattered field: rough surface only

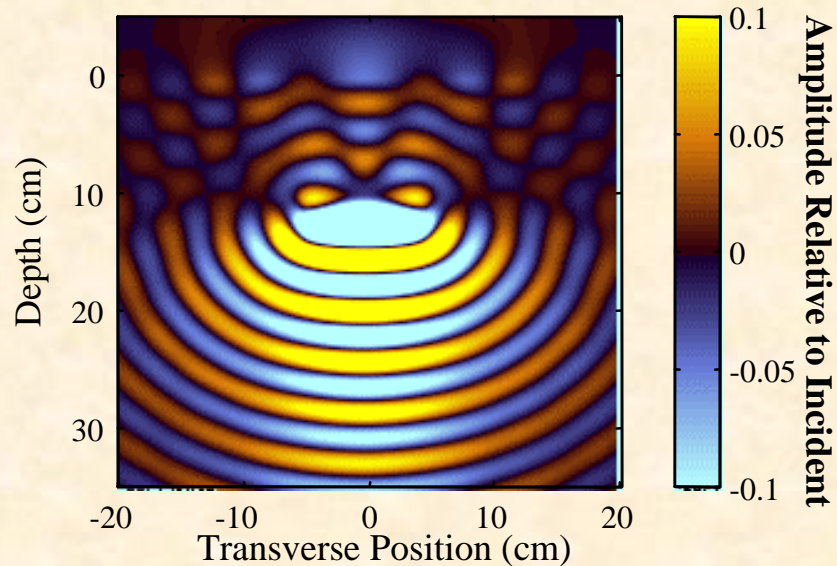


Mine scattered field: rough surface

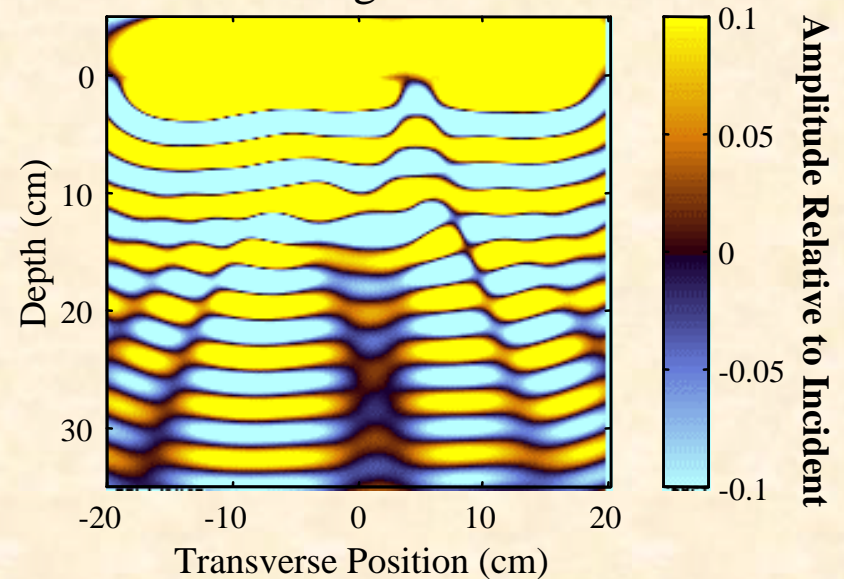


1920 MHz

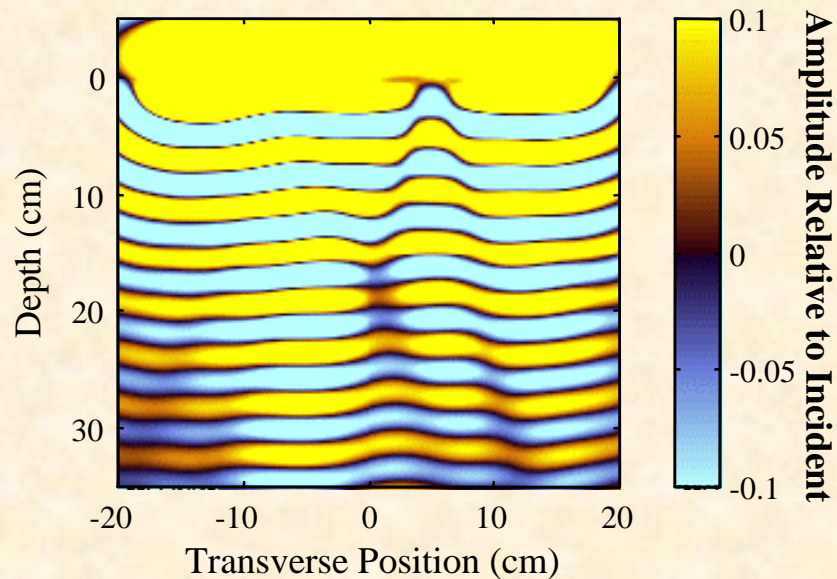
Mine scattered field: smooth surface



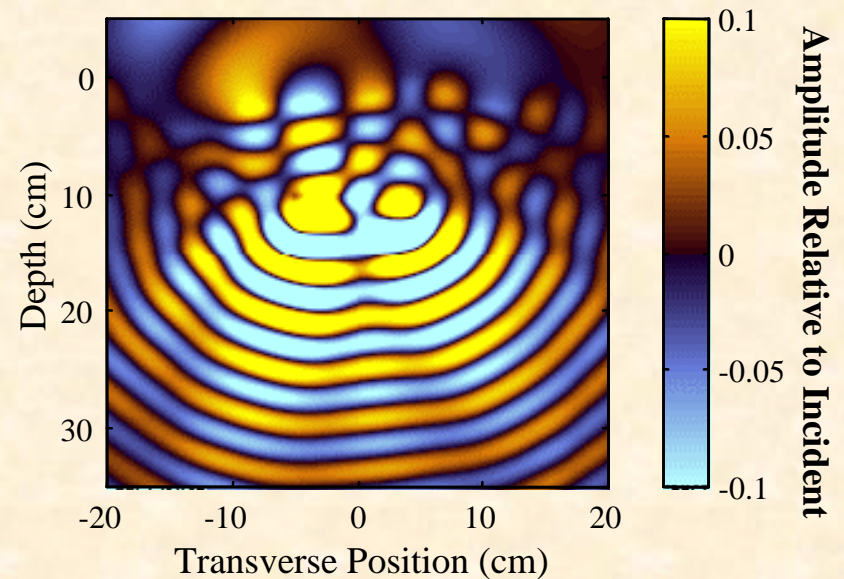
Scattered field: rough surface with mine



Scattered field: rough surface only

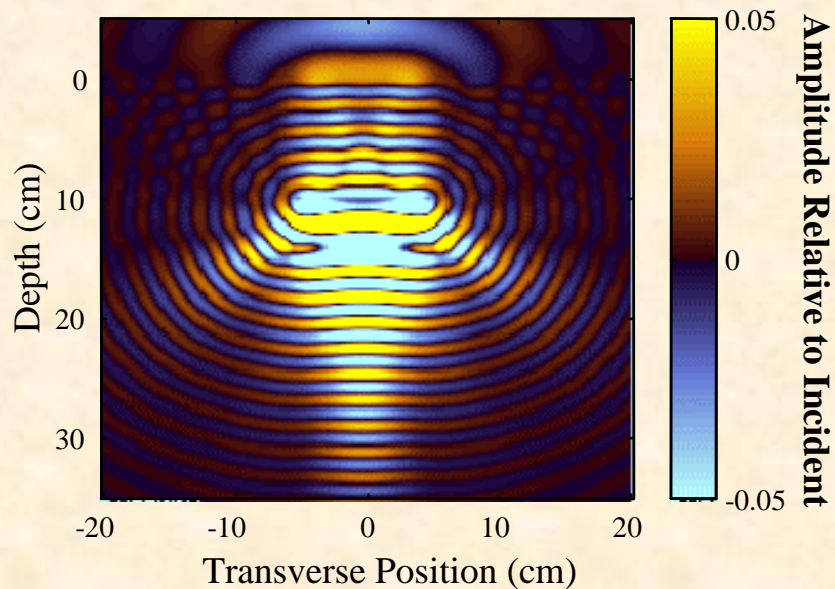


Mine scattered field: rough surface

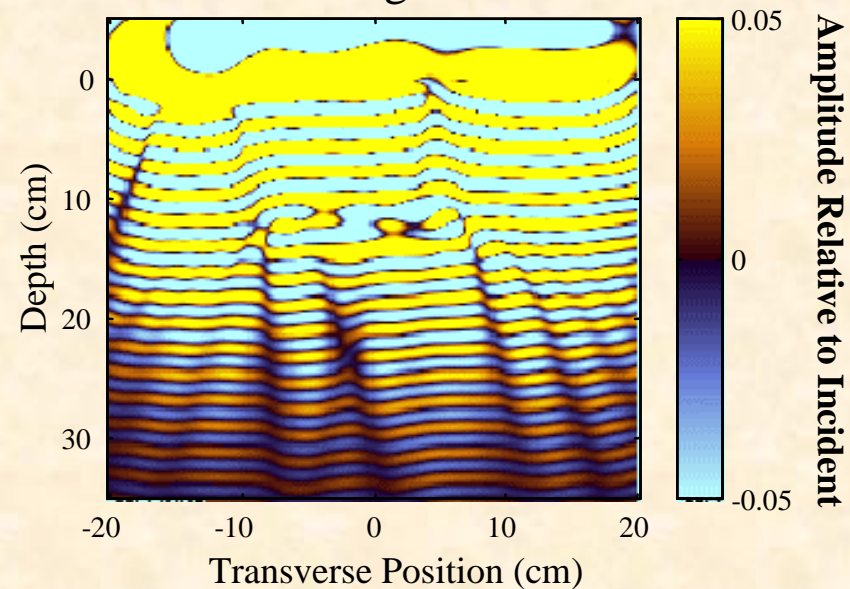


3840 MHz

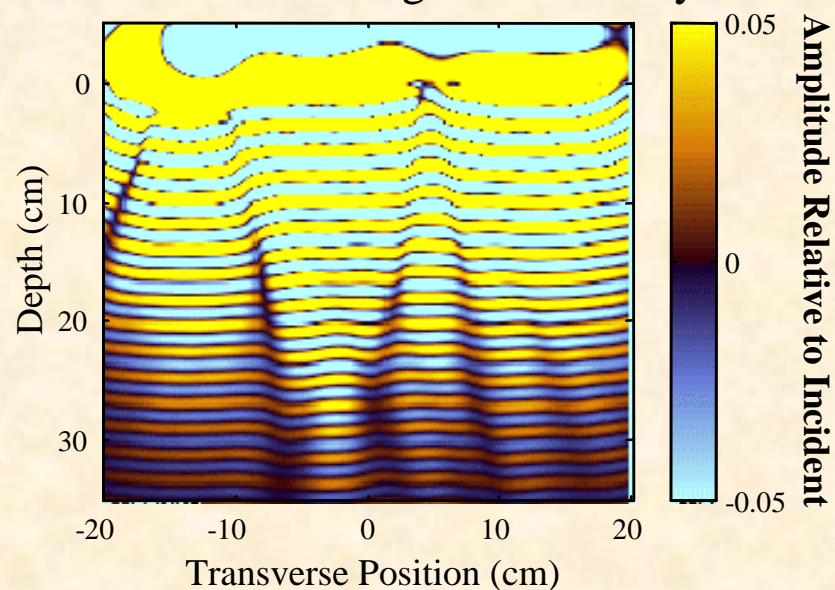
Mine scattered field: smooth surface



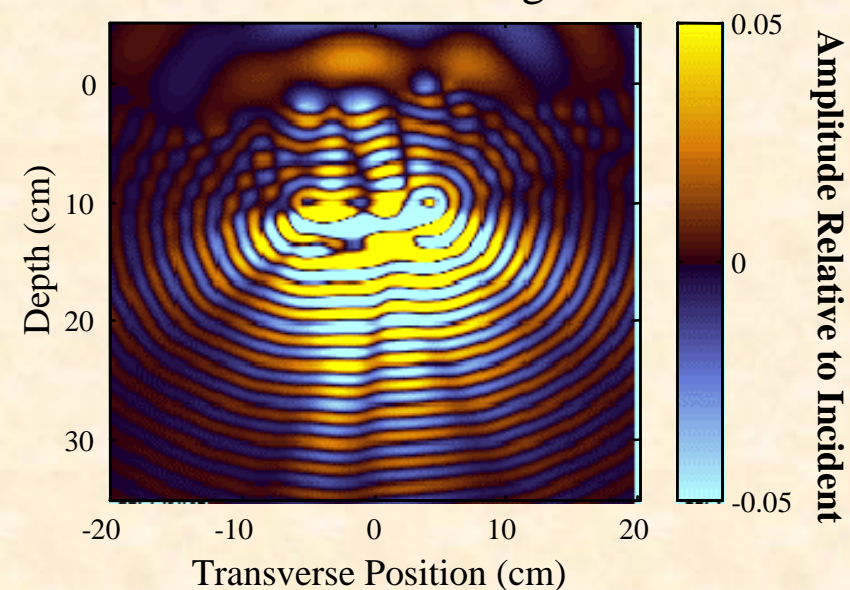
Scattered field: rough surface with mine



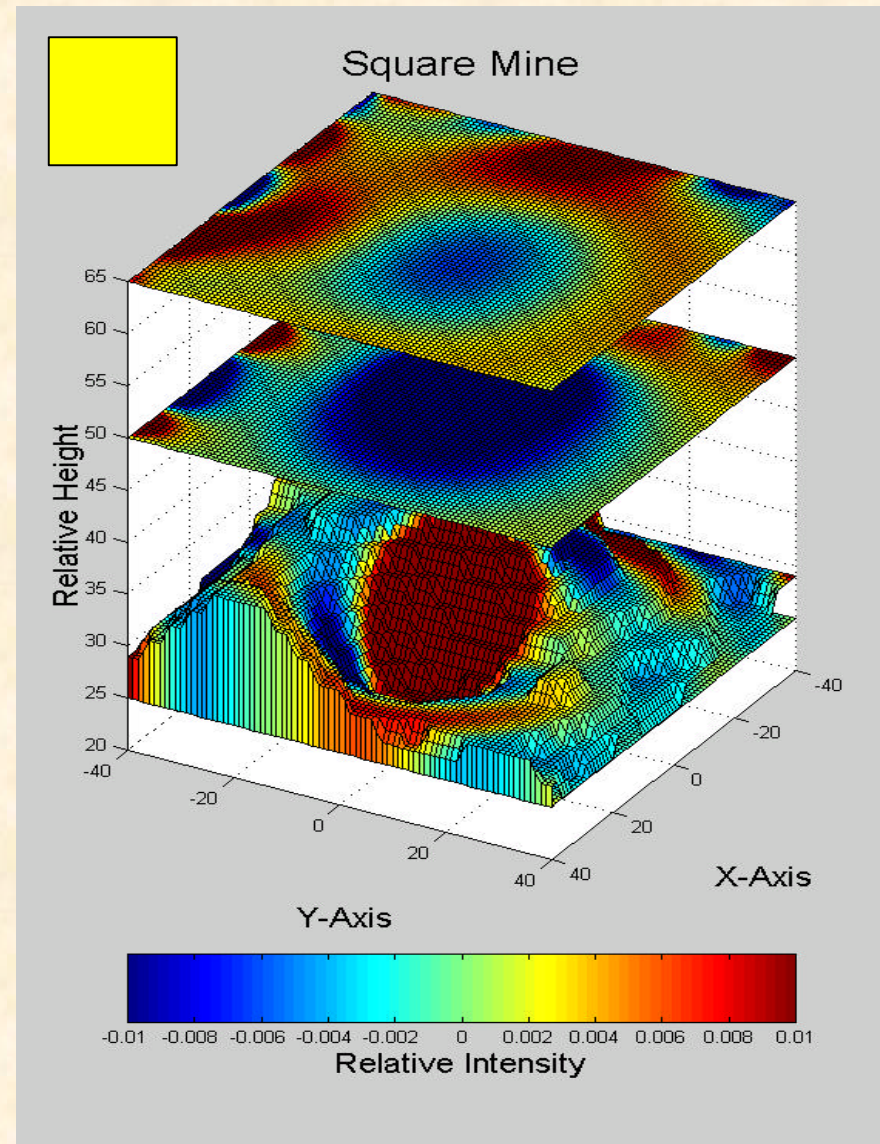
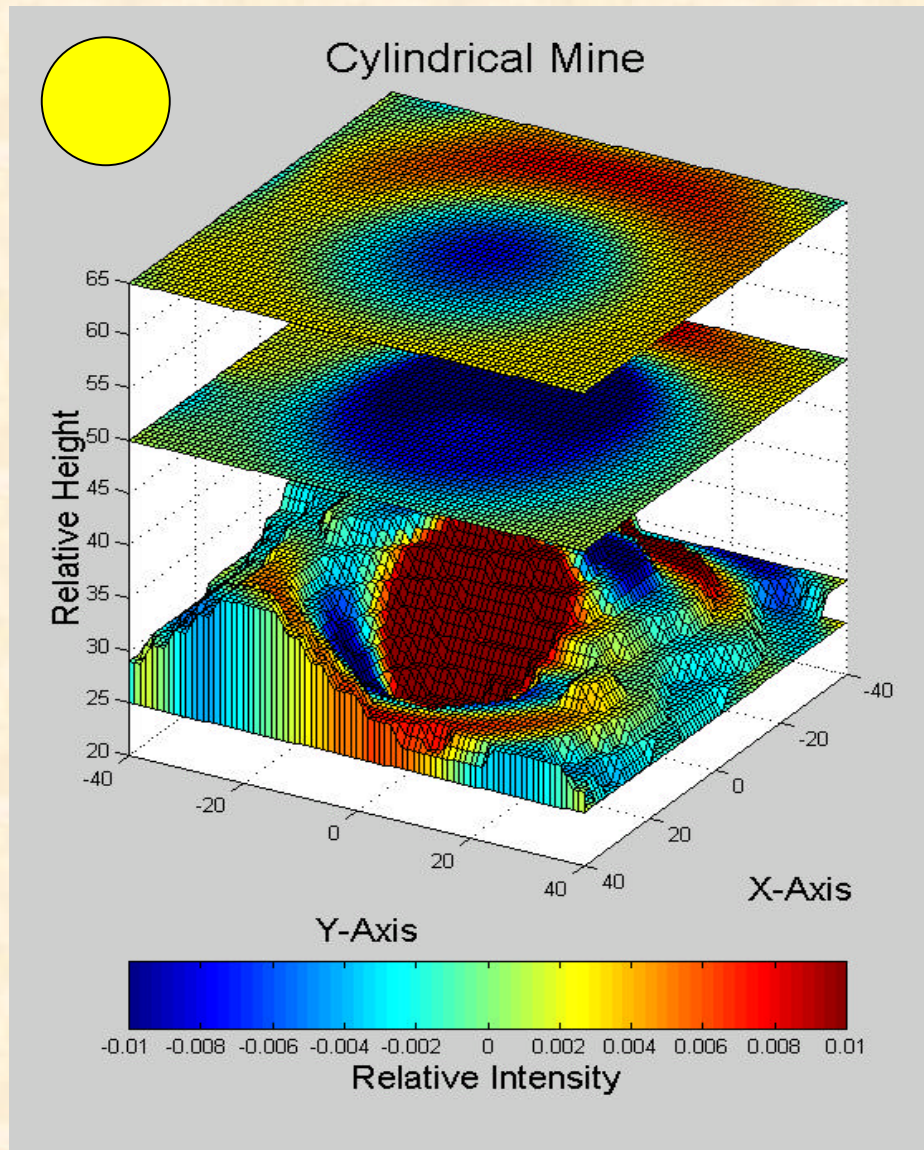
Scattered field: rough surface only



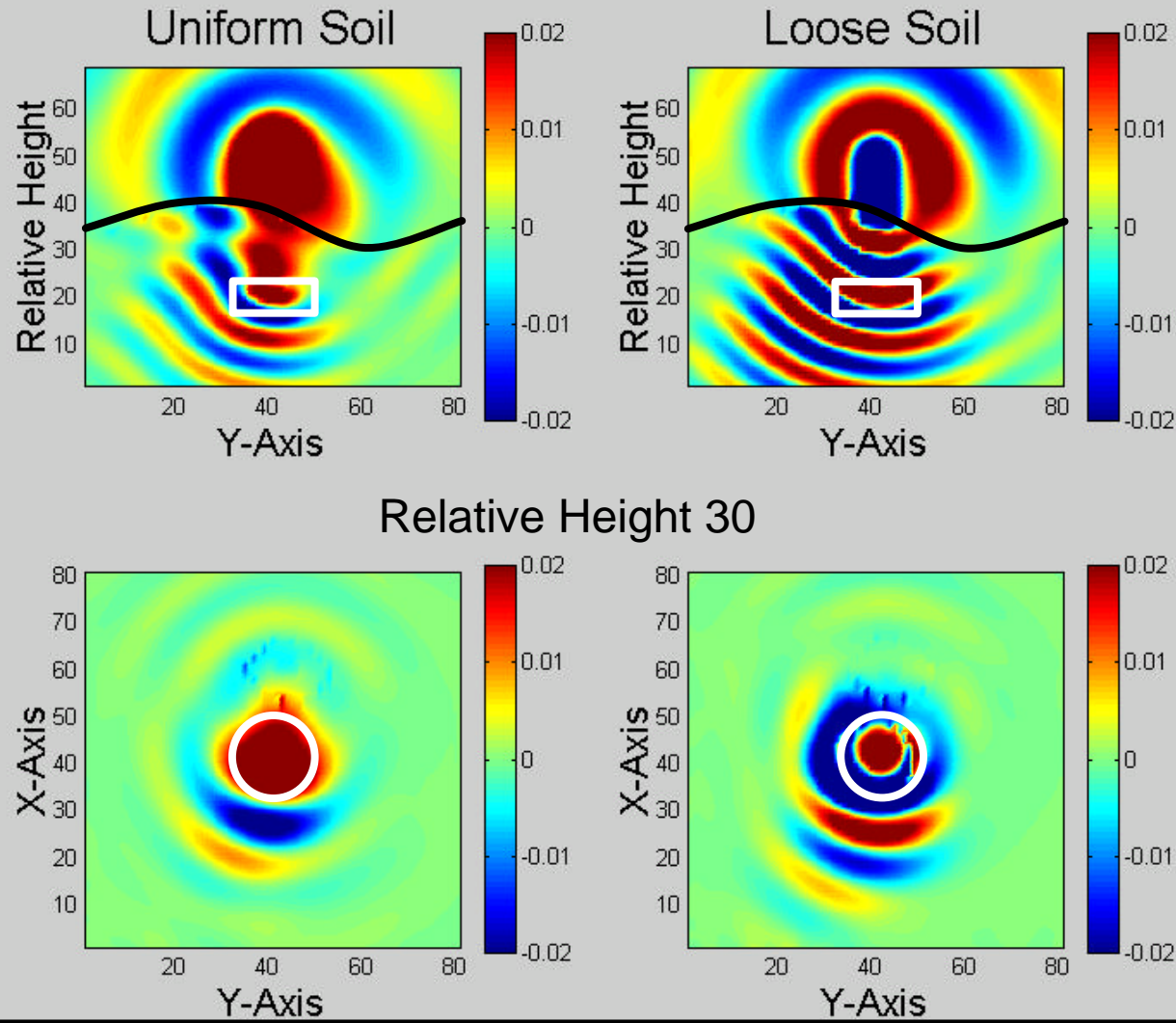
Mine scattered field: rough surface



3D Target Scattered E-Field, Real Part, 960 MHz



Cylindrical Scatterer Under Rough Surface: Scattered Field Only



TNT in 25% moisture Bosnian soil at 960 MHz

Conclusions

- **FDFD is effective in efficiently analyzing subsurface scattering of objects buried under rough surfaces**
- **Plane wave FDFD is validated by FDTD and semi-analytic methods**
- **Subsurface scattering effects are often non-intuitive**

Hyperspectral Imaging Using AOTF and NIR Sensing of Buried Objects and Landmines

J. Mobley¹, G. H. Miller¹, P. M. Kasili^{1,2}, C. DiMarzio³, and T. Vo-Dinh^{1,2(*)}

¹Oak Ridge National Laboratory, Oak Ridge, TN 37831-6101

²University of Tennessee, Knoxville, TN 37916

³Northeastern University, Boston, MA 02115

(*) Author for Correspondence: tvo@ornl.gov

Abstract

The detection of landmines and buried objects requires methods that can cover large areas rapidly while providing the required sensitivity to detect the optical and spectroscopic contrasts in soil properties that can reveal their presence. These conditions on contrast and coverage can be met by capturing images of the soil at wavelengths which are sensitive to the properties modified by the presence of buried objects. In this work we investigate both imaging and scanning methods which may have some utility for the detection problem. In the imaging approach, we capture hyperspectral reflection images using an acousto-optic tunable filter (AOTF) and fluorescence images using a long-pass filter. For the scanning method, we acquire data point-by-point over a two-dimensional grid with a single emitter/detector pair. The results illustrate the potential of these two approaches for detection of landmines and buried objects.

I. Introduction

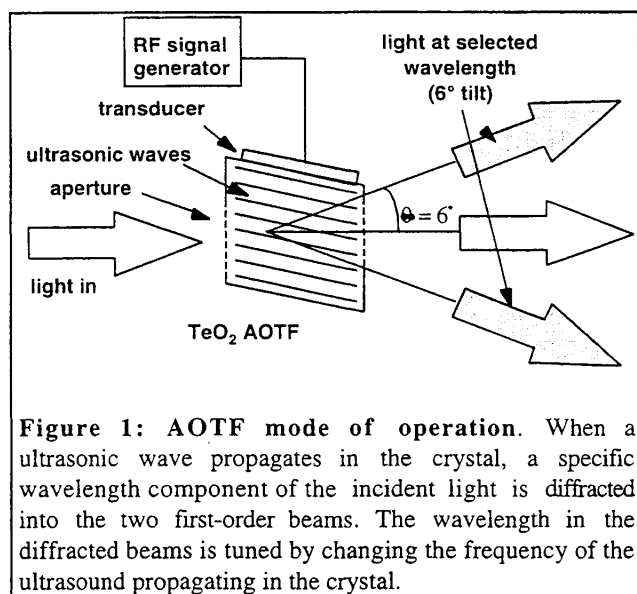
Indirect methods of detecting buried objects involve looking for differences between undisturbed soil and soil directly above a buried object. The detection of landmines and buried objects requires methods that can cover large areas rapidly while providing the sensitivity to the contrasts in soil properties that can reveal the presence of foreign objects. The conditions on coverage can be met by acquiring images of large areas or rapidly scanning over the ground. The contrast is provided by acquiring the data at specific wavelengths or over continuous spectral bands which are particularly sensitive to the properties affected by the presence of buried objects. In this work we look at both imaging and point-by-point (scanning) methods which can be adapted for use in the detection problem. The images were acquired by a charge-coupled device (CCD). Two types of images were captured: (1) narrowband reflection images using an acousto-optic tunable filter (AOTF) and (2) fluorescence images using a long-pass filter. The scanning method employed a near-infrared (NIR) emitter/detector pair where data were captured point-by-point over a two-dimensional grid.

II. Imaging-Based Methods for Object Detection

IIa. Reflection Imaging with an Acousto-Optic Tunable Filter

By taking two-dimensional images of a target zone, a large area can be spectrally mapped quickly given that the contrast between the background soil and burial sites are sufficiently strong to allow rapid detection. One method of increasing contrast is to image at/over wavelengths where the differences in reflection or fluorescence properties are maximized. These types of images can be captured rapidly by using an acousto-optic tunable filter (AOTF) for wavelength discrimination. The AOTF is

a narrowband optical filter with a electronically-selectable passband and a two-dimensional aperture. It can quickly jump between remote wavelengths or be rapidly scanned across some continuous spectral range. Band-limited (i.e., a spectral range wider than the resolution width of the filter) imaging can be performed by scanning the AOTF across a specific range of

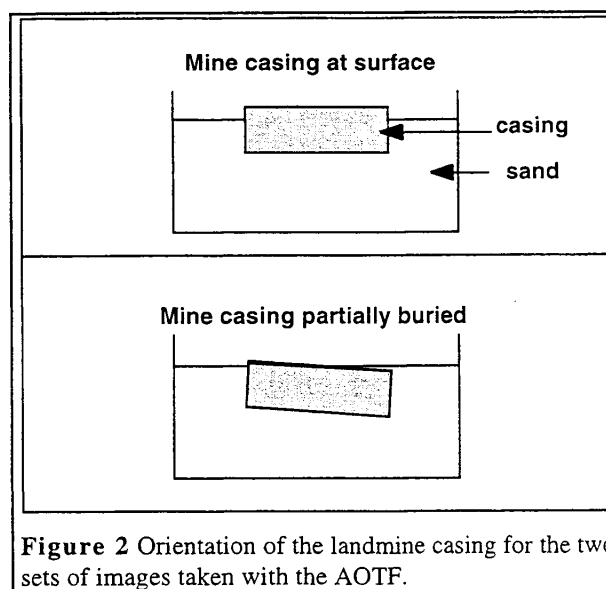


wavelengths while exposing the image-capturing elements (e.g., film or charge-coupled device array (CCD)). Our laboratory has previously reported on the use of AOTF/CCD based systems for hyperspectral imaging[1-3].

The AOTF operates by propagating ultrasonic waves through an anisotropic crystal. The strains induced by the ultrasound act to spatially modulate the indices of refraction of the crystal. The device is essentially a three-dimensional grating whose spacing can be tuned by changing the wavelength of the ultrasound in the crystal. Since the AOTF has no moving parts and can be rapidly moved to the desired wavelength, it is ideal for use in portable instrumentation where spectral selectivity and timely feedback are required. Figure 1 shows the how the

light is discriminated by the TeO₂ AOTF used here.

In this work our imaging system was limited to wavelengths shorter than 700 nm. Thus to demonstrate the capabilities of an AOTF-based system to make wavelength-resolved images, we have examined a landmine casing both above and partially buried in sand (see Figure 2) illuminated by an incandescent lamp and imaged at wavelengths from 525 to 700 nm. A white-light image of the casing is shown in Figure 3. The imaging system is diagrammed in Figure 4 and employs a TeO₂ AOTF (Brimrose, Inc. tunable from 450 to 700 nm), a thermoelectrically cooled CCD (Santa Barbara Instruments, Model ST-6) and a camera lens with adjustable aperture and focus. At each wavelength, in addition to the landmine images, an image of a white card was also recorded along with the landmine casing for use in data normalization. The normalization process is diagrammed in Figure 5.



The normalized images for the exposed casing are displayed in Figure 6. The casing has five indentations on its surface (four around the rim and one arrow-shaped) which were filled with sand in the wavelength-resolved images. The best images in

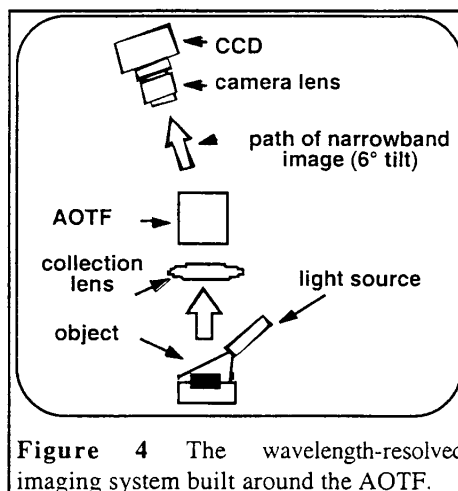
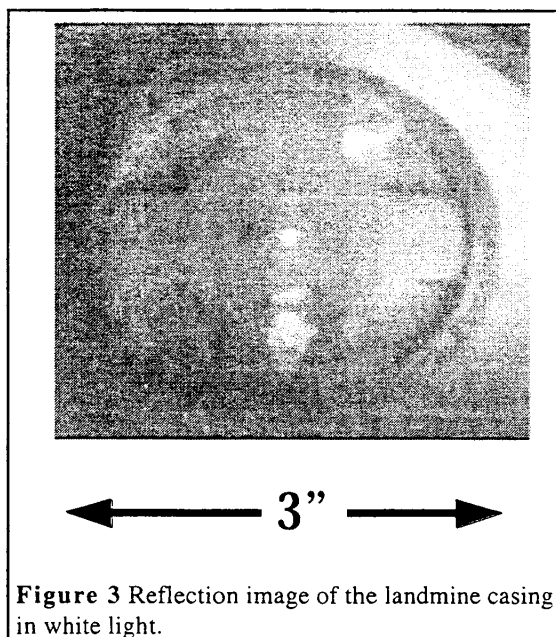


Figure 4 The wavelength-resolved imaging system built around the AOTF.

terms of clarity and contrast between the sand and casing are the three images taken at 600 nm, 625 nm, and 650 nm respectively. The images of the partially buried casing, as displayed in Figure 8 (Figure 7 gives the orientation of the casing as seen in Figure 8), showed similar results. For these, the 625-nm and 650-nm images display the best contrast. The resulting contrasts in these reflected-light pictures are of course related to the respective colors of the casing and sand. The casing is mostly green and as one moves to longer wavelengths (starting from 525 nm) the casing reflects less light while the background sand reflects more evenly across the spectrum. Thus the contrast between the two increases as one moves from green wavelengths to the red end of the spectrum. Future work will involve investigating local reflection properties of the soil at NIR wavelengths. By performing narrowband or wavelength-resolved imaging, the contrast in soil properties can be maximized.

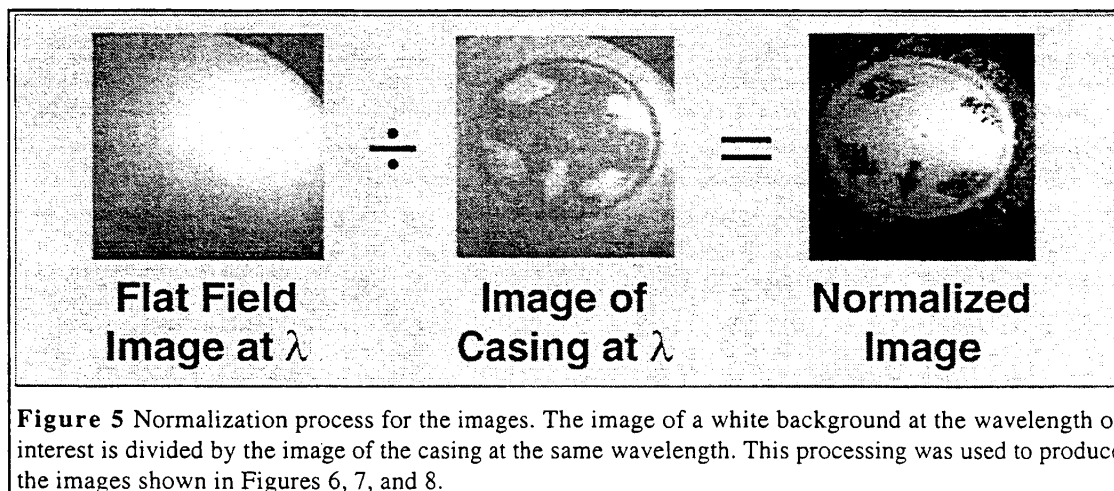


Figure 5 Normalization process for the images. The image of a white background at the wavelength of interest is divided by the image of the casing at the same wavelength. This processing was used to produce the images shown in Figures 6, 7, and 8.

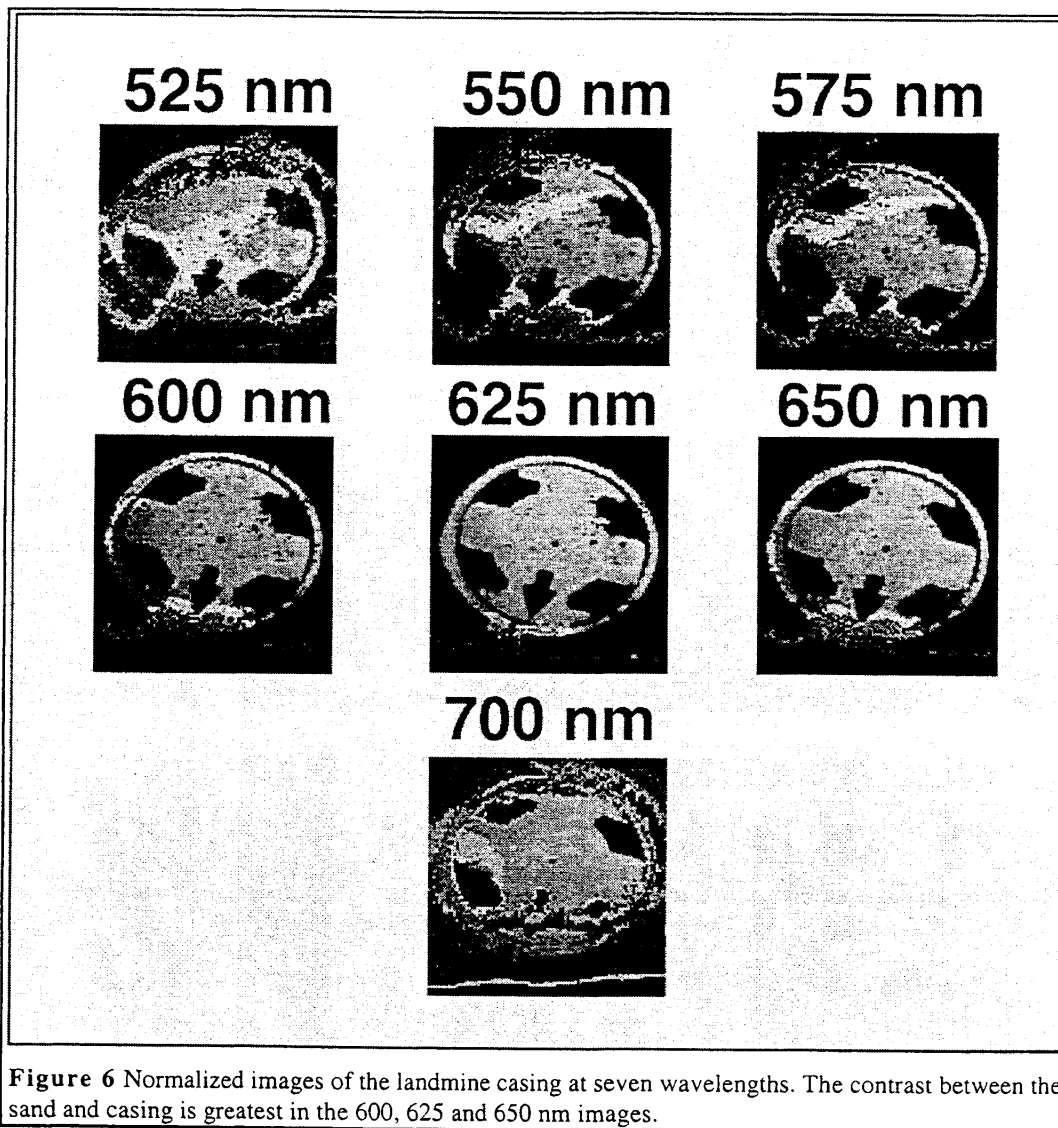


Figure 6 Normalized images of the landmine casing at seven wavelengths. The contrast between the sand and casing is greatest in the 600, 625 and 650 nm images.

IIb. Fluorescence Imaging

In addition to reflected light, we also captured images of the fluorescence emission from the partially-buried object. In this mode, the AOTF was replaced by a 525-nm long pass filter, the focusing lens/camera were axially aligned with the collection lens and the object was illuminated by the 488-nm line of an Argon laser. The captured image is shown in Figure 9. The high contrast between the casing and sand is readily apparent. If a condition of the soil above a buried mine can be detected via fluorescence, this method could be useful in low-light situations where the fluorescence can be measurably detected over ambient reflected light.

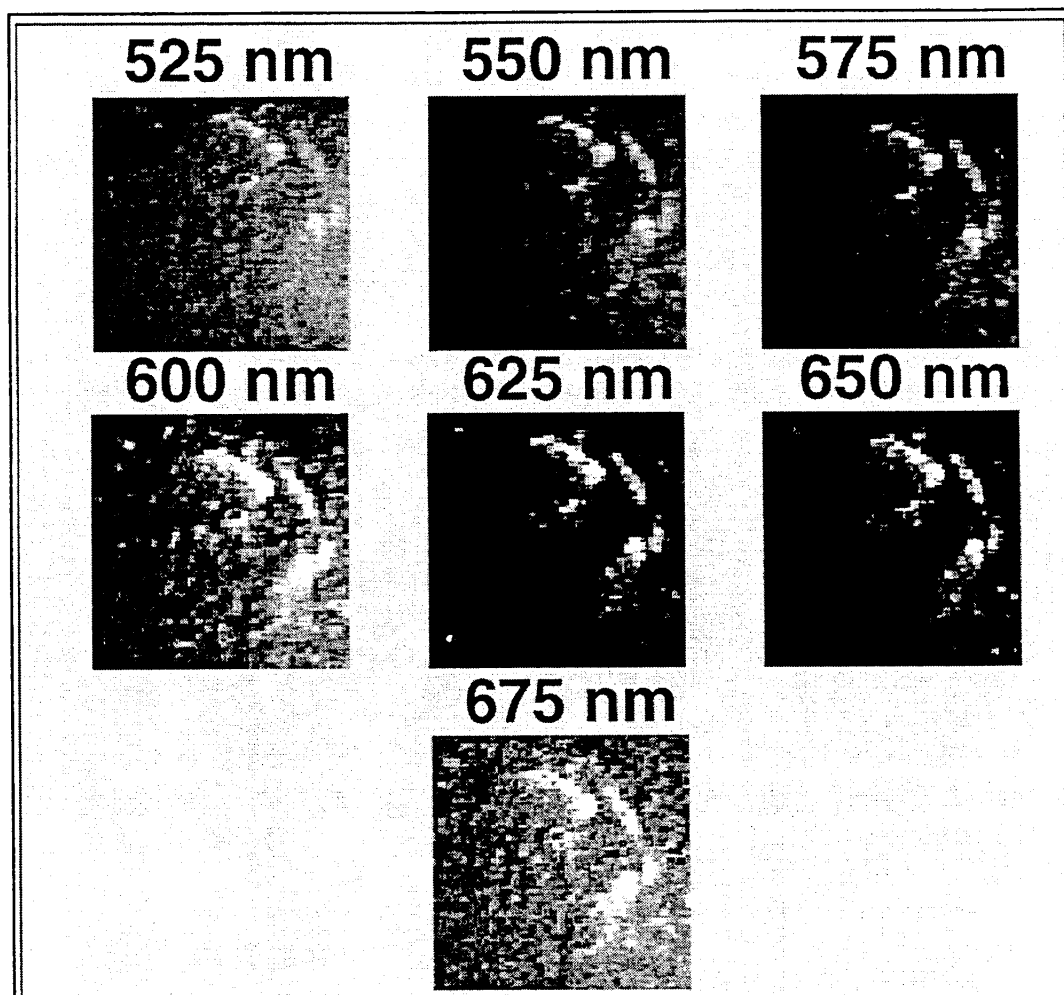
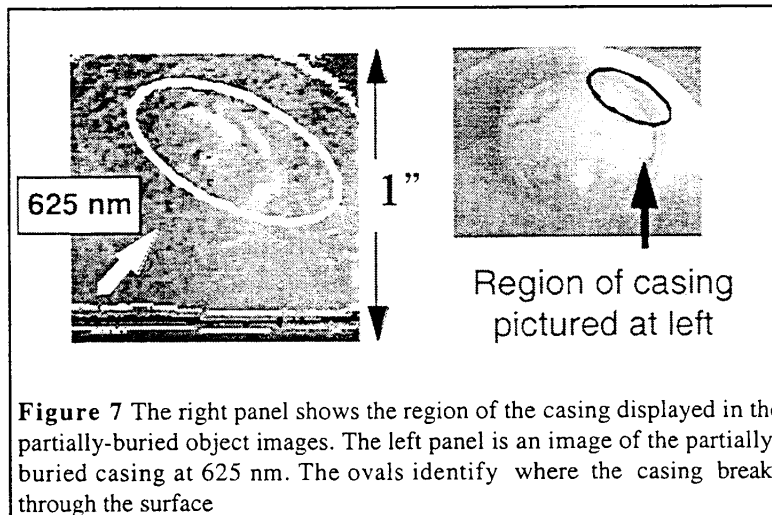
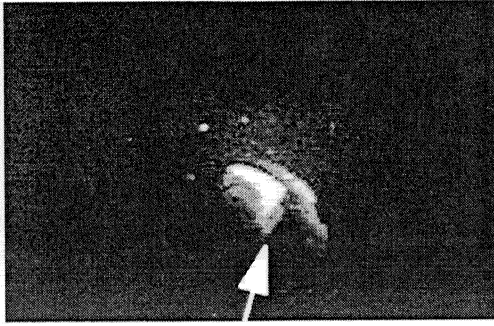


Figure 8 Normalized reflection images of the partially-buried landmine casing at seven wavelengths. The degree of contrast between the sand and casing is greatest in the 625 and 650 nm images.



Exposed portion of casing

Figure 9 Fluorescence image of the partially-buried landmine casing. The sample area was illuminated with the 488 nm line of an Argon laser. A 520 nm long pass filter was used to remove the background.

III. Point-by-Point Scanning in the NIR

In contrast to the direct imaging methods, we also performed a point-by-point reflection-mode scan in a 2-D grid pattern at NIR wavelengths over the surface of sand with wet and dry areas. The idea behind this experiment is that the soil above the buried object may contain less moisture - the flow of moisture being drawn up to the surface (due concentration gradients stemming from surface evaporation) may be affected by the presence of the object. A casing for a landmine was buried under 5 mm of material in a tray that measured 30x16x8 mm (LxWxD). In order to simulate the dry and wet areas found in the field, a piece of plastic wrap was used to separate the dry material from the wet. Using an infra-red emitter/detector matched pair operating at 915 nm, a light-

emitting diode (LED) and a photodetector were mounted side-by-side on perforated board and soldered in place. The orientation was chosen to maximize the detection of reflected IR from the infrared LED. The photodetector has a sensitive area of about 1 mm² while the LED has a collimating lens mounted in place and a radiant power of 0.5 mW. The detector and experiment are depicted in Figure 10.

The photodetector was used in the passive mode and voltages were measured with a digital multi-meter. The device was positioned using an optical rail system marked in mm. The LED and photodetector were held about 30 mm above the surface of the soil and a reading was taken, in a darkened

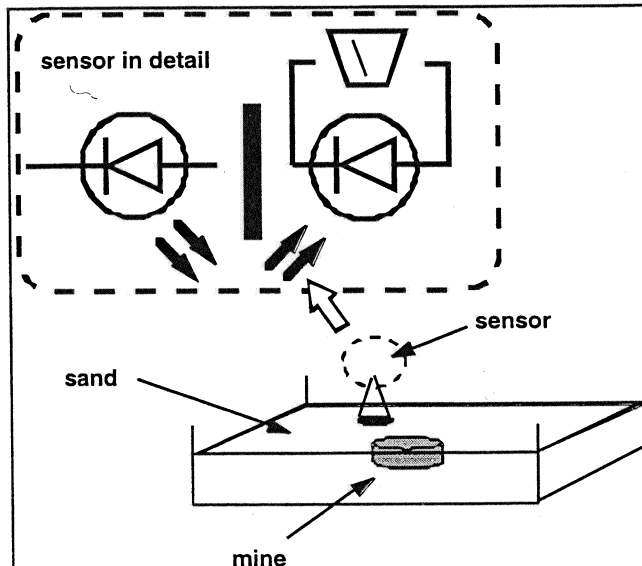


Figure 10 Diagram of the sensor and the experimental arrangement used for the IR scan.

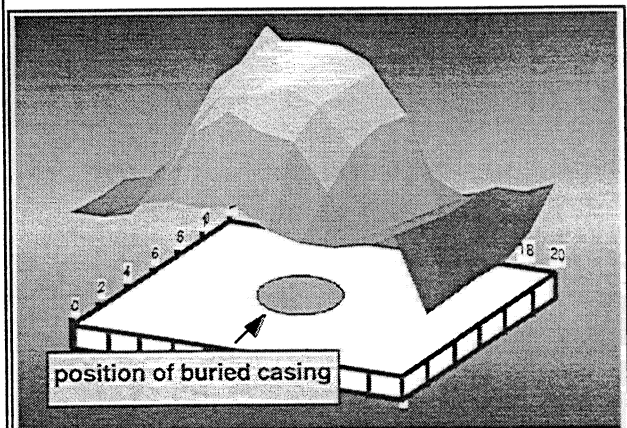


Figure 11 Result for the IR scan of the buried object. The x and y axes are position and the z-axis is the photodiode voltage in linear units.

room, every 20 mm in the x and y direction. The surface scanned in this procedure covers a 12x20 cm area. The resulting image is illustrated in Figure 11.

The experiment provided an image of medium resolution where the wet sand reflected enough radiant energy to produce a signal on the order of 27 - 40 mV and the dry sand that of 50 - 60 mV. The variation of the signal at any given point was about one millivolt. The device was able to differentiate between wet and dry material and produce an image that roughly corresponds with that of the buried landmine. Further work will focus on improving the sensor and more accurately simulating field conditions in the laboratory.

Conclusion

In this paper, we have reported on some preliminary work evaluating methods that could be used in the detection of buried objects. We examined imaging techniques using both reflected and fluorescent light as well as point-by-point scanning with a single infrared emitter-detector pair. Each of the methods shows some promise and will be evaluated under more stringent conditions in the future.

Acknowledgments

This research is sponsored by the US Army MURI Project (US ARMY ARO/DAAG-55-97-1-0013) and by the Office of Biological and Environmental Research, U.S. Department of Energy under contract DE-AC05-96OR22464 with Lockheed Martin Energy Research Corporation.

References

1. F. Moreau, D. M. Hueber, and T. Vo-Dinh, Remote Spectral Imaging System (RSIS) Based on an Acousto-Optic Tunable Filter (AOTF), *Instr. Sci Tech.*, 24 (3), pp. 179-193 (1996).
2. F. Moreau, S. M. Moreau, D. M. Hueber, and T. Vo-Dinh, Fiber-Optic Remote Multisensor System Based on an Acousto-Optic Tunable Filter (AOTF), *Appl. Spectrosc.*, 50 (10), pp. 1295-1300 (1996).
3. A. D. Campiglia, D. M. Hueber, F. Moreau, and T. Vo-Dinh, Phosphorescence Imaging System Using an Acousto-Optic Tunable Filter and a Charge-Coupled Device, *Analytica Chimica Acta*, 246, pp. 361-372 (1997).

SCATTERING at LOW GRAZING ANGLES from LARGE SCALE TWO DIMENSIONAL RANDOM ROUGH SURFACES USING the STEEPEST DESCENT FAST MULTIPOLE METHOD (SDFMM)

W. C. Chew, M. El-Shenawee, V. Jandhyala, and E. Michielssen
Center for Computational Electromagnetics
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
1406 W. Green St., Urbana IL 61801

Tel: 217-333-7309, Fax: 217-333-5962, e.mail: w-chew@uiuc.edu

Introduction and Background

The analysis of low grazing angle (LGA) electromagnetic scattering from two dimensional random rough surfaces is a challenging research problem. In spite of several analytical and numerical techniques that have been developed during the last few years, the problem of LGA scattering from practical random rough surfaces remains unsolved. The memory and CPU requirements of classical computational techniques prohibit the analysis of the LGA scattering phenomena, and the accuracy provided by many of these techniques simply is not adequate to solve this problem.

The objective of this work is to analyze LGA electromagnetic scattering for practical random rough surfaces using an enhanced SDFMM. This technique is a hybridization of the Multilevel Fast Multipole Algorithm (MLFMA) and the Steepest Descent Path (SDP) method. It has been successfully used to analyze electromagnetic scattering from two dimensional quasi planar surfaces. Recently, the SDFMM has been enhanced by parameter selection ruler that provides *a priori* error estimates [1]. The SDFMM dramatically accelerates the iterative solution of the method of moments equations for a large class of structures. When using an iterative solver, both the memory and the computational cost for the SDFMM are of $O(N)$ while they are of $O(N^2)$ for the conventional MOM technique.

For the LGA scattering problem, accuracy is an important factor in any numerical technique because of the shadowing effects that occur at these angles. In the SDFMM, there are tradeoffs between accuracy and both the CPU time and the required memory that can be fully controlled by the proper choice of SDFMM parameters. This makes the SDFMM a prime candidate for solving the LGA scattering problem.

Results and Discussion

A random rough surface is modeled as a finite object chosen large enough such that a practical LGA antenna beam does not excite the surface edges. An example considered in this paper is a surface of $115\lambda \times 28\lambda$ which results in a number of MOM current unknowns equal to 617,096 (using eight unknowns per wavelength). A Gaussian beam (with half beam width equal to 4λ) excites the surface at an angle of 80° from the normal direction. The surface is assumed to be perfect conductor with Gaussian statistics. The roughness parameters are 0.5λ for the rms height and 1.0λ for the correlation length. Preliminary results show that the SDFMM is efficient for analyzing LGA scattering problem.

Reference

- [1] M. El-Shenawee, V. Jandhyala, E. Michielssen, and W. C. Chew, Proceedings of the IEEE APS-URSI '98, pp. 182, Atlanta, GA, June 1998.

GEM-3: A Monostatic Broadband Electromagnetic Induction Sensor

I.J. Won, Dean A. Keiswetter, David R. Hanson
Elena Novikova and Thomas M. Hall

Geophex, Ltd.

ABSTRACT

We have designed, fabricated, and field-tested a new, unique, monostatic, broadband, electromagnetic sensor for subsurface geophysical investigation. The sensing unit consists of a pair of concentric, circular coils that transmit a continuous, broadband, digitally-controlled, electromagnetic waveform. The two transmitter coils, with precisely computed dimensions and placement, create a zone of magnetic cavity (viz., an area with a vanishing primary magnetic flux) at the center of the two coils. A third receiving coil is placed within this magnetic cavity so that it senses only the weak, secondary field returned from the earth and buried targets.

This monostatic configuration has many advantages including (1) compact sensor head, (2) a large transmitter moment, (3) high spatial resolution, (4) no spatial distortion of an anomaly common to bistatic sensors, (5) circular symmetry that greatly simplifies mathematical description, and, therefore, (6) simplified forward and inverse modeling processes. Three prototype GEM-3 units have been built and tested at various environmental sites, including those containing unexploded ordnance and land mines.

Introduction

Based on the data quality and quantity, portability, and non-intrusiveness, the electromagnetic induction (EMI) method is perhaps the most popular method for shallow geophysical exploration. During the past several years, Geophex has developed a new generation of airborne and man-portable EMI sensors. GEM-2, first debuted in early 1995, is a 9-lb hand-held, bistatic EMI sensor that can operate either in the frequency- or the time-domain mode. In the frequency-domain, GEM-2 is field-programmable to operate at simultaneous, multiple frequencies or in swept-frequency mode between 90 Hz and about 24 kHz. The GEM-2 operating principles and case histories are described in Won et al. (1996).

One of the major motives for GEM-3 was to increase lateral and spatial resolution. A bistatic sensor, because of the long path between the source and receiver, is inherently poor in spatial resolution particularly for small, shallow targets. Figure 1 shows GEM-3 in operation as of August, 1996. For this prototype, the electronic console including the data logger is very similar to that of GEM-2. Figure 2 shows the electronic block diagram of GEM-3.

GEM-3 Operating Principle

The GEM-3 sensor contains a pair of concentric transmitter coils and a small receiver coil at the center. We call this concentric geometry a "monostatic" configuration because all coils are essentially co-located. The GEM-3 is a transmitter-bucked sensor, as will be discussed in a later section. All coils are molded into a single, light, circular disk in a fixed geometry, rendering a very portable package. The disk, along

with a handle boom, is made of a Kevlar-skinned foam board. Attached to the other end of the boom is a removable electronic console (fig. 1).

For a frequency-domain operation, the GEM-3 prompts for a set of desired transmitter frequencies. Built-in software converts these frequencies into a digital "bit-stream," which is used to construct the desired transmitter waveform for a particular survey. This bit-stream represents the instruction on how to control a set of digital switches (called H-bridge) connected across the transmitter coil, and generates a complex waveform that contains all frequencies specified by the operator. This method of constructing an arbitrary waveform from a digital bit-stream is known as the pulse-width modulation (PWM) technique.

The base period of the bit-stream for GEM-3 is set to 1/30th of a second for areas having a 60-Hz power supply, as does the U.S. The period is 1/25th of a second at 50-Hz areas, as in Europe and Japan. The GEM-3 H-bridge switches at a rate of 96 kHz and, therefore, the bit-stream contains 3,200 steps within the 1/30-second base period. Any integral number of the base period may be used for a consecutive transmission in order to enhance the signal-to-noise ratio (SNR). For detailed explanation of how the GEM-2 constructs a transmitter waveform and actual examples, refer to Won et al. (1996).

Bistatic and Monostatic Sensors

The terms "bistatic" and "monostatic" may be new to some readers; hence, a brief discussion is appropriate. As a simple analogy, an underground miner who uses a headlamp operates a monostatic sensor because his light source and eyes are essentially co-located. A camera with a built-in flash

Electromagnetic Induction Spectroscopy

I.J. Won, Dean Keiswetter, and Elena Novikova
Geophex, Ltd.

ABSTRACT

An object, made partly or wholly of metals, has a distinct combination of electrical conductivity, magnetic permeability, and geometrical shape and size. When the object is exposed to a low-frequency electromagnetic field, it produces a secondary magnetic field. By measuring the broadband spectrum of the secondary field, we obtain a distinct spectral signature that may uniquely identify the object. Based on the response spectrum, we attempt to "fingerprint" the object. This is the basic concept of *Electromagnetic Induction Spectroscopy (EMIS)*.

EMIS technology can be particularly useful for detecting buried landmines and unexploded ordnance. By fully characterizing and identifying an object without excavation, we should be able to reduce significantly the number of false targets. EMIS is applicable to many other problems where target identification and recognition (without intrusive search) are important. For instance, an advanced EMIS device at an airport security gate may be able to identify a particular weapon according to its maker and type. Other potential applications may include industrial sorting processes and robotics.

Introduction

When an electrically conductive and/or magnetically permeable object is placed in a time-varying electromagnetic field, a system of induced current flows through the object. By observing the small secondary magnetic field emanating from the induced current, we attempt to detect the object; this is the foundation of the time-proven electromagnetic induction (EMI) method. EMI physics is completely described by Maxwell's four equations, although analytical solutions beyond the simplest geometry are rare due to mathematical complexity.

Our main interest in this article is the frequency dependence of the EMI response. By measuring an object's EMI response in a broad frequency band, we attempt to detect and characterize the object's geometry and material composition. We name this potential new technology exploiting the spectral EMI response *Electromagnetic Induction Spectroscopy (EMIS)*.

For low-frequency geophysical applications, the displacement current resulting from the dielectric property may be ignored up to a megahertz and higher. At zero frequency, we observe the induced magnetism from a permeable (i.e., magnetic) object; this is the foundation of the magnetic method. In this sense, the magnetic method is a subset of the EMI method at zero-frequency with the earth's magnetic field as the inducing source.

From numerous surveys that we have conducted using our multifrequency sensors (GEM-2 and GEM-3; Won et al., 1996 and 1997), we have accumulated significant evidence that a metallic object undergoes continuous changes in response as the transmitter frequency changes. For instance, we have commonly noticed that the observed anomalies have

opposite polarities at certain frequencies, depending on whether the target is ferrous or nonferrous. It appears that the phase also depends on the target's geometrical shape. These observations suggest strongly that the EM anomaly measured in a broad band may offer the ability to both detect and identify a target.

Based on elementary EM theory, an object must exhibit different responses at different frequencies. The reason why this fact has not been exploited is due to the lack of practical broadband EMI instruments to study the phenomenon. Most commercial EMI sensors (including common metal detectors) operate at single frequency or, rarely, at a few discrete frequencies. However, with the recent development of the GEM-series sensors, it is now practical to exploit broadband EMI spectra in order to detect and identify the targets.

Because of this spectral limitation, there has been more interest in the "geometrical sounding" method (an expanding transmitter-receiver array at single frequency) than in the "frequency sounding" method (a fixed transmitter-receiver geometry with varying frequencies). We use such sounding methods to obtain enough data to solve for unknown parameters of a preconceived mathematical model. In contrast, EMIS looks for a spectroscopic identity by matching an observed EMI spectrum to a spectrum of known objects.

On the research level, however, there have been many experimental works that studied broadband EMI phenomena. To mention a few, Ryu et al. (1972) measured at 14 discrete frequencies between 200 Hz and 10 kHz with a horizontal loop and measured tilt angle and ellipticity to explore groundwater in California. Ward et al. (1974) tested a similar 14-frequency system between 10.5 Hz and 86 kHz. Won (1980) showed broadband theoretical models and experimental results using a saltwater tank and graphite targets in the labora-

Multifrequency Electromagnetic Signature of the Cloud Chamber, Nevada Test Site

Dean Keiswetter and I.J. Won
Geophex, Ltd.
Raleigh, North Carolina

ABSTRACT

Multifrequency electromagnetic (EM) data acquired over the Cloud Chamber (CC), Nevada Test Site (NTS), clearly delineate the lateral bounds of the facility and detect the line-of-sight pipe - a 0.4-m diameter stainless-steel pipe buried nine meters below ground level. The purpose of the ground-based electromagnetic surveys was to demonstrate that the CC could be detected using the GEM-2, and acquire multi-frequency data for inversion processing and algorithm development. We present herein GEM-2 data acquired at the CC and discuss the importance of wideband EM data.

Introduction

Geophex acquired broadband electromagnetic data over the Cloud Chamber (CC) facility, Nevada Test Site (NTS), during 1995 and 1996 to support a Department of Energy (DOE) funded, multi-year, multi-agency research program to geophysically characterize underground structures. This manuscript presents contoured maps of multifrequency EM data and describes the sensor and acquisition methods. Witten et al. (1997) presents details of multifrequency-based inversion theory, and imaging results of these data.

Cloud Chamber

The CC is located in north central Yucca Flats, an alluvial basin on the NTS that has been used repeatedly for underground nuclear tests (Office of External Affairs, 1988). The shallow geologic strata are composed of Tertiary and Quaternary volcanic and clastic rocks (Dockery et al., 1985).

The CC, built in 1968 as part of a diagnostics experimental program, was used to measure ionized particles from the 7.4-kiloton (KT) HUPMOBILE event (Office of External Affairs, 1988). The quonset hut CC structure is approximately 42 m in length, 10 m wide at its base, and 5 m high at its maximum. Semicircular steel members support a wooden frame as shown in fig. 1. The reinforced-concrete floor of the CC is 9 m below ground surface.

Multifrequency Electromagnetic Approach

The electromagnetic induction method can be used to target different depths of interest by changing either the spacing between transmitter and receiver coils, or the frequency of the transmitted field (Patra and Mallick, 1980; Won, 1980). The first method is known as geometrical sounding and involves recording data using several transmitter-receiver coil spacings centered over fixed location; the depth of exploration increases with the coil spacing. The two coils systems, although typically connected by an umbilical cord, are physically separate and require two or more field operators.

The second method is known as frequency sounding and involves changing the transmitter frequency, but keeping the transmitter-receiver separation constant (fig. 2). The depth of exploration (also called the depth of penetration or skin depth) is mainly determined by the source frequency and ground conductivity (Won, 1980). In simple terms, low frequency signals travel further than high frequencies through conductive media and, thus, detect deep structures. Conversely, high frequencies detect shallow features better than low frequencies. Broadband frequency sounding is therefore analogous to depth sounding and can be used to create a pseudo 3-D subsurface image (Won, 1983) or for 3-D conductivity inversion processing (Lee et al., 1987; Witten et al., 1997).

Frequency sounding possesses inherent advantages over geometric sounding for depth imaging because it is logistically and operationally simple (due to a one-person operation) and improves spatial resolution (by averaging vertically instead of laterally; fig. 2). Theoretical and practical discussions on these methods may be found in Grant and West (1965), Keller and Frischknecht (1966), and Kaufman and Keller (1983).

GEM-2 Sensor and Data Acquisition Methods

We acquired broadband electromagnetic data using the GEM-2, (fig. 3; Won et al., 1996), a multifrequency induction sensor developed at Geophex. The GEM-2 contains a transmitter, receiver, and bucking coil in a single lightweight, portable, and compact package. The signal-processing electronics, analog circuitry (for signal conditioning), and power supply are housed in a removable console.

The primary field generated by the GEM-2 sensor is controlled via a digitally constructed waveform. The transmitted signal can vary from a complex waveform containing any number of frequencies (user-defined) to a simple square wave. Although the design of the GEM-2 allows it to operate in the either the time- or frequency-domain, only frequency domain data were recorded during this project. The maximum current

Comparison of Magnetic and Electromagnetic Anomalies Caused by Underground Structures

I.J. Won and Dean Keiswetter
Geophex, Ltd.
Raleigh, North Carolina

ABSTRACT

Magnetic and electromagnetic (EM) methods are perhaps the most convenient and popular geophysical survey methods for detecting buried manmade objects, which is due to their non-intrusiveness, light field logistics, high survey speed, and the quality of information. One should always consider the two methods as the precursor to any geophysical survey. Often, the data resulting from the two methods are sufficient for characterizing buried objects.

In this paper, we present magnetic and EM data collected at four sites: (1) Cloud Chamber at the Nevada Test Site, (2) Anacostia Metro Tunnels in Washington, D.C., (3) Cold Test Pit at Idaho National Engineering Laboratory, and (4) Unexploded ordnance site at Jefferson Proving Ground, Indiana. The first two sites may be considered typical underground facilities. The last two sites, however, contain small buried objects (storage tanks, ordnance, etc.) specifically prepared to test various geophysical methods for detection and, possibly, discrimination.

We find through these and numerous other comparisons that broadband EM data are superior to magnetic data in terms of the amount and the quality of information. The monopolar EM anomaly is invariably easier to interpret, and thus can locate a buried target more accurately than the dipolar magnetic anomaly. In addition, the EM method senses both electrically conductive and magnetically permeable targets. In contrast, the magnetic method responds only to permeable, or ferrous, metals. In that sense, the magnetic method should be considered a subset of the EM method, or a special "passive" EM method at zero frequency.

Introduction

From the viewpoints of non-intrusiveness, field logistics, and survey speed, the magnetic and electromagnetic (EM) methods stand far ahead of other geophysical techniques, including gravity, electrical resistivity, seismic refraction or reflection, and ground-probing radar (GPR).

Intrusiveness: Magnetic and EM sensors are commonly man-portable. The methods do not require ground contact (e.g., seismic and electrical resistivity) or demand a stabilized survey platform (e.g., gravity). The GPR is sensitive to ground roughness, particularly at high frequencies.

Field Logistics: Portable magnetic and EM sensors are usually light in weight, simple to operate, and consume relatively little electrical power. The methods do not involve an array of sensors connected by tangled cables and wires, as would be common for seismic, electrical resistivity, and GPR methods.

Survey Speed: Magnetic and EM sensors can operate continuously in time and in motion. With high-capacity data-logging electronics common to modern portable sets, the sensors can collect over tens of thousand survey points per hour. For a typical high-resolution survey, this translates to an areal coverage of one acre (0.4 hectare) or more per hour. The data can be downloaded to a portable computer and the survey results can be viewed, typically in a color contour format, on

site within a few minutes.

Based on these advantages, one should always consider conducting magnetic and EM surveys as the precursor to any ground geophysical surveys for detecting buried objects, including underground facilities. If nothing else, they are quick and easy and provide a first geophysical view of the site. Quite often, the two methods provide enough data to proceed with planned site activities.

Most manmade objects, except for plastics, contain metals that are easy to detect by the magnetic and/or EM method. Quite often, the two methods can be complementary. In this article, we compare the viability of the two methods and present example data collected over known buried objects. We limit our discussions in this article to ground-level geophysical surveys, although similar arguments can be made for airborne geophysical surveys.

Nature of Magnetic Anomalies

Whether induced or permanent, a buried, ferrous object exhibits basically a dipolar magnetic source. Any object may be represented as a sum of magnetic dipoles having various magnetic moments and dipole orientations. The surface manifestation of this source on a survey plane (i.e., the ground) can be varied according to the source distribution, topography, and geomagnetic latitude.

Modeling the effects of non-uniform soil moisture on detection efficacy of mine-like objects with GPR

Carey Rappaport, Scott Winton, Dongping Jin, and Lori Siegel
Center for Electromagnetics Research
Northeastern University
Boston, MA

ABSTRACT

We present dispersive soil finite difference time domain wave propagation simulations to study the effects of realistic multilayered soil moisture on modulated short pulse Ground Penetrating Radar mine detection exciations. Our conclusions suggest the amount of water and application time -- as a function of soil hydraulic conductivity -- that is necessary for modifying the ground to most effectively detect non-metallic anti-personnel mines.

Keywords: GPR, mine detection, dielectric target imaging.

1 INTRODUCTION

There is considerable interest in detecting buried dielectric mines by modifying the electromagnetic characteristics of soil. It has been suggested that since wetting the ground increases its effective dielectric constant and therefore also increases the dielectric contrast relative to non-metallic mines, this modification may improve detectability.

When water is applied to the ground, the soil becomes saturated on the surface, but remains relatively dry below the wetting front. The wetting front descends with continual watering with time. If insufficient water or time is used in the soil environment modification, there is a multi-layer effect to the electromagnetic scattering of incident radar pulses. In particular, with a rough ground surface, a layered soil moisture content profile may make the target detection problem more difficult than for a uniform, dry soil volume.

2 DISPERSIVE MEDIA FDTD

The finite-difference time-domain (FDTD) method [1] is a straightforward means of modeling wide-band wave propagation in inhomogeneous media. For wave propagation in most real materials, the constitutive parameters are frequency dependent. This dispersion -- which is just a multiplication in the frequency domain -- become a convolution operation in the time domain. Dispersion can be modeled in the time domain in three main ways: 1) Discrete convolution of constitutive relation $D(t) = \epsilon(t) * E(t)$ for Debye and Lorentz models of complex permittivity [2]; 2). Discretization of a differential equation approximation relating $D(t)$ to $E(t)$ [3]; and 3) Direct discrete modeling using the Z-transform of either constitutive relation $J(\omega) = \sigma(\omega)E(\omega)$ or $D(\omega) = \epsilon(\omega)E(\omega)$ [4,5].

For soil propagation problems, the third method, modeling conductivity, has been found to be most suitable. By expressing the soil conductivity as a ratio of polynomials in Z-domain, the conversion to discrete time domain is quite straightforward, since factors involving Z^{-1} correspond to sample time delays. We can approximate it using the single pole approximant[6].

$$\frac{J(Z)}{E(Z)} = \sigma(Z) = \frac{b_0 + b_1 \cdot Z^{-1} + b_2 \cdot Z^{-2}}{1 + a_1 \cdot Z^{-1}} \quad (1)$$

By setting $Z = \exp(i\omega t)$, the real part of Eqn. (1) can be fit to experimentally measured soil conductivity values. The imaginary part of Eqn. (1), divided by $\omega\epsilon_0$ added to an average dielectric constant parameter ϵ_{Av} is then fit to measured dielectric constant. This model can be compared to the standard Debye model, which consists of sums of terms with single poles. A two-term Debye model has been fit to Puerto Rican clay loam with 1.2 g/cc density for three typical moisture levels [3]. This soil has measured conductivity that varies by almost one order of magnitude across two decades of frequency 30 to 3840 MHz [7]. A comparison of the modeled dielectric constant and conductivity for the two different models is shown in Figure 1.

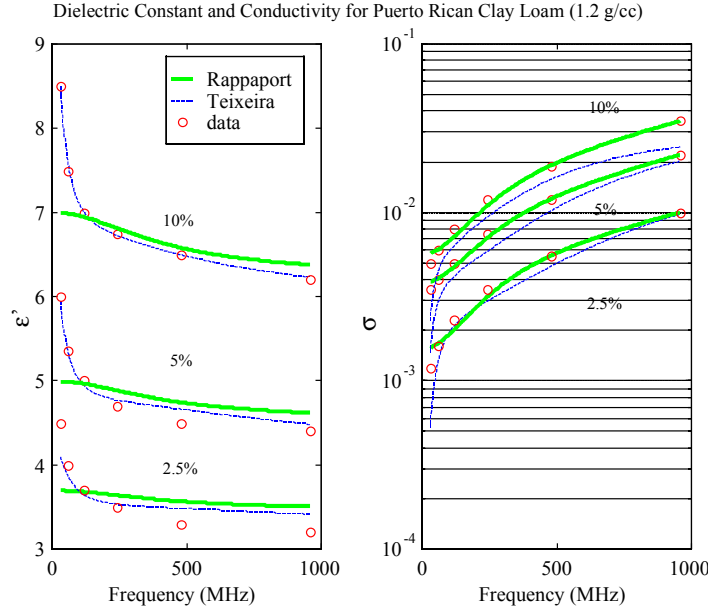


Figure 1 Puerto Rican clay loam, density = 1.2g/cc

Although the Debye model fit is better than that of the Z-transform model for the dielectric constant, the latter is a closer match for conductivity. A more important performance metric than these electrical parameters is the complex wave number, k , which specifies wave velocity and decay rate. The real and imaginary parts of k are shown in Figure 2. Note that the agreement between the values corresponding to measured decay rate α and the Z-transform model for α is far superior to that of the Debye model. This is particularly surprising since the former specifies only one pole, while the two-term Debye model has two poles, and thus requires more storage and computation.

Table 1: Puerto Rican clay loam, density = 1.2g/cc

Moisture (%)	2.5	5	10	20
b_0	0.2401	0.6344	0.9162	1.1482
b_1	-0.4336	-1.1647	-1.6766	-2.0672
b_2	0.1937	0.5308	0.7611	0.9218
a_1	-0.88	-0.88	-0.88	-0.88
ϵ_{Av}	2.18	4.39	8.74	11.84

While the measured soil data gives important information at particular moisture levels, it is necessary to interpolate model parameter values between measured levels to accurately represent realistic continuously varying wetting profiles. Using the best fit values for b_0 , b_1 , b_2 , a_1 , and ε_{Av} for the specified moisture levels – presented in Table 1 – a simple set of interpolating functions have been generated. We interpolate using the function: $b_i = p_i \ln m + q_i$, $i=0,1,2$, with the goal to find values for p_i and q_i [6].

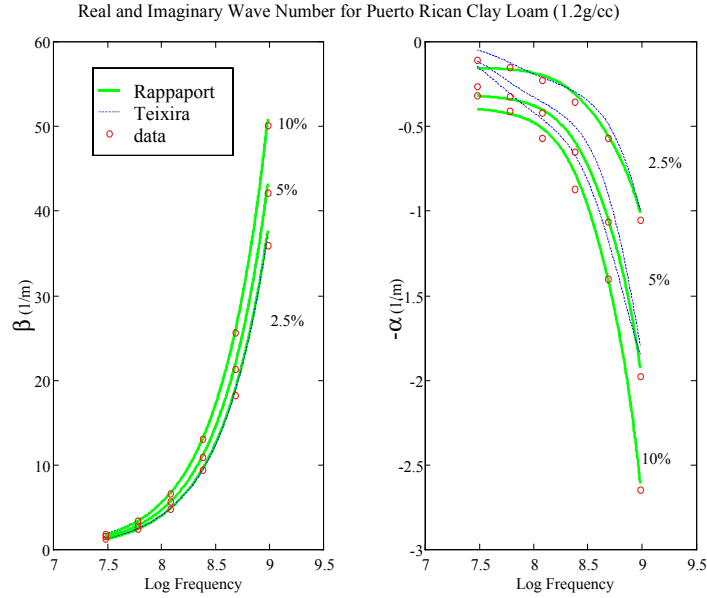


Figure 2 Puerto Rico clay loam, density = 1.2g/cc

The interpolating functions are shown in Figure 3.

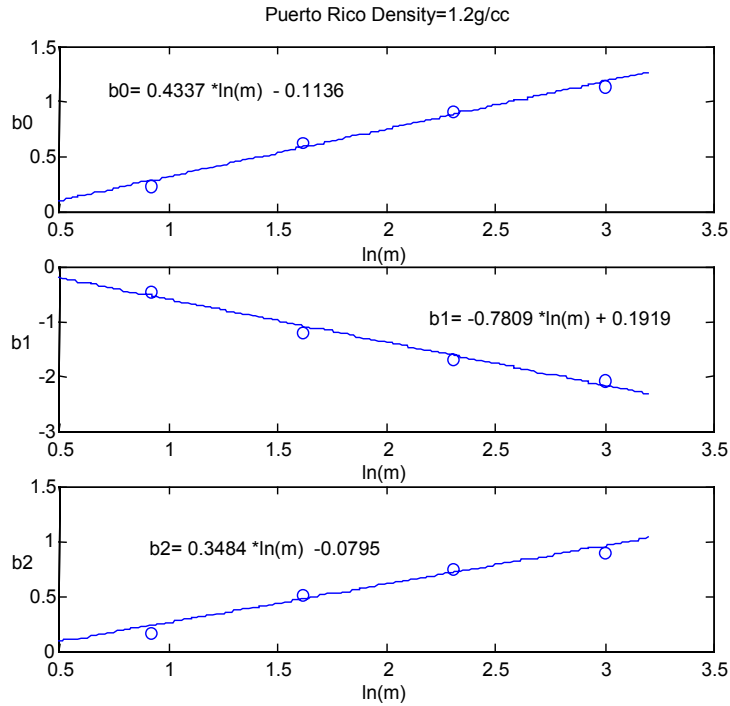


Figure 3. Continuous moisture variation parameter functions.

To model the effect of non-uniform watering of dry soil, we first determine the wetting profile. Water movement in a vertical column of a medium is described by the advection-dispersion equation in the z -direction, as [8]:

$$\frac{d\Theta}{dt} = \frac{d}{dz} \left(D(\Theta) \frac{d\Theta}{dz} \right) + \frac{d}{dz} K(\Theta) \quad (2)$$

Where z is the depth of the water column, $\nu(z)$ is the moisture content, D is the dispersion coefficient of water, K is the hydraulic conductivity of the soil (a measure of the rate of water to pass through). By numerically solving Eqn. (2) we can plot the moisture profile as a function of depth and time.

Figure 4 shows profiles for 0.1, 1, 2, 3, 4, and 5 minutes of watering for soil with saturation hydraulic conductivity of 0.2. From these profiles and the parameter functions derived above, the material characteristics of the wave simulation can be specified to arbitrary accuracy.

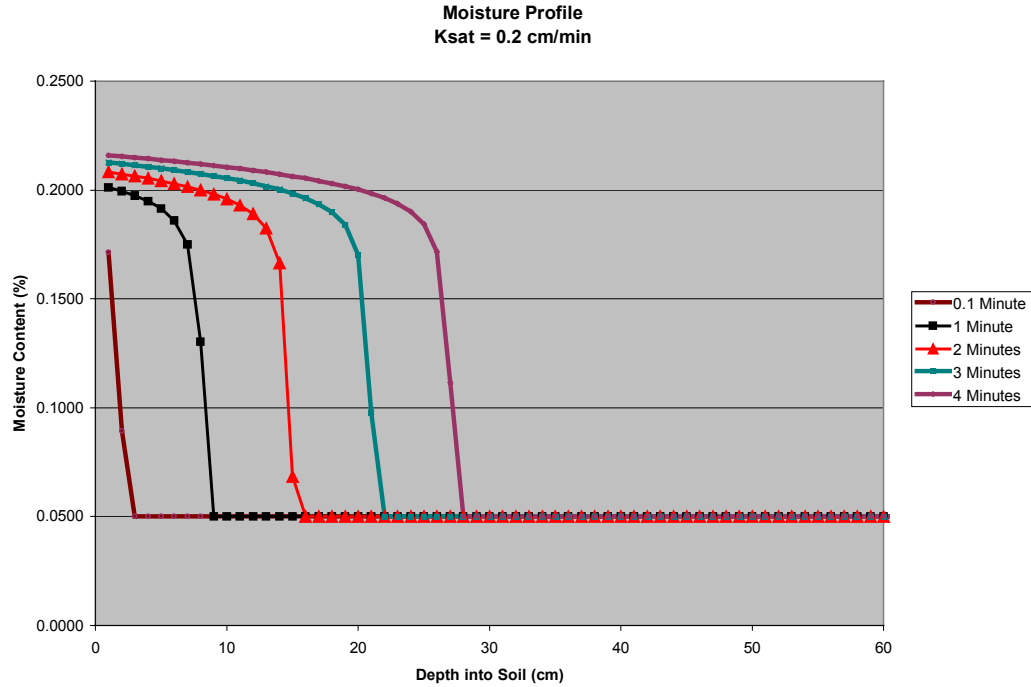


Figure 4. Moisture profile for several watering duration times.

3 SCATTERING SIMULATION

Using the dispersive FDTD code along with the model for continuously varying moisture, we have simulated the two dimensional TE (single longitudinal magnetic field component into the plane of calculation) scattered field from non-metallic mine-like targets for three different cases: 1) constant moisture at 5% (by weight), typical of the residual background moisture in clay loam; 2) constant 20% moisture, representative of clay loam after a long steady rain; and 3) a continually varying moisture profile from 20% down to 5% for the one minute profile as described above. The geometry of the test cases is shown in Figure 5, with a two-dimensional 1GHz modulated gaussian point source 30 cm above the nominal ground surface (which can be chosen to be either planar or rough), and a 10 cm by 5 cm plastic mine-like target buried 10cm below the ground surface. The entire computational grid is terminated with a soil-tuned Perfectly Matched Layer absorbing boundary condition [9].

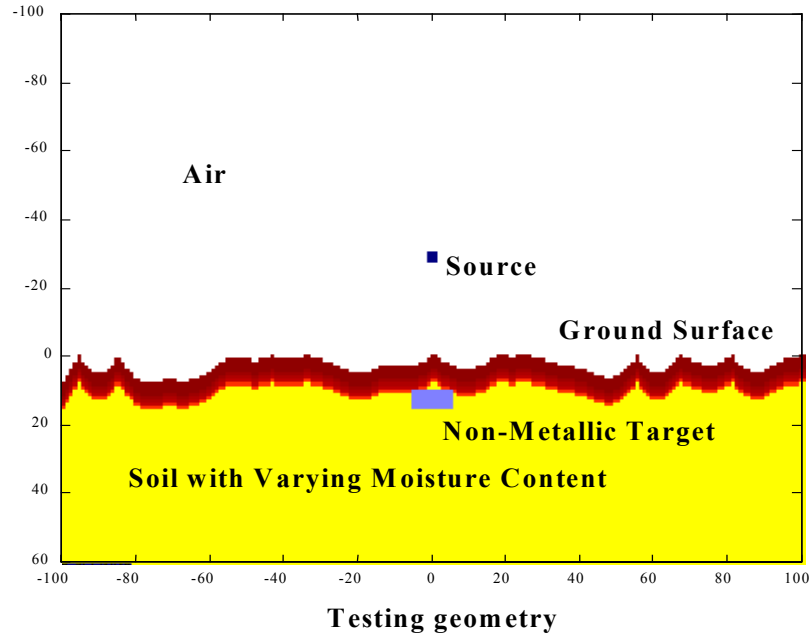


Figure 5 FDTD scattering geometry

Figure 6 shows the total field with target, the incident field with no target, and the scattered field due just to the target, for a flat ground surface with constant residual moisture at a time when the initial circular wave has propagated out past the lattice boundary. The first image is representative of the actual field available to be measured when the mine is present. Clearly, only the field above the ground can be measured, but the field distribution below the ground surface around the mine gives information about how the surrounding environment effects the target scattering. The second image would in general not be available, since it represents the given environment after the mine has been removed. The third image represents the field scattered solely by the mine-like target. It is the job of the mine detection algorithm to recognize the features of the third image in the first image, by ignoring the fields scattered by the soil boundary of the second image. For the 5% constant background soil moisture with planar boundary, picking out these characteristics is relatively easy.

For uniformly wet soil, the dielectric contrast of the non-conducting, non-metallic mine target with the surrounding background is greater than for the previous dry soil case. The target scattering is stronger, and, since the wave velocity is slower in wetter soil, the delay between ground surface scattering and target scattering is greater. The target should thus be easier to detect. Figure 7 shows this case with 20% constant moisture and a planar boundary. It is apparent that the scattered amplitude in the third image is greater, and that it appears later compared to the ground scattered wave. The longer delay makes it easier to separate the target scattered signal from the surface signal.

A more realistic moisture model for a manually applied water scenario is considered in generating Figure 8. For this case, the previously developed continuously varying moisture model causes a stronger surface scattered signal, but weaker target scattered signal. Also, the target delay is not as great as for the uniform wet soil. In fact, this realistic wetting situation combines the worst aspects of the two previous constant cases: the ground clutter is greatest, while the target signature is weakest.

These three cases are repeated in Figures 9-11 for a rough ground surface. With this configuration, the ground clutter confuses the target signal considerably. For 5% moisture, Figure 9, the short delay between the target scattered wave and the ground scattered signal makes it more difficult to distinguish the two. In

Figure 10, the stronger scattered wave from the ground, and the subsequent multiple scattering effects of local surface perturbations makes the target signal much harder to recognize. For the realistic moisture profile, Figure 11, it is almost impossible to separate the target scattered field from the ground scattered field in the total field image at the top.

4 CONCLUSIONS

The effects of non-uniform wetting of soil on the scattered field of small buried non-metallic mine-like targets has been explored. For smooth, flat ground surfaces, there are target detection advantages to both uniformly wet and dry soil backgrounds. However, when water is added to dry soil in an attempt to enhance the dielectric contrast, if insufficient water is applied, the ability to detect target signals is degraded. With rough ground surfaces, wetter, high dielectric contrast soil contribute more surface scattering clutter, which may overwhelm the larger target signal. For continuously varying soil moisture on a rough surface, the target signal is much harder to recognize than for either uniformly moist soil background. The main conclusion is that adding water will only enhance target detectability if the ground is saturated for several minutes.

5 ACKNOWLEDGEMENTS

The authors gratefully acknowledge support from Army Research Office MURI grant DAAG55-97-1-0013.

REFERENCES

- [1] K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media", *IEEE Trans. Antennas Propagat.*, Vol. AP-14, pp. 302-307, 1966.
- [2.] Luebbers, R., Hunsberger, F., Kunz, K., Standler, R., and Schneider, M., "A Frequency-Dependent Finite Difference Time Domain Formulation for Dispersive Materials," *IEEE Trans.on Electromag. Comp.*, pp. 222-227, Vol.32, March 1990.
- [3]. F. Teixeira, W. Chew, M. Straka, M. Oristaglio, and T. Wang, "Finite Differece Time Domain Simulation of Ground Penetrating Radar on Dispersive Inhomogeneous, and Conductive Soils," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 36, no. 6, November 1998, pp. 1928—1936.
- [4] C. M. Rappaport and W. H. Weedon, "A general method for FDTD modeling of wave propagation in arbitrary Frequency dispersive media," *IEEE Transactions on Antennas and Propagation*, " March, 1997
- [5]. C. M. Rappaport, S. C. Winton, "Modeling dispersive soil for FDTD computation by fitting conductivity parameters," *12th Annual Review of Progress in Applied Computational Electromagnetics Symposium Digest*, March 1996, pp. 112--118.
- [6] C. M. Rappaport, S. Wu, S. C. Winton, "FDTD wave propagation in dispersive soil using on a single pole conductivity model," accepted for publication in *IEEE Transactions on Magnetics*, 1999.
- [7.] Hipp, J., "Soil Electromagnetic Parameters as Functions of Frequency, Soil Density, and Soil Moisture," *Proceedings of the IEEE*, Vol. 62, pp. 98—103, January 1974.
- [8] R.L. Bras, *Hydrology: an Introduction to Hydrological Science*, Addison-Wesley: Reading, 1990.C. M.
- [9.] Rappaport, "Interpreting and improving the PML absorbing boundary condition using anisotropic Lossy Mapping of space," *IEEE Transactions on Magnetics*, vol. 32, no. 3, pp. 968-974, 1996.

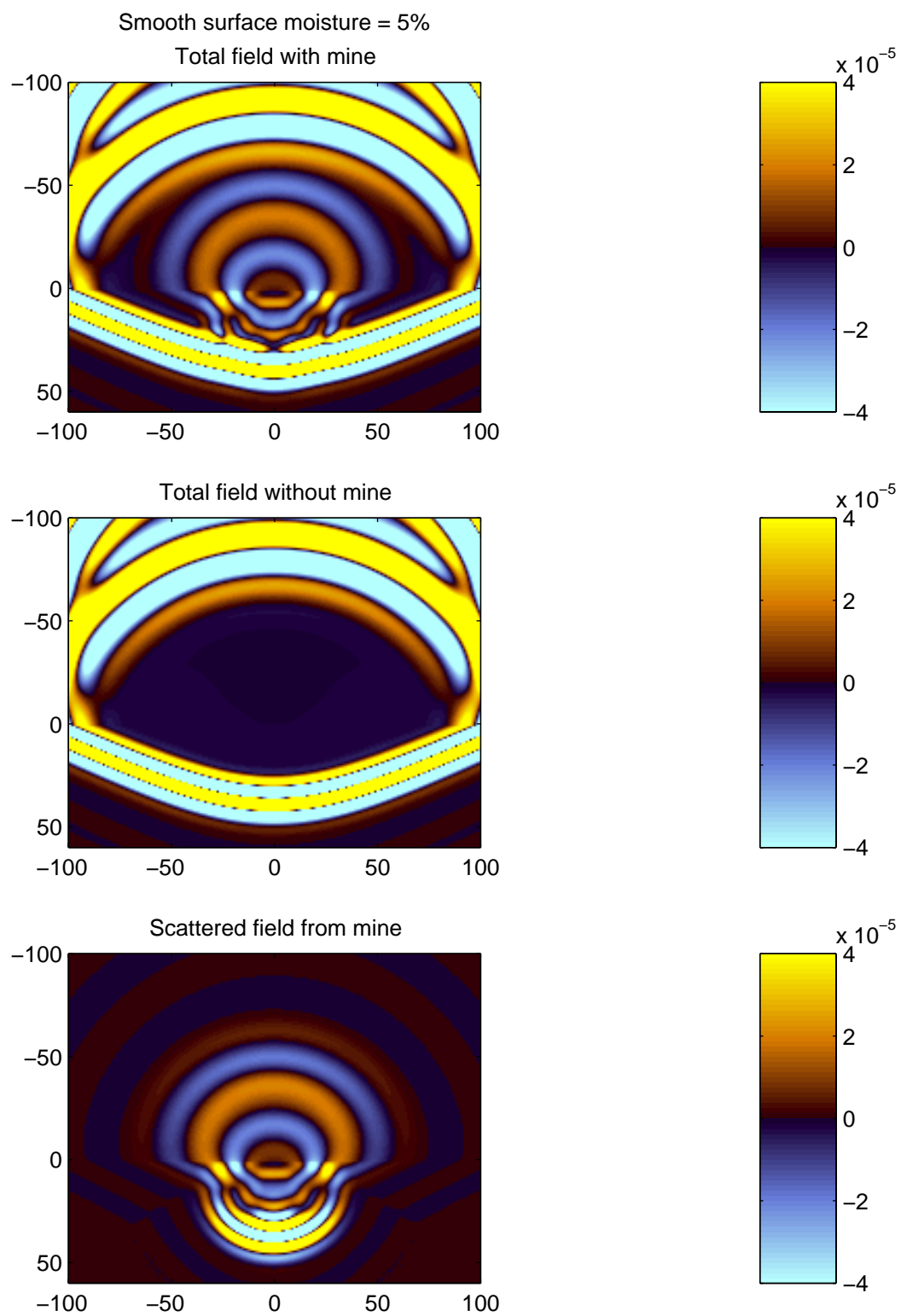
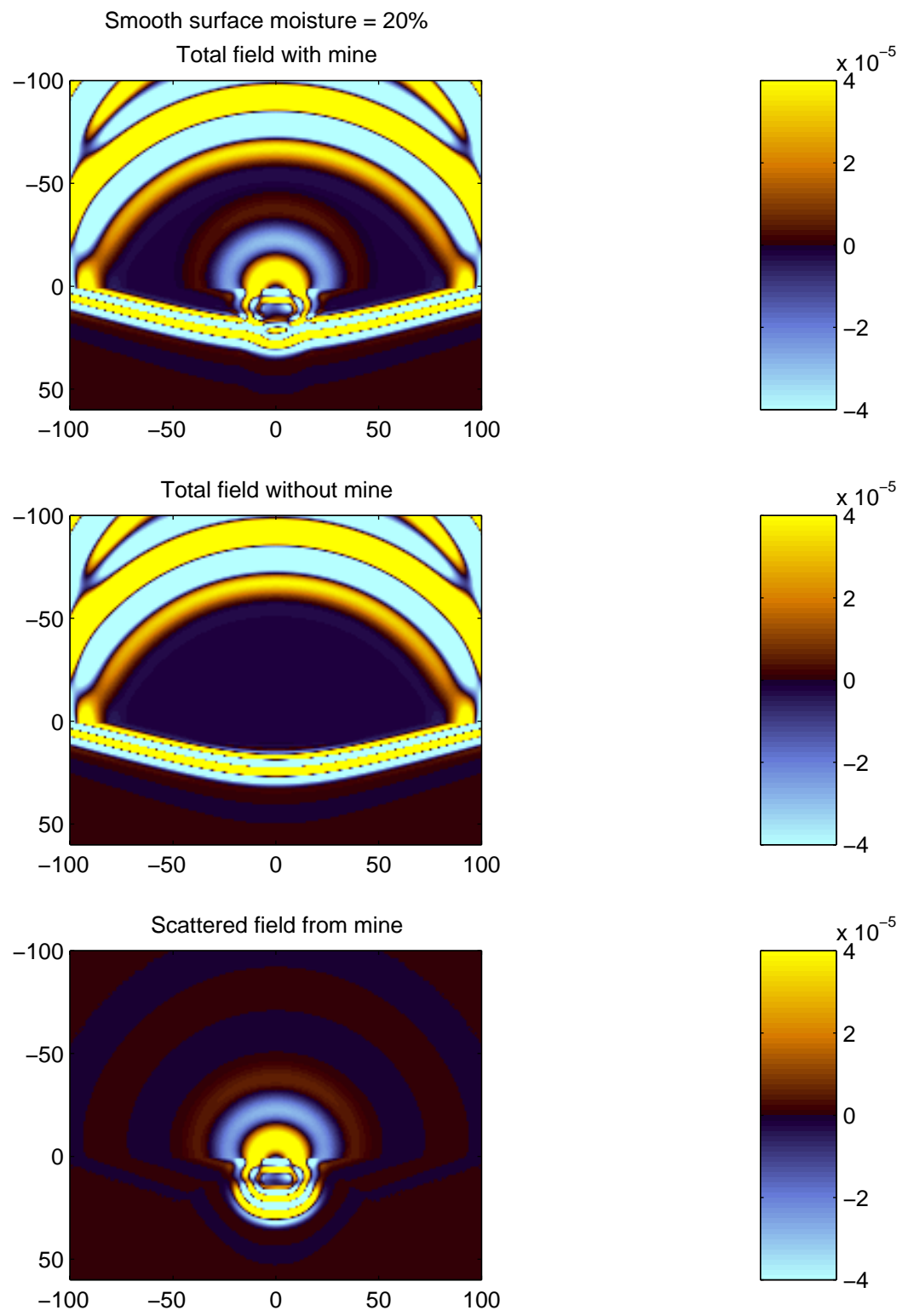


Figure 6



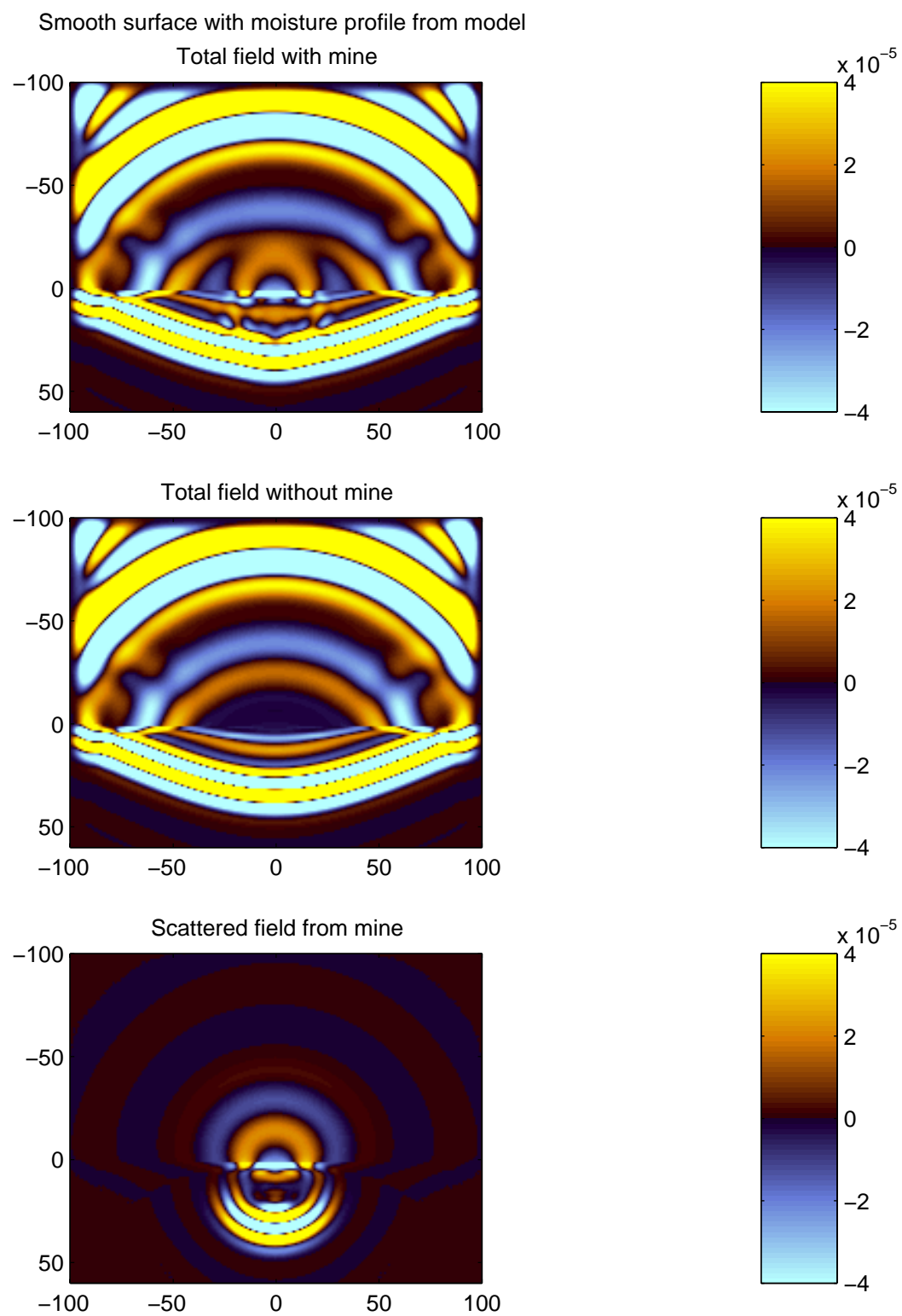


Figure 8

Rough surface moisture = 5%

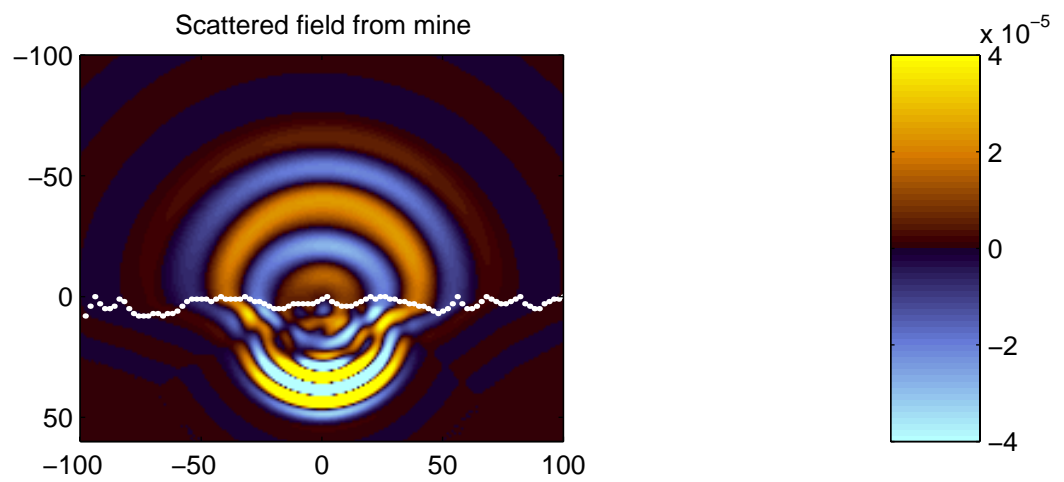
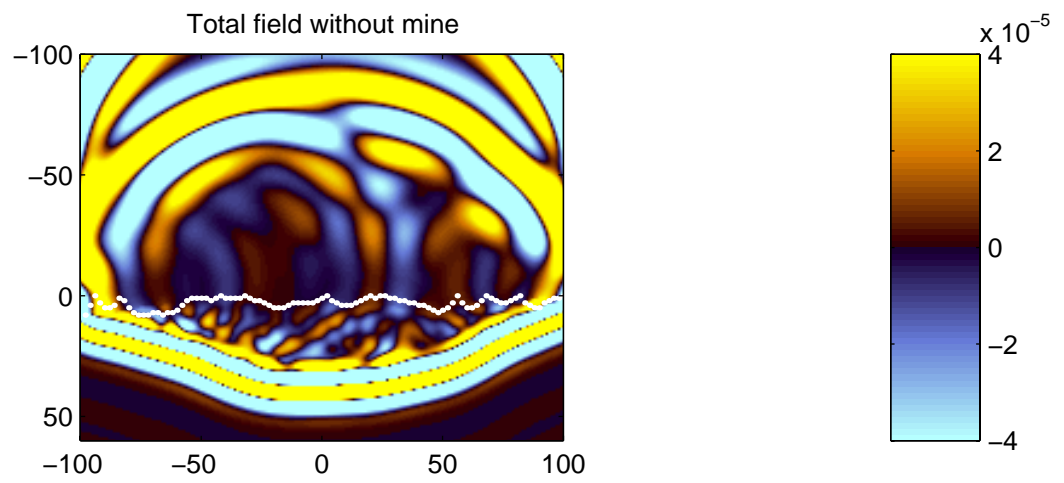
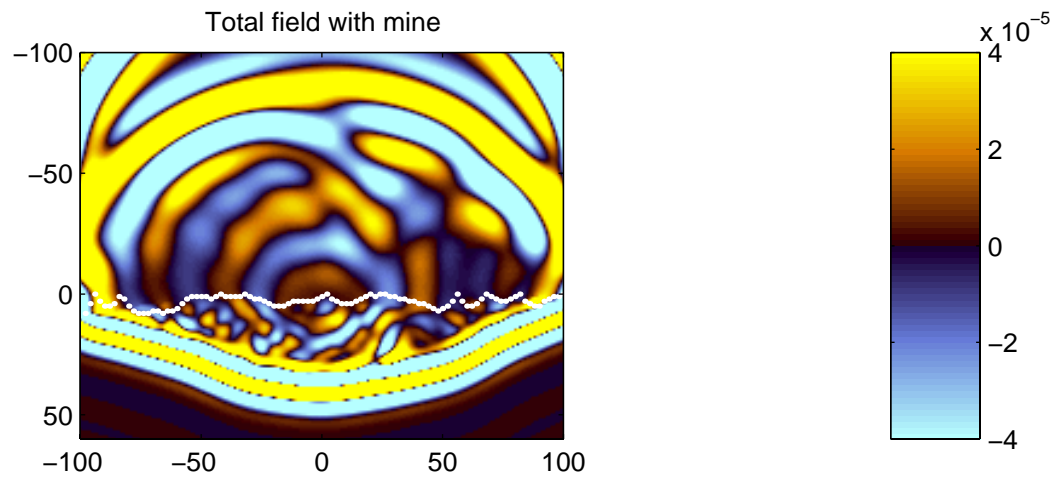
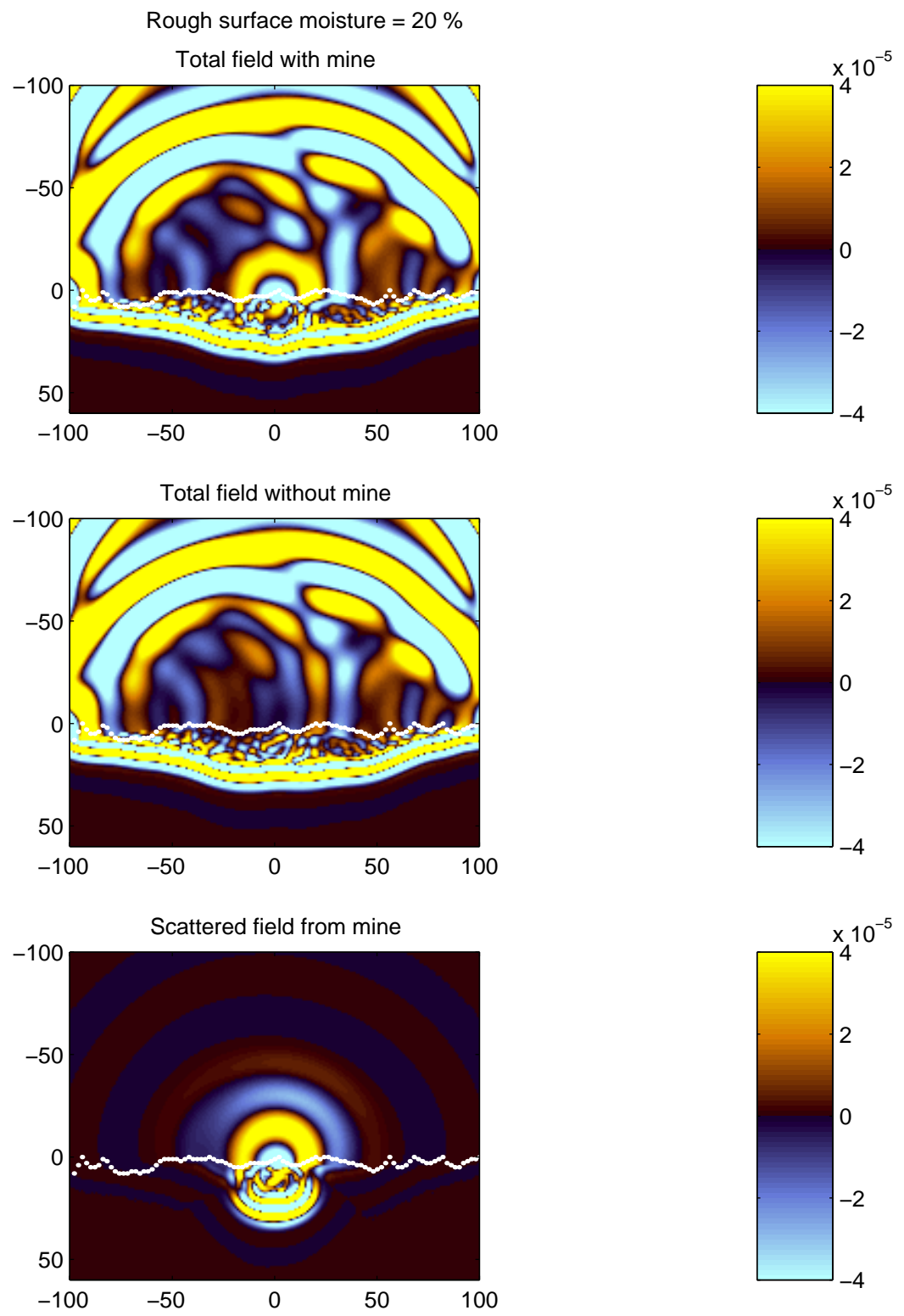
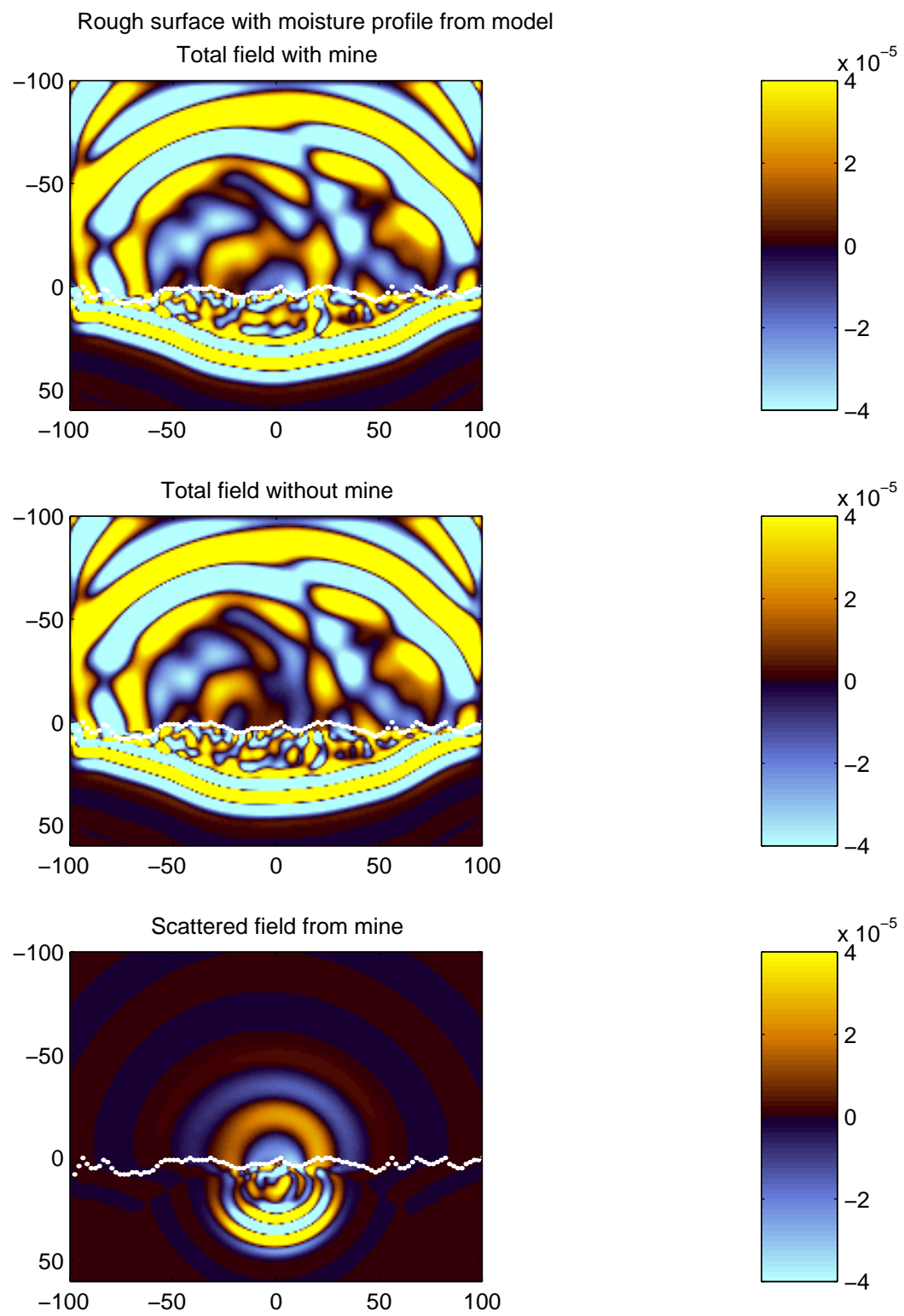


Figure 9





Multiscale Methods for the Segmentation and Reconstruction of Signals and Images

Michael K. Schneider, *Student Member, IEEE*, Paul W. Fieguth, *Member, IEEE*, William C. Karl, *Member, IEEE*, and Alan S. Willsky, *Fellow, IEEE*

Abstract—This paper addresses the problem of both segmenting and reconstructing a noisy signal or image. The work is motivated by large problems arising in certain scientific applications, such as medical imaging. Two objectives for a segmentation and denoising algorithm are laid out: it should be computationally efficient and capable of generating statistics for the errors in the reconstruction and estimates of the boundary locations. The starting point for the development of a suitable algorithm is a variational approach to segmentation [1]. This paper then develops a precise statistical interpretation of a one-dimensional (1-D) version of this variational approach to segmentation. The 1-D algorithm that arises as a result of this analysis is computationally efficient and capable of generating error statistics. A straightforward extension of this algorithm to two dimensions would incorporate recursive procedures for computing estimates of inhomogeneous Gaussian Markov random fields. Such procedures require an unacceptably large number of operations. To meet the objective of developing a computationally efficient algorithm, the use of recently developed multiscale statistical methods is investigated. This results in the development of an algorithm for segmenting and denoising which is not only computationally efficient but also capable of generating error statistics, as desired.

Index Terms—Denoising, multiscale statistical models, segmentation.

I. INTRODUCTION

MUMFORD and Shah have developed a theoretical framework in which to address the problem of simultaneous image denoising and segmentation [2], [3]. In this framework, the goal is to decompose a given noisy image into piecewise smooth regions bounded by contours on which the image intensity is allowed to change abruptly. This is accomplished by minimizing a particular functional jointly over the image boundaries and a reconstruction of the image. The minimizing reconstruction of the original image has been denoised through smoothing everywhere except along the detected boundary contours. The

Mumford and Shah functional has many nice mathematical and psychovisual properties [4]; however, it is difficult to compute minimizers because of the discrete nature of the image boundary terms.

Various people have tried to address these computational difficulties by making small alterations in the Mumford and Shah framework [5]–[7]. In this revised setting, there are also two objects computed by the algorithm: a reconstruction of the original image and a continuous-valued edge-strength function. The reconstruction is a denoised version of the unprocessed, noisy image which does not suffer from the edge blurring effects of some simple linear reconstruction algorithms (e.g., low-pass filtering). The edge-strength function provides information about the optimal spatially-varying amount of smoothing that should be applied to produce the reconstruction. The function varies between the values 0 and 1, taking on the value 1 where no smoothing should be done and 0 in areas where full smoothing is performed. While the edge-strength function itself is not an explicit estimate of image edges, it is demonstrably a more robust indicator of edge likelihood than standard gradient maps. In particular, the edge-strength function displays substantial robustness to noise and automatically avoids problems of dynamic range exhibited by gradient maps of noisy images (see Section VI). As a consequence, generating explicit edge contours from edge-strength functions can be accomplished robustly as shown in [1] using thresholding and [8] using curve evolution.

The approach to segmentation and denoising taken in this paper begins with a novel Bayesian interpretation of the revised Mumford and Shah variational approach to segmentation. This interpretation follows in the footsteps of some recent work in which close connections are made between certain variational and statistical approaches to image processing [9], [10]. The principal advantage of the Bayesian framework is that it provides a theoretical structure for the interpretation and computation of error statistics. Error statistics provide a quantitative measure of the quality of the reconstruction and estimate of the edge-strength function. Such a measure of quality is very important in certain scientific applications. One of the main contributions of this paper is the development of a segmentation algorithm which produces not only a reconstruction of the image and an estimate of the edge-strength function but also error statistics. Furthermore, this paper contains a careful evaluation of the nature and quality of the information provided by these error statistics.

While the Bayesian interpretation can be equally well applied to both one and two-dimensional (2-D) signals, there is a significant difference in computational complexity in solving the

Manuscript received January 13, 1997; revised August 25, 1999. This material is based upon work supported by a National Science Foundation Graduate Research Fellowship, by ONR under Grant N00014-91-J-1004, by AFOSR under Grants F49620-98-1-0349 and F49620-96-1-0028, by Boston University under Grant GC123919NGN, and by NIH under Grant NINDS 1 R01 NS34189. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jeffrey J. Rodriguez.

M. K. Schneider and A. S. Willsky are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: mikesch@mit.edu).

P. W. Fieguth is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, Ont., Canada N2L 3G1.

W. C. Karl is with the Department of Electrical and Computer Engineering and the Department of Biomedical Engineering, Boston University, Boston, MA 02215 USA.

Publisher Item Identifier S 1057-7149(00)01512-8.

resulting problems. As a consequence, a thorough analysis of the problem is first presented in a one-dimensional (1-D) setting. Markov random field prior (MRF) models appear in the corresponding 2-D Bayesian estimation problems. Unfortunately, algorithms for solving estimation problems involving 2-D MRF priors require a large number of computations to generate exact estimates and a prohibitively large computational load to calculate error statistics. Our objective here is to develop an algorithm with constant per-pixel complexity that also produces useful error statistics.

There are two possible approaches to achieving such an objective, namely *approximating the solution* (i.e., replacing the solution to the estimation problem with one which is easier to compute) or *approximating the problem* (i.e., replacing the estimation problem with one which has similar characteristics but which can be solved exactly using an efficient algorithm). An approach of the former type is described in [11]. In this paper, an approach of the latter type is developed by altering the prior model appearing in the problem formulation. In particular, this paper examines the usefulness of multiscale prior models for image segmentation. Multiscale models, which were introduced and studied in [9], admit algorithms with constant per-pixel complexity for the calculation of both estimates and error variances. They have also been shown to be useful in defining alternative approaches to problems in computer vision which are often posed in a variational context [9], [10], [12], [13]. These previous investigations, however, dealt with problems that resulted in linear estimation algorithms. In contrast, image segmentation is fundamentally a nonlinear problem, and thus, this paper represents the first work on using multiscale stochastic models to solve a nonlinear problem in image processing and computer vision. The algorithm that results not only has a modest computational load but also yields good performance.

The work presented in this paper builds primarily upon the two areas of variational methods for segmentation and multiscale methods for image processing. Section II summarizes the relevant material on variational methods for segmentation. Section III provides an overview of the multiscale modeling framework. The overview is not meant to be comprehensive, however, and many details are not discussed but may be found in the literature [9], [10], [12]–[15]. The subsequent Sections IV and V discuss the specifics of the 1-D scenario, and Section VI is devoted to the 2-D scenario.

II. VARIATIONAL METHODS IN IMAGE SEGMENTATION

A family of functionals proposed by Ambrosio and Tortorelli for image segmentation and denoising [5], [6] lies at the core of a segmentation and denoising algorithm developed by Shah [1] and extended by Pien and Gauch [7] among others. A member of this family of functionals, parameterized by ρ , is of the form

$$E(f, s) = \int \int_{\Omega} \left\{ r^{-1}(g - f)^2 + \lambda |\nabla f|^2 (1 - s)^2 + \frac{\beta}{2} \left(\rho |\nabla s|^2 + \frac{s^2}{\rho} \right) \right\} dx dy \quad (1)$$

where $\Omega \subset \mathbf{R}^2$ is the image domain, $g: \Omega \rightarrow \mathbf{R}$ is the image data, $f: \Omega \rightarrow \mathbf{R}$ is a piecewise smooth approximation to g , and $s: \Omega \rightarrow [0, 1]$ is an estimate of the edge-strength function, indicating high probability for the presence of an edge where it takes values close to one. The first and second terms constrain the approximating surface f to match the data as best as possible and also to be smooth in those places where s is close to zero. The third term ensures that s remain reasonably smooth and does not tend to 1 everywhere. As shown in [6], the minima of (1) converge to a minimum of the Mumford and Shah functional as $\rho \rightarrow 0$. The general approach Shah and Pien use to minimize (1) is coordinate descent: one alternates between fixing s and minimizing

$$\int \int_{\Omega} (r^{-1}(g - f)^2 + \lambda |\nabla f|^2 (1 - s)^2) dx dy \quad (2)$$

over possible f , and fixing f and minimizing

$$\int \int_{\Omega} \left(\lambda |\nabla f|^2 (1 - s)^2 + \frac{\beta}{2} \left(\rho |\nabla s|^2 + \frac{s^2}{\rho} \right) \right) dx dy \quad (3)$$

over possible s . Based on empirical evidence, Shah [1] and Pien and Gauch [7] have noted that this coordinate descent scheme converges to a reasonable solution and that the results are not significantly affected by the initial condition or whether one starts by estimating f or s .

III. MULTISCALE MODELS

Consider a prototypical quadratic minimization problem: minimize

$$E(f) = r^{-1} \|g - f\|^2 + \lambda \|Lf\|^2 \quad (4)$$

where f and g are vectors consisting of a lexicographic ordering of pixels in an image and L is a matrix chosen to ensure that the minimizer of (4) is smooth (e.g., L could take first differences of nearest neighbors as an approximation of a derivative). Minimization of (4) is equivalent to maximization of

$$p(f, g) = e^{-(r^{-1} \|g - f\|^2 + \lambda \|Lf\|^2)}. \quad (5)$$

One can now view $p(f, g)$ as a joint Gaussian probability distribution. In fact $p(f, g)$ is the joint distribution for f and g given by the measurement and model equations

$$g = f + \sqrt{r}v \quad \sqrt{\lambda}Lf = w \quad (6)$$

where v and w are independent zero-mean Gaussian random vectors with identity covariance. For a given value of g , the maximum of $p(f, g)$ over f occurs at the conditional mean of the Gaussian, $E[f|g]$. What's more, $E[f|g]$ is the Bayes least-squares estimate of f , according to the distribution induced by the modeling equations (6). Thus, one can view the problem of minimizing (4) from the perspective of optimization or of statistical estimation [9], [10], [16].

The main advantage of the Bayesian interpretation is that it casts the problem into a probabilistic framework in which it is natural to examine the accuracy of the resulting estimates. This is especially relevant in scientific applications such as remote

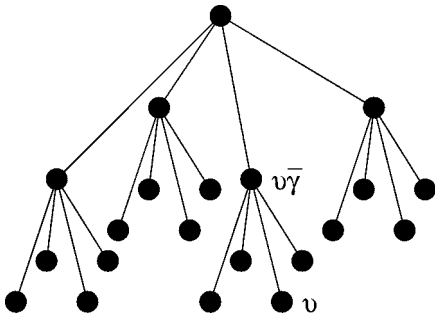


Fig. 1. Nodes of a tree, such as the quad-tree pictured here (typically one uses a binary tree for 1-D applications and a quad-tree for 2-D applications), are the index sets of the multiscale processes discussed in this paper. The operation of $\bar{\gamma}$ on an index is described as follows: if ν is the index of some node, then $\nu\bar{\gamma}$ denotes the parent of that node.

sensing, in which one may be, for example, assessing if features in a reconstruction are meaningful or statistically insignificant artifacts. In addition, this statistical formulation brings into focus the role played by the regularization term as a prior model, opening up the possibility of using alternate models that offer certain advantages.

Specifically, one can consider modeling a 1-D or 2-D phenomenon as the finest scale of a stochastic process on a tree. Doing so provides important computational advantages when performing estimation. The tree modeling framework used in this paper was introduced in [14] and further developed in [9], [10], [12], [13], [15]. In these references, one can find more detailed discussions of the following concepts.

The multiscale processes that are of interest in this paper are specified in terms of an autoregression on a tree (see Fig. 1). The root-to-leaf recursion takes the form

$$x(\nu) = A(\nu)x(\nu\bar{\gamma}) + B(\nu)w(\nu) \quad (7)$$

where the notation $\nu\bar{\gamma}$ refers to the parent of node ν . This recursion is a generalization of the standard state-space recursion for modeling a phenomenon evolving in time. In (7), the $w(\nu)$ and the state $x(\text{root})$ at the root node are independent zero-mean Gaussian random vectors, the w 's with identity covariance and $x(\text{root})$ with prior covariance P_{root} . The A and B matrices are deterministic quantities which define the statistics of the process on the tree. Observations $y(\nu)$ of the state variables have the form

$$y(\nu) = C(\nu)x(\nu) + v(\nu) \quad (8)$$

where the $v(\nu)$ are independent Gaussian random vectors, and the matrices $C(\nu)$ are deterministic. The least-squares estimates of process values at all nodes on the tree given all observations and the associated error variances can be calculated with an $O(N)$ recursive algorithm [14], [17], where N is the number of finest scale nodes on the tree. The algorithm consists of a fine-to-coarse recursion in which data in successively larger subtrees are fused up to the root node of the tree, and a subsequent coarse-to-fine recursion which produces both the optimal estimates and their error covariances. This algorithm generalizes the standard Kalman filter and Rauch–Tung–Striebel smoother.

The class of processes representable as the finest scale of a tree process includes some very important processes. In particular, Luetgten [18] has shown that any 1-D Gauss–Markov process can be represented on a tree using a model with a three-dimensional (3-D) state. The 1-D results in this paper make use of this model. The 2-D results make use of and extend the $1/f$ -like model used previously for the computation of optical flow and ocean surface reconstruction [10], [13]. This $1/f$ -like model and its extension are presented in Section VI.

IV. STATISTICAL INTERPRETATION OF SEGMENTATION IN ONE DIMENSION

The first step in applying the multiscale modeling framework to segmentation in 1-D is to develop a statistical interpretation of the discretized version of (1). One possibility for discretizing this functional is to replace the functions $f(x)$, $g(x)$, and $s(x)$ with regularly spaced collections of samples $f(i)$, $g(i)$, and $s(i)$; the integrals with sums over i ; and the derivatives with first differences. The result in 1-D is the discrete functional

$$\begin{aligned} E(f, s) = & r^{-1} \sum_{i=1}^n (f(i) - g(i))^2 \\ & + \lambda \sum_{i=1}^{n-1} (1 - s(i))^2 (f(i+1) - f(i))^2 \\ & + \frac{\beta}{2} \left(\rho \sum_{i=1}^{n-2} (s(i+1) - s(i))^2 + \frac{1}{\rho} \sum_{i=1}^{n-1} s(i)^2 \right) \end{aligned} \quad (9)$$

where n denotes the number of data points $g(i)$ ¹. As was done in [1], [7], one can use coordinate descent to minimize (9), thereby decomposing this complex problem into two simpler ones.

The problem of fixing s and finding the f that minimizes (9) is equivalent to finding the f that minimizes the discrete functional

$$\begin{aligned} E_s(f) = & r^{-1} \sum_{i=1}^n (f(i) - g(i))^2 \\ & + \sum_{i=1}^{n-1} \lambda (1 - s(i))^2 (f(i+1) - f(i))^2. \end{aligned} \quad (10)$$

A slightly more compact form can be written by collecting the samples $f(i)$, $g(i)$, and $s(i)$ into vectors f and $g \in \mathbf{R}^n$ and $s \in \mathbf{R}^{n-1}$. Specifically, let L_n be the $(n-1) \times n$ matrix that takes first differences of n samples, and let $S = \text{diag}(1 - s_1, \dots, 1 - s_{n-1})$. Then, (10) simplifies to $E_s(f) = \|f - g\|_{r^{-1}I}^2 + \lambda \|L_n f\|_{S^T S}^2$, where $\|x\|_W^2 = x^T W x$. Finding the minimum of E_s for fixed invertible S is then equivalent to finding the least-squares estimate of f assuming the following measurement and prior model

$$g = f + \sqrt{r}v_f \quad L_n f = \frac{1}{\sqrt{\lambda}} S^{-1} w_f \quad (11)$$

¹In this discretization, s has a length which is one sample less than that of f . This is because the samples of the discretized s lie between the samples of the discretized f . More details are provided in [19].

where v_f and w_f are independent zero-mean Gaussian random vectors with identity covariance. Notice that for i such that $s(i) \approx 1$, the multiplier of $w_f(i)$, $1/(1 - s(i))$, is very large. Thus, at these locations, the variance of $f(i+1) - f(i)$ in the prior model is high, and a least-squares estimator will allow big jumps to occur in the estimate of f . This is exactly what one wants the estimator to do at edge locations.

The problem of fixing f and finding s that minimizes (9) is equivalent to minimizing

$$\lambda \sum_{i=1}^{n-1} (f(i+1) - f(i))^2 (1 - s(i))^2 + \frac{\beta\rho}{2} \sum_{i=1}^{n-2} (s(i+1) - s(i))^2 + \frac{\beta}{2\rho} \sum_{i=1}^{n-1} s(i)^2. \quad (12)$$

Defining $a(i) = \lambda(f(i+1) - f(i))^2$, $b = \beta/2\rho$, $c = \beta\rho/2$, and $\gamma(i) = a(i)/(a(i) + b)$, one finds that, after completing the square, (12) can be rewritten

$$\sum_{i=1}^{n-1} ((a(i) + b)(\gamma(i) - s(i))^2 + a(i)(1 - \gamma(i))^2 - b\gamma(i)^2) + c \sum_{i=1}^{n-2} (s(i+1) - s(i))^2. \quad (13)$$

Ignoring terms which do not depend on s , one observes that minimizing (12) over s is equivalent to minimizing

$$E_f(s) = \sum_{i=1}^{n-1} (a(i) + b)(\gamma(i) - s(i))^2 + c \sum_{i=1}^{n-2} (s(i+1) - s(i))^2. \quad (14)$$

By defining the diagonal matrix

$$A = \text{diag}(\sqrt{\lambda(L_n f)_1 + b}, \dots, \sqrt{\lambda(L_n f)_{(n-1)} + b})$$

and the vector

$$\gamma = \left(\frac{\lambda(L_n f)_1^2}{\lambda(L_n f)_1 + b} \cdots \frac{\lambda(L_n f)_{(n-1)}^2}{\lambda(L_n f)_{(n-1)} + b} \right)^T$$

where $(L_n f)_k$ corresponds to the k th row of $L_n f$, one can rewrite (14) as

$$E_f(s) = \|\gamma - s\|_{A^T A}^2 + c \|L_{n-1} s\|^2. \quad (15)$$

In the original functional (1), s is constrained to lie within $[0, 1]$. If one removes this constraint, the problem of finding the s that minimizes (15) is equivalent to the problem of estimating s given the following measurement and prior model:

$$\gamma = s + A^{-1} v_s \quad L_{n-1} s = \frac{1}{\sqrt{c}} w_s \quad (16)$$

where v_s and w_s are independent zero-mean Gaussian random vectors with identity covariance. Notice that γ plays the role of an observation of s and that its components take on values near one where the difference between consecutive samples of f is large and near zero where the difference is small. Observe also that $\gamma(i)$ lies within $[0, 1]$; thus, the first term in (15) provides an increased penalty for s that does not stay within $[0, 1]$. This

is desirable because a solution to the unconstrained minimization of (15) that lies within $[0, 1]$ is an optimal solution of the constrained problem. As it turns out, this is often the case, as discussed in Section V.

As an aside, we note that one of the benefits of formulating a minimization problem in terms of statistics is that it yields natural interpretations of the parameters. These interpretations, in turn, can be used to form a loose set of guidelines for picking parameter values suitable for a particular segmentation application. Specifically, simple calculations lead to the following observations which can be used to pick the signal-dependent parameters λ , b , c , and r .

- λ is inversely proportional to the variability in the reconstructed signal f at locations where no edges are present. In particular, $E[(f(i+1) - f(i))^2 | s(i) = 0] = 1/\lambda$.
- r is the measurement noise variance term, determined in many applications by sensor specifications.
- c is proportional to the width of the edges. Specifically, the model for the edge-strength function s implies $E(s(i+1) - s(i))^2 = 1/c$.
- b controls the degree of edginess. In particular, choosing a value for it is related to the issue of defining what level of variability in f we wish to call an edge. Since the observations appearing in the model for the edge-strength function s have the form

$$\gamma(i) = \frac{\lambda(f(i+1) - f(i))^2}{\lambda(f(i+1) - f(i))^2 + b}$$

one desires that for i at edge locations

$$b \gg \lambda E[(f(i+1) - f(i))^2 | s(i) = 0] = 1, \\ b \ll \lambda(f(i+1) - f(i))^2.$$

These rules guided the choice of parameter values used for the numerical results presented in the subsequent section. More discussion concerning the parameters, including Monte Carlo results, can be found there and in [19].

V. NUMERICAL RESULTS

Based on the Gauss-Markov estimation problem formulations specified by (11) and (16), one can compute estimates \hat{f} and \hat{s} using any one of a variety of efficient methods. These include direct methods for solving the associated normal equations and Kalman filter smoothing. For the simulation results that follow, estimates \hat{f} and \hat{s} as well as error variances P_f and P_s were computed by a multiscale recursive estimation algorithm [13], [17], [18] (see also Section III). One detail concerning the implementation of this and other algorithms is that they require the specification of prior variances P_0 on $f(0)$ and $s(0)$, the first samples of the probabilistic models for f and s . However, the precise interpretation of the variational formulation as an estimation problem corresponds to viewing the initial value as unknown, which is equivalent to an infinitely large prior variance. While it is possible to accommodate this in the estimation formulation with no effect on algorithmic complexity, it is common to use an alternate approach in which one closely approximates the solution to the original problem by setting the prior covariance P_0 to a relatively large number. For the purposes of this

TABLE I
DESCRIPTION OF PARAMETERS IN 1-D SEGMENTATION ALGORITHM

Parameter	Description	Value for the Results in Figs.	
		2 and 4	3
λ	λ adjusts the smoothness in f away from edges. See (16).	1	25000
b	b affects the edginess of the estimate of the edge-strength function. See (16).	10	25
c	c sets the allowed variability in s . See (16).	100	1
r	r is the assumed noise variance in the data	1	$(0.2)^2$
P_0	P_0 is the prior covariance for initial process values.	100	
ϵ	Estimates of s are clipped to lie within $[0, 1 - \epsilon]$.	1.0×10^{-4}	
Δ	The algorithm stops after the percent change of the functional (9) falls below Δ .	0.01%	

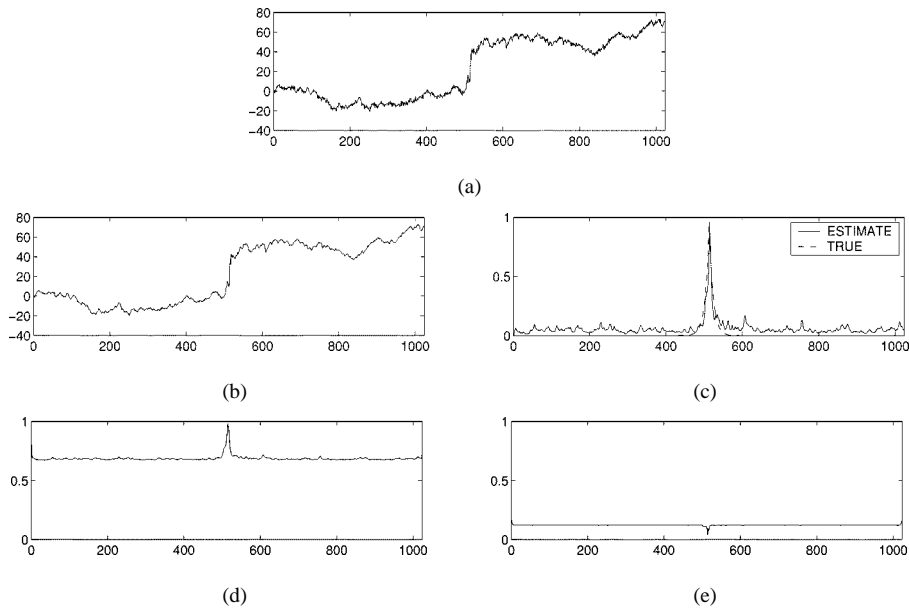


Fig. 2. The results for segmenting the data pictured in (a), a noisy observation of a process whose statistics are dictated by (11) for the true edge-strength function given by (17). The measurement noise is white and Gaussian with unit variance. All parameters are set as in Table I. (a) Data g , (b) reconstruction \hat{f} , (c) estimate of the edge-strength function \hat{s} ; (d) reconstruction error standard deviations $\sqrt{P_f}$; and (e) estimate of edge-strength function error standard deviations $\sqrt{P_s}$.

paper, P_0 is set large relative to $1/\lambda$ and $1/c$. This segmentation algorithm requires the specification of two parameters in addition to that of P_0 :

- ϵ : Since estimating f requires that $1/(1-s)^2$ be well-behaved, we must also enforce a constraint on the range of s . A simple solution is to clip each estimate of the edge-strength function so that for some small ϵ , $s \in [0, 1 - \epsilon]$. This solution proves adequate.
- Δ : As the segmentation algorithm is iterative in nature, one must specify a stopping criterion. For all of the results in this section, the algorithm stops when the percentage change in the functional (9) falls below Δ .

Our experience has been that a single pair of values of ϵ and Δ can be used to produce good segmentations of a variety of signal types. The parameter Δ specifies when to stop, but one also needs an initial guess with which to start the iterative algorithm. For the results in this section, the algorithm starts by estimating

f using an initial estimate of the edge-strength function $s^0 = 0$. A list of all of the algorithm's parameters are listed in Table I.

To illustrate the operation of the algorithm, some examples follow. These, in turn, are followed by some Monte Carlo experiments designed to assess quantitatively the performance of the algorithm.

A. Examples

Fig. 2 illustrates a segmentation of a synthetic signal. The data g in Fig. 2(a) consists of a signal f to which unit intensity white Gaussian measurement noise has been added. The signal f is a realization of a Gaussian process described by (11). The process is started with initial condition $f(0) = 0$ and generated using for s the double-sided exponential function

$$0.97 \times \exp(-|i - 512|/10). \quad (17)$$

The values of the parameters used are listed in Table I. Now, recall that where the edge-strength function is approximately one,

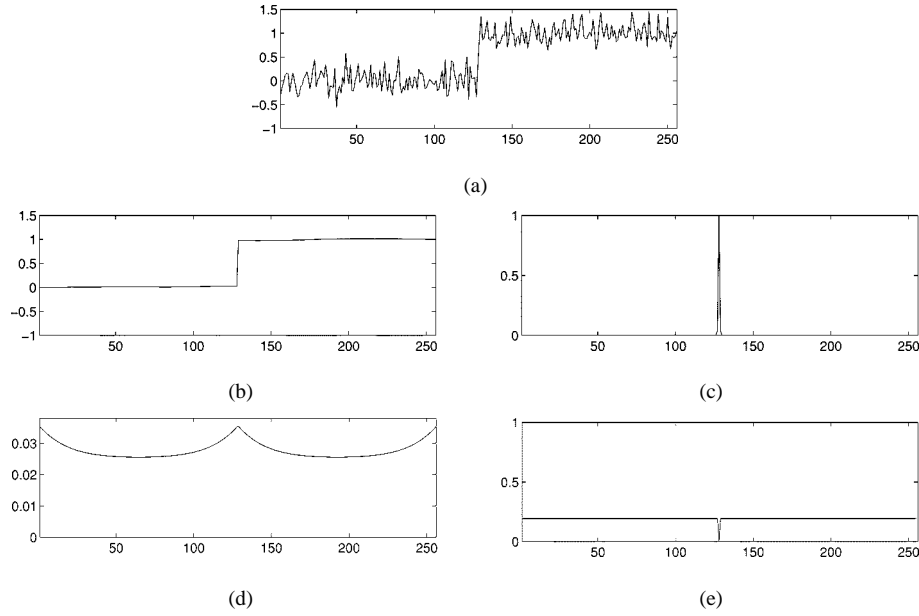


Fig. 3. Segmentation of a noisy unit step as plotted in part (a). The measurement noise is white and Gaussian with standard deviation 0.2. All parameters are given in Table I. (a) Data g , (b) reconstruction \hat{f} , (c) estimate of the edge-strength function \hat{s} , (d) reconstruction error standard deviations $\sqrt{\hat{P}_f}$, and (e) estimate of edge-strength function error standard deviations $\sqrt{\hat{P}_s}$

the variance of the increment in the model of f increases. This is clearly evident in Fig. 2(a) in which the particular realization of f displays a clear jump in its value in the vicinity of the edge-strength function's peak.

The data in Fig. 2(a) are then fed into the iterative segmentation and denoising algorithm. The results displayed in the remaining parts of Fig. 2 are after five iterations of the algorithm, at which point, the value of the functional (9) was changing by less than $\Delta = 0.01\%$. No clipping was necessary during the course of the run, and thus, the results are true to the discrete form of the variational formulation (9). The final reconstruction is a smoother version of the data, but the edge has not been smoothed away. The final estimate of the edge-strength function has a strong peak at the location of the edge. These are the desired results of the segmentation algorithm. In addition, the estimation error variance for f in Fig. 2(d) displays the characteristic one would expect: away from the expected edge, considerable lowpass filtering is effected, reducing the noise variance. However, in the vicinity of the edge, one expects greater variability and, in essence, the estimator performs less noise filtering, resulting in a larger error variance. Note also that the variance in the estimate of the edge process is almost constant, with a slight drop in the vicinity of the edge, i.e., where f changes abruptly, reflecting greater confidence that an edge is present in this vicinity. The results for the preceding example are good, but not completely convincing by themselves since f is matched to the algorithm by its construction. Consider a prototypical signal not matched to the model, namely a step edge. Fig. 3 displays results for data consisting of a unit step embedded in white Gaussian noise. The estimates are shown after 12 iterations. Once again, no clipping was necessary in the iterative process. The results demonstrate that the algorithm works as desired. It removes almost all of the noise away from the edge, while accurately preserving the discontinuity. Note that P_f increases near the outer ends because there are fewer measure-

ments in the vicinity and λ is large. As in the case of the first example, however, the error statistics reflect the fact that, near the edge, one expects less noise reduction in estimating f and has higher confidence in the estimate of the edge process because of the abrupt change in value of f .

B. Monte Carlo Experiments

Two different sets of Monte Carlo experiments are presented here. The first set provides some information for interpreting the error statistics. The second set shows that the algorithm is robust to parameter settings.

1) *Error Statistics:* In this section, a more careful look is taken at the error statistics provided by the segmentation algorithm in order to assess their accuracy and utility. Since the full iterative algorithm is nonlinear, the exact error variances in estimating f and s are not easily computed. The statistics calculated by our algorithm represent approximations that result from the linear estimation problems for each of the two separate coordinate descent steps for f and s . Fig. 4 presents Monte Carlo results comparing the error statistics computed by the segmentation algorithm with the actual error variances. Each experiment in this simulation corresponds to (a) generating a realization f of the process described by (11) for the fixed edge-strength function s given by (17) and with the initial value $f(0)$ set to 0; (b) adding white Gaussian measurement noise with unit intensity; and (c) applying the segmentation algorithm using the parameters in Table I to obtain the estimates \hat{f} of the realization f and \hat{s} of the edge-strength function s as well as P_f and P_s , the error variances for these estimates that the algorithm generates.

The quantities of interest for each run are $e_f = (\hat{f} - f)$, $e_s = (\hat{s} - s)$, and the error statistics P_f and P_s computed by the algorithm. From 100 independent runs, the following quantities are estimated: $E e_f^2$, $\text{Var}(e_f)$, $E P_f$, $E e_s^2$, $\text{Var}(e_s)$, and $E P_s$. These are plotted in Fig. 4 along with Monte Carlo error bars

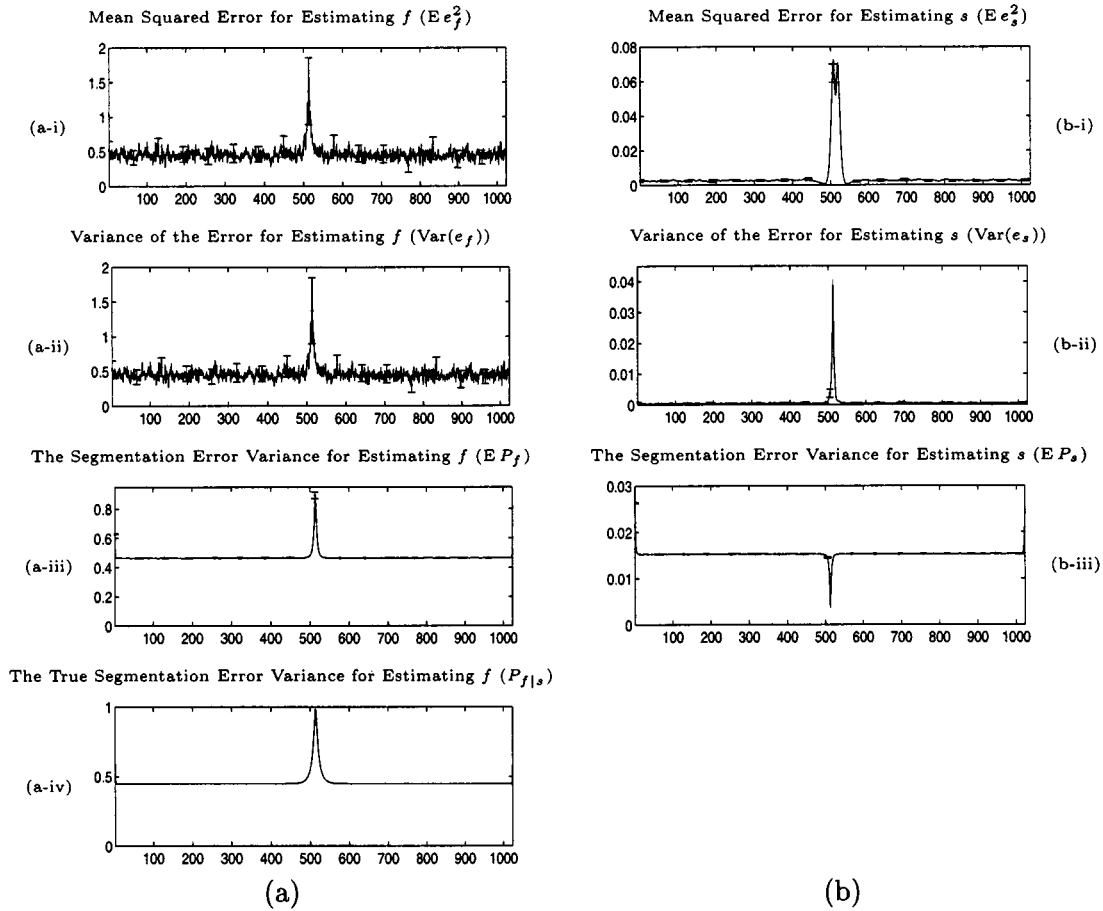


Fig. 4. Comparison of various error statistics compiled using Monte Carlo techniques for segmenting synthetic data. The data are realizations of a process whose statistics are given by (11) for the exponential edge-strength function given by (17). Part (a) of this figure displays statistics concerning the reconstruction errors, $e_f = (\hat{f} - f)$, and the reconstruction error standard deviations generated by the algorithm, P_f . Part (a-iv) displays the optimal error standard deviations for estimating f given that the true edge-strength function given by (17) is known. Part (b) of this figure displays the statistics concerning the errors in estimating the edge-strength function, $e_s = (\hat{s} - s)$. (a-i) Mean squared error for estimating f ($E e_f^2$), (a-ii) variance of the error for estimating f ($\text{Var}(e_f)$), (a-iii) segmentation error variance for estimating f ($E P_f$), (a-iv) true segmentation error variance for estimating f ($P_{f|s}$), (b-i) mean squared error for estimating s ($E e_s^2$), (b-ii) variance of the error for estimating s ($\text{Var}(e_s)$), and (b-iii) segmentation error variance for estimating s ($E P_s$).

set at 2 standard deviations. Comparing Fig. 4(a)-i for $E e_f^2$ and Fig. 4(a)-ii for $\text{Var}(e_f)$, one sees that these are quite close in value, indicating that the estimate produced by our algorithm is essentially unbiased. Comparing these two figures with the plot of $E P_f$, one observes that the error variance computed by our algorithm has essentially the same shape, reflecting the fact that it accurately captures the nature of the errors in estimating f . Fig. 4(a)-iv shows a plot of the error variance for an estimator that is given perfect knowledge of the edge process. Comparing this to Fig. 4(a)-iii, one notices that the segmentation algorithm performs nearly as well as if s were known perfectly and did not have to be estimated.

The error statistics for the edge-strength function are depicted in Fig. 4(b). $E e_s^2$ and $\text{Var}(e_s)$ being small relative to one indicate that the estimate of the edge-strength function is quite accurate and that the error does not vary considerably from sample path to sample path. In addition, the shapes of these plots have several interesting features related to the behavior of the estimator in the vicinity of the edge. Note first that, as can be seen in Fig. 2, the algorithm tends to estimate edge-strength functions that are slightly narrower than the actual edge-strength function. This is

actually preferable for segmentation, for which the peak locations in the estimates of the edge strength-functions are more important than the estimates' shapes. Because of this bias toward tighter edge localization, \hat{s} is a slightly biased estimate of s given by (17), as evidenced by the broader peak of $E e_s^2$ as compared to $\text{Var}(e_s)$.

A second interesting point is that $E e_s^2$ increases slightly in the vicinity of the edge, while the variance computed by the estimation algorithm, $E P_s$, decreases. The reason for this can be explained as follows. Specifically, the estimator believes that it has more information about s when the gradient of f is large. Thus, in the vicinity of an edge, the estimator indicates a reduction in error variance for estimating s . However, if the estimate of the *location* of the edge is in error, then the difference $e_s = (\hat{s} - s)$ will exhibit large, very localized errors, both positive and negative (just as one would see in the difference of two discrete-time impulses whose locations are slightly different). Thus, rather than providing an accurate estimate of the size of the estimation error variance in this vicinity, this dip in the error variance should be viewed as a measure of confidence in the presence of an edge in the vicinity.

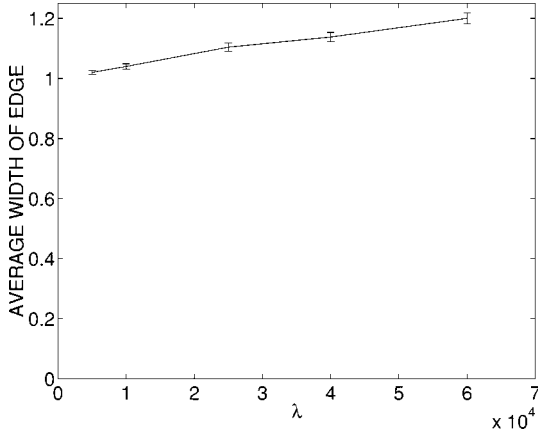


Fig. 5. Average value of W for the step edge example of Fig. 3 but for different values of λ .

2) *Parameter Selection*: Fig. 5 presents results from a Monte Carlo simulation designed to characterize how well the algorithm can segment signals for different values of the parameter λ . In this experiment, the algorithm is segmenting a unit step edge. The quantity computed from each segmentation is W , the number of values of the estimated edge-strength function that lie above a given threshold set close to one. This corresponds roughly to the sum of the widths of the edges in the segmentation. In the case of the step edge, the desired value of W is one. Except for λ , the algorithm's parameters take on the values listed in Table I for Fig. 3. The threshold is set to 0.9. The results in Fig. 5 are for 500 runs, and the error bars are set at one standard deviation.

Recall that the amount of smoothness the algorithm expects in f where there is no edge is directly related to λ . For $\lambda = 2.5 \times 10^3$, the algorithm generates the very flat step estimate of Fig. 3. However, one can not set λ too high because, as Fig. 5 shows, the average value of W increases with λ . Remember that W is a measure of how many points are very likely to be edges. If λ is set too high, the algorithm will put edges in many places and set the estimate of the function f almost constant between edges. Although the results of the algorithm depend on λ , the slope of the curve in Fig. 5 is not very steep. This indicates that small perturbations in the value of λ will not severely diminish the performance of the algorithm.

Similar experiments have been performed in which other parameters were adjusted. These can be found in [19]. To summarize, the effect of these parameters on the results coincide with the simple guidelines presented in Section IV. Furthermore, these Monte Carlo experiments indicate that the results are robust to small changes in the parameter settings.

VI. A MULTISCALE METHOD FOR IMAGE SEGMENTATION

As discussed in [19], solving the exact 2-D counterparts to the estimation problems in the 1-D segmentation and denoising algorithm is not an easy task. In particular, the calculations in 2-D correspond to solving estimation problems with prior models that are 2-D Markov random fields (MRF) or, equivalently, to solving elliptic partial differential equations and computing the diagonal elements of the inverses of elliptic operators [9]. There

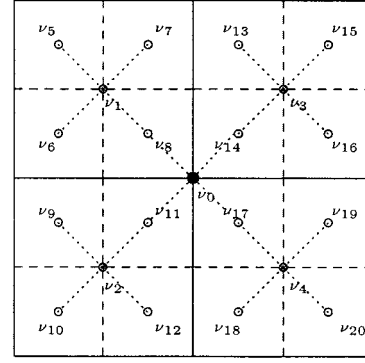


Fig. 6. How discontinuities are incorporated into the $1/f$ -like multiscale model.

exist no known algorithms which can compute the necessary quantities for this general problem with fewer than $O(n^3)$ operations, where n is the linear dimension of the image.

Since the objective is to generate estimates and error variances with constant computational complexity per pixel (i.e., with $O(n^2)$ operations), one is confronted with the need to develop approximations. The approach taken here is to alternately estimate a reconstruction and an edge-strength function assuming multiscale prior models. When the prior is a multiscale model, there is an efficient algorithm for producing the estimates, and the associated error variances of the reconstruction and edge-strength function [14], [17] (see also Section III). These two estimation steps replace the two coordinate descent step in Shah and Pien's variational approach to segmentation as discussed in Section II. The estimation is performed in an overlapped domain, as described in [20]. Thus, the multiscale models do not directly model f and s but the corresponding lifted versions \mathbf{f} and \mathbf{s} . The two models, one for each of the lifted fields \mathbf{f} and \mathbf{s} , are described in Section VI-A. Note that there are two principal distinctions between our multiscale modeling approach and some other quad-tree approaches to solving image processing problems. The first is that the variables on the tree are states that decorrelate regions of the image, and the second is that the associated estimation algorithm passes information everywhere along the tree so as to produce globally optimal estimates of the lifted fields.

A. Multiscale Models for Segmentation

Consider first, the model used for the estimation of \mathbf{s} . As discussed in [9], [10], [12], [13], [16], the smoothness penalty associated with the gradient, used for example in (2) and (3), corresponds to a fractal penalty in that it is roughly equivalent to a $1/f$ -like prior spectrum for the random field being modeled. This type of spectrum has a natural scaling law; namely, the variances of increments at finer and finer scales decrease geometrically. In [9], [10], [12], [13], it was demonstrated that a very simple multiscale model having this same scaling property leads to estimates that are very similar to those produced using the original smoothness penalty. A model of this type is used for the lifted version \mathbf{s} of the edge-strength function. Specifically

$$\mathbf{s}(\nu) = \mathbf{s}(\nu\bar{\gamma}) + d(\nu)B_s w_s(\nu) \quad (18)$$

TABLE II
DESCRIPTION OF PARAMETERS, IN THE MULTISCALE METHOD

Parameter	Description	Value for the Results in Figs.		
		7	8	9
b	b affects the edginess of the estimate of the edge-strength function.	30		10
λ	λ affects the determination of what is an edge and what isn't.	15		50
B_s	B_s adjusts the multiscale smoothness penalty placed on s . See (18).	1		0.87
B_f	B_f adjusts the multiscale smoothness penalty placed on f . See (19).	1/40		0.75
r	r is the assumed noise variance in image data.	1	6	1
P_{root}	P_{root} is the multiscale model prior covariance for the process value at the root node.	1×10^6		
\mathcal{O}	Each component of \mathcal{O} specifies the amount of overlap at a particular scale.	$\begin{pmatrix} 16 & 10 \\ 7 & 4 & 2 \\ 2 & 0 & 0 \end{pmatrix}$		$\begin{pmatrix} 50 & 31 \\ 18 & 11 & 7 \\ 4 & 2 & 1 \\ 0 & 0 & 0 \end{pmatrix}$
ϵ	Estimates of s are clipped to lie within $[0, 1 - \epsilon]$.	0.01		
I	I is the number of iterations of estimating f and s .	2		

where B_s is a constant, and the $w_s(\nu)$ are independent unit variance Gaussian random variables. As described in [12], [15], the $d(\nu)$ are constants that decrease from one scale to the next finer scale and depend on the amount of overlap used. The measurements and measurement error variances used in conjunction with the model for \mathbf{s} in (18) are exactly analogous to those specified by γ and A^{-2} in 1-D. The only difference in 2-D is that a sum of the squares of the first differences in each direction replaces the square of the first difference in one direction. Details are specified in [19].

The multiscale model for the lifted version of the reconstruction \mathbf{f} is, to some extent, similar to the one for the edge-strength function. However, significant modification to this model is needed in order to capture the presence of discontinuities, as indicated by the edge estimates. In particular, in the 1-D case, as captured in (11), the increments of f have a variance which is inversely proportional to the corresponding value of $(1 - s)^2$. Thus, the variance of the increment of f is large near an edge (i.e., where s is approximately one in value). In a similar manner, one needs to capture the idea that increments of \mathbf{f} , as one moves to finer scales, should have variances that reflect the presence of edges (i.e., that are again inversely proportional to $(1 - s)^2$). This is done as follows. Note that each node on the tree can be thought of as representing the center of a subregion of the image domain. A 2-D example is depicted in Fig. 6. The dots in this figure correspond to the center points of the regions associated with different nodes on the tree. The dots are shaded according to the scale of the corresponding node on the tree; the darker the dot, the coarser the scale. Thus, for example, the node ν_0 represents the entire large square region, while the node ν_3 at the next finest scale represents the upper-right quadrant of this large square. Now, if there is an edge located between ν_0 and ν_3 (i.e., if the values of s at image domain

pixels between these nodes indicate the presence of an edge), the variance of the scale-to-scale increment of \mathbf{f} between these two nodes should increase. More precisely, the model for \mathbf{f} is specified by the recursion

$$\mathbf{f}(\nu) = \mathbf{f}(\nu\overline{\gamma}) + \eta(\nu)d(\nu)B_f w_f(\nu) \quad (19)$$

where B_f is a constant, $w_f(\nu)$ are independent unit variance Gaussian random variables, and $\eta(\nu)$ is the sum of $1/(1 - \hat{s}(i, j))^2$ for estimates of the edge-strength function values which fall on the line connecting ν and $\nu\overline{\gamma}$. In this manner, additional uncertainty is put into the recursion for \mathbf{f} at the appropriate locations.

B. Numerical Results

This section presents numerical results on two test images, a synthetic image of a circle and an MRI brain scan. Table II lists the algorithm's parameters and values². The execution times for each example were on the order of minutes when run using MATLAB on a Sparc Ultra.

The segmentation of a synthetic 64×64 image of a circle is presented in Fig. 7. The circle image provides a simple example for which one can observe the desired outputs of the algorithm. The reconstruction contains very little of the noise present in the original data. Furthermore, the edges of the circle have not been smoothed over in the reconstruction. The estimate of the edge-strength function is close to zero everywhere except at the edge of the circle, where the estimate tends to unity. This indicates the algorithm has successfully identified edges and has ignored the spurious changes in intensity in the data due to noise.

²The parameters b , λ , B_s , B_f , and r have interpretations analogous to those presented in Section IV for the parameters b , λ , c , and r appearing in the 1-D problem. The interpretations were used to guide the choice of parameters used in the numerical examples of the multiscale method for segmentation.

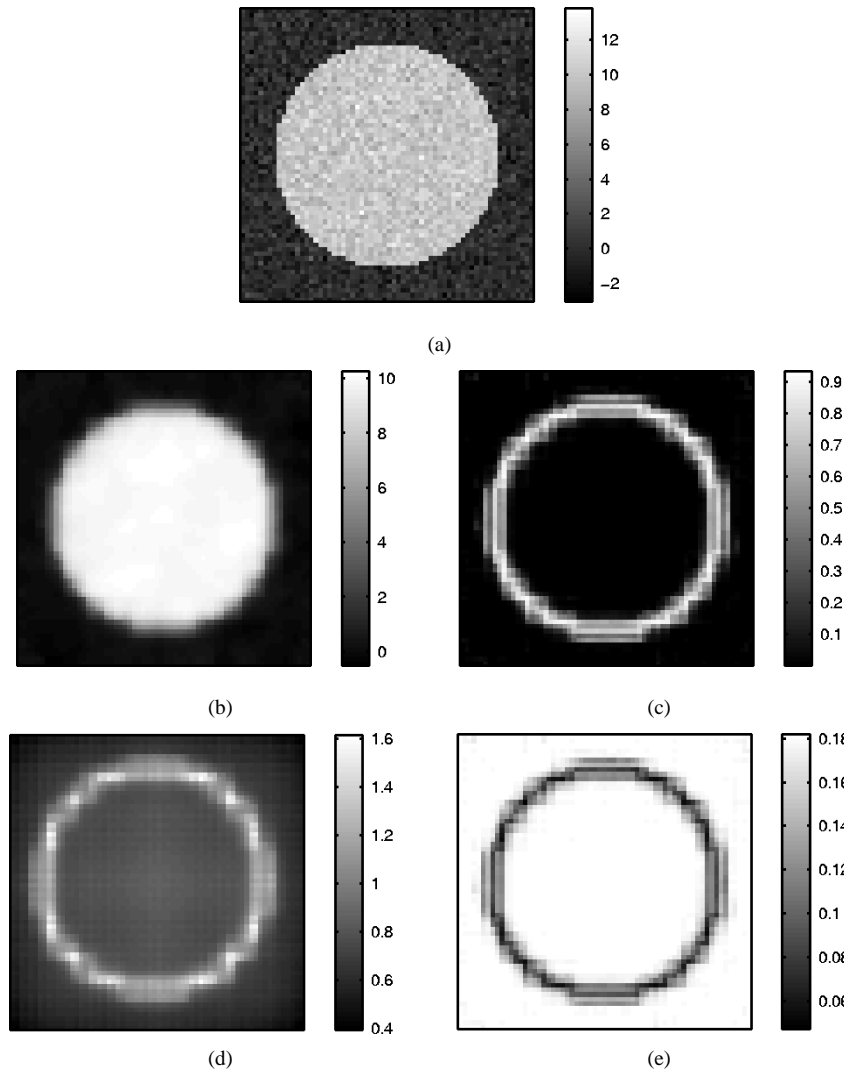


Fig. 7. Segmentation of a synthetic circle image computed using the multiscale method. (a) Data g , (b) reconstruction \hat{f} , (c) estimate of the edge-strength function \hat{s} ; (d) reconstruction error standard deviations $\sqrt{P_f}$; and (e) estimate of edge-strength function error standard deviations $\sqrt{P_s}$.

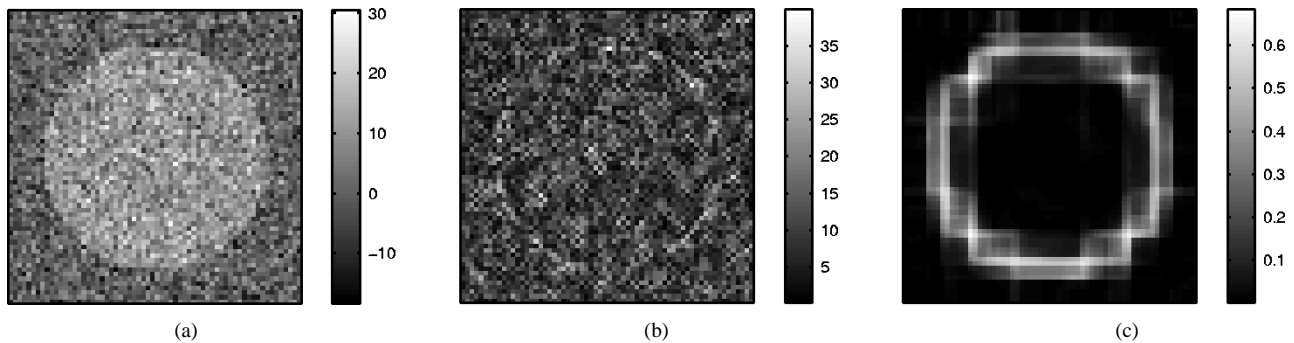


Fig. 8. This figure illustrates the difference between a gradient map and the estimate of the edge-strength function for a noisy circle image. (a) Data g , (b) magnitude of the gradient, and (c) estimate of the edge-strength function \hat{s} .

Fig. 8 illustrates the difference between a gradient map and the estimate of the edge-strength function. As for the example in Fig. 7, the unprocessed image is a circle embedded in white noise. The standard deviation of the noise is six. The magnitude of the gradient is very noisy, but the edge-strength function is only moderately noisy and clearly indicates the location of the underlying circle.

In addition to the reconstruction and estimate of the edge-strength function, the multiscale algorithm computes the standard deviations of the error in the reconstruction and estimate of the edge-strength function. Notice that the error standard deviations for the reconstruction increase near edges and, for the estimate of the edge-strength function, decrease near edges, as in the 1-D case. Thus, one expects that the error standard

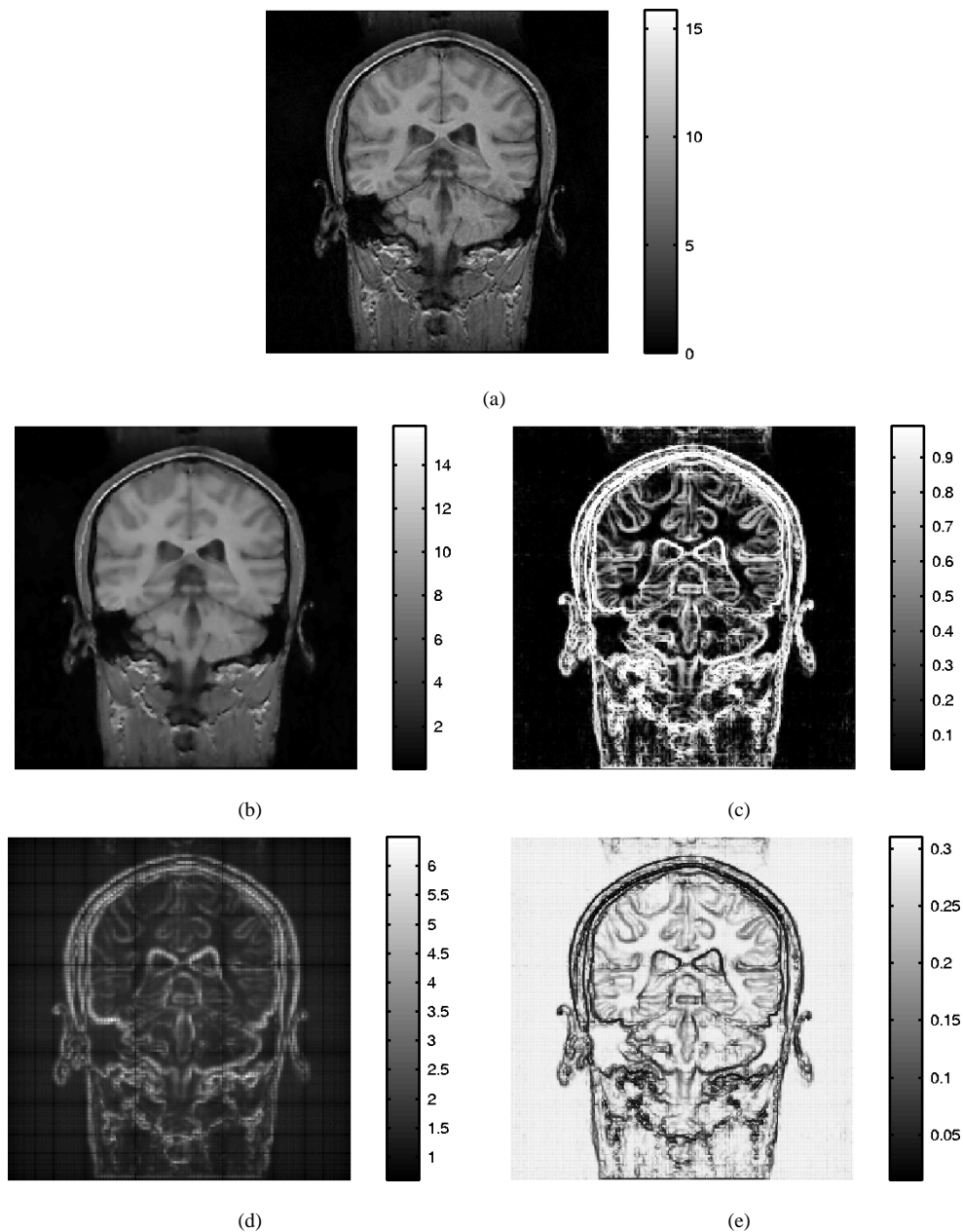


Fig. 9. MRI segmentation computed using the multiscale method. (a) Data g , (b) reconstruction \hat{f} , (c) estimate of the edge-strength function \hat{s} ; (d) reconstruction error standard deviations $\sqrt{P_f}$; and (e) estimate of edge-strength function error standard deviations $\sqrt{P_s}$.

deviations in Figs. 7 and 9 are of similar significance to those generated in 1-D. Another consequence of the above-mentioned properties of the error standard deviations is that they can be used not only to estimate one's confidence in segmenting the image but also to improve one's estimate of the boundary locations. This is a consequence of the error standard deviations marking the edges in the image as well or better than the estimate of the edge-strength function.

Fig. 9 displays a multiscale segmentation of a 256×256 MRI brain scan³. There are many different potential uses of MRI brain segmentation. Associated with each of these different uses are different goals of the segmentation. One of these goals is to

³An MRI image with pixel values ranging over more than 800 integers was shifted and scaled to produce the data for this figure. The dynamic range of the scaled MRI data is close to that of the circle images.

demarcate the boundaries of the ventricles, the two hollow regions in the middle of the brain. Another goal is to determine the boundary between gray and white matter in the brain. The estimate of the edge-strength function displayed in Fig. 9 does a good job at indicating likely boundaries both of the ventricles and of the gray and white matter.

VII. CONCLUSION

In this paper, we have described a new approach to the reconstruction and segmentation of noise corrupted signals and images. The points of departure for our work are the variational formulation of Mumford and Shah [2] and, more explicitly, the relaxed variational formulation of Ambrosio and Tortorelli [5].

The latter formulation leads to the simultaneous computation of both a reconstructed signal or image and an edge-strength function which provides a measure of the likely locations of edges or abrupt changes in the image or signal values.

The contributions of this paper are several. First, we provide precise statistical interpretations of the steps involved in estimating both the reconstructed signal or image and the edge-strength function. This interpretation not only establishes an intellectual bridge to statistical estimation, but it also leads directly to the extension of the algorithm to produce error variances for the reconstructed image and edge-strength function.

Moreover, this statistical interpretation leads us to another significant extension and new algorithm, in the case of 2-D image processing. In particular, replacing the random field image models implied by the variational formulations in [1], [7] with a multiscale prior model with similar characteristics yields an extremely efficient algorithm for the simultaneous estimation of the reconstructed image and edge-strength function. Furthermore, the new algorithm can efficiently compute the variances of the errors in these estimates. The algorithm for computing both estimates and error variances given a multiscale model has constant computational complexity per pixel. In contrast, the best known algorithm for computing estimates and error variance given a random field model has a complexity per pixel that increases as the square root of the total number of pixels.

In addition to these substantial computational advantages, we have also demonstrated the efficacy of these algorithms through several experimental studies. In 1-D, we demonstrated the accuracy of the error variance calculations that our algorithm automatically produces for the reconstructed image. In 2-D, we also demonstrated the ability of our algorithm to produce reliable estimates even in the presence of very high noise. In particular, we demonstrated that the estimates of edge-strength functions produced by our approach are much more robust to noise than image gradient maps. Thus, the edge-strength function estimates can be subsequently processed to produce robust and accurate edge estimates, e.g., by thresholding [1] or more sophisticated methods such as curve evolution [8].

ACKNOWLEDGMENT

The authors would like to thank several of the reviewers for their constructive criticisms and suggestions.

REFERENCES

- [1] J. Shah, "Segmentation by nonlinear diffusion, II," in *Proc. IEEE Computer Vision Pattern Recognition Conf.*, 1992, pp. 644–664.
- [2] D. Mumford and J. Shah, "Boundary detection by minimizing functionals, I," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 1985, pp. 22–26.
- [3] —, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, pp. 577–684, 1989.
- [4] J. Morel and S. Solimini, *Variational Methods in Image Segmentation*. Boston, MA: Birkhäuser, 1995.
- [5] L. Ambrosio and V. M. Tortorelli, "Approximation of functionals depending on jumps by elliptic functionals via Γ -convergence," *Commun. Pure Appl. Math.*

- [6] —, "On the approximation of free discontinuity problems," *Bollettino Della Unione Matematica Italiana*, vol. 6-B, pp. 105–123, 1992.
- [7] H. Pien and J. Gauch, "Variational segmentation of multi-channel MRI images," in *Proc. IEEE Int. Conf. Image Processing*, Nov. 1994, pp. 508–512.
- [8] H. H. Pien, M. Desai, and J. Shah, "Segmentation of MR images using curve evolution and prior information," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 116, no. 8, pp. 1233–1245, 1997.
- [9] M. Luetgten, W. C. Karl, and A. S. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow," *IEEE Trans. Image Processing*, vol. 3, pp. 41–64, Jan. 1994.
- [10] P. Fieguth, W. Karl, A. Willsky, and C. Wunsch, "Multiresolution optimal interpolation and statistical analysis of TOPEX/POSEIDON satellite altimetry," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 280–292, Feb. 1995.
- [11] J. Kaufhold, M. Schneider, A. W. Willsky, and W. C. Karl, "A statistical method for efficient segmentation of MR imagery," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 11, no. 8, pp. 1213–1231, 1997.
- [12] P. Fieguth, "Application of multiscale estimation to large scale multidimensional imaging and remote sensing problems," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, June 1995.
- [13] M. Luetgten, "Image processing with multiscale stochastic models," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, May 1993.
- [14] K. Chou and A. Beneveniste, "Multiscale recursive estimation, data fusion, and regularization," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 464–478, Mar. 1994.
- [15] P. W. Fieguth, A. S. Willsky, and W. C. Karl, "Efficient multiresolution counterparts to variational methods for surface reconstruction," *Comput. Vis. Image Understand.*, vol. 70, pp. 157–176, May 1998.
- [16] R. Szeliski, *Bayesian Modeling of Uncertainty in Low-Level Vision*. Norwell, MA: Kluwer, 1989.
- [17] K. Chou, "A stochastic modeling approach to multiscale signal processing," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, May 1991.
- [18] M. R. Luetgten, W. C. Karl, A. S. Willsky, and R. R. Tenney, "Multiscale representations of Markov random fields," *IEEE Trans. Signal Processing*, vol. 41, pp. 3377–3396, Dec 1993.
- [19] M. K. Schneider, "Multiscale methods for the segmentation of images," M.S. thesis, Mass. Inst. Technol., Cambridge, May 1996.
- [20] W. W. Irving, P. W. Fieguth, and A. S. Willsky, "An overlapping tree approach to multiscale stochastic modeling and estimation," *IEEE Trans. Image Processing*, vol. 6, pp. 1517–1529, Nov. 1997.



Michael K. Schneider (S'93) received the B.S.E. degree in electrical engineering from Princeton University, Princeton, NJ, in 1994, and the M.S. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1996. He is currently pursuing the Ph.D. degree at MIT.

His research interests include statistical signal and image processing and numerical linear algebra.



Paul W. Fieguth (S'87–M'96) received the B.A.Sc. degree from the University of Waterloo, Waterloo, Ont., Canada, in 1991, and the Ph.D. degree from the Massachusetts Institute of Technology (MIT), Cambridge, in 1995, both in electrical engineering.

He joined the University of Waterloo in 1996, where he is currently Assistant Professor in systems design engineering. Prior to this, he held postdoctoral positions with the Department of Computer Science, University of Toronto, Toronto, Ont., and in the Laboratory for Information and Decision Systems,

MIT. His research interests include statistical signal and image processing, hierarchical algorithms, and interdisciplinary applications of such methods, particularly to remote sensing.



William C. Karl (M'91) received the S.M., E.E., S.B., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge.

He was Staff Research Scientist with Brown-Harvard-MIT Center for Intelligent Control Systems and the MIT Laboratory for Information and Decision Systems, Cambridge, from 1992 to 1994. He joined the faculty of Boston University, Boston, MA, in 1995, where he is currently Assistant Professor of electrical, computer, and systems engineering. Since

January 1996, he has held a joint appointment in the Department of Biomedical Engineering. His research interests are in the areas of multidimensional and multiscale signal and image processing, geometric estimation, detection, and medical signal and image processing.

Dr. Karl is associate editor of the *TRANSACTIONS ON IMAGE PROCESSING*. He is a member of Sigma Xi.



Alan S. Willsky (S'70-M'73-SM'82-F'86) received the S.B. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1969 and 1973, respectively.

He joined the MIT faculty in 1973 and his present position is Professor of electrical engineering. From 1974 to 1981, he served as Assistant Director of the Laboratory for Information and Decision Systems, MIT. He is also a founder and member of the board of directors of Alphatech, Inc., and is currently a member of the U.S. Air Force Scientific Advisory

Board. He has held visiting positions at Imperial College, London, U.K., L'Université de Paris-Sud, and the Institut de Recherche en Informatique et Systèmes Aléatoires, Rennes, France. He is the author of the research monograph *Digital Signal Processing and Control and Estimation* and is the co-author of the undergraduate text *Signals and Systems*. His research interests are in development and application of advanced methods of estimation and statistical signal and image processing. Methods he has developed have been successfully applied in a wide variety of applications including failure detection in high-performance aircraft, advanced surveillance and tracking systems, electrocardiogram analysis, computerized tomography, and remote sensing.

Dr. Willsky received the Donald P. Eckman Award from the American Automatic Control Council in 1975. He was Program Chairman for the 17th IEEE Conference on Decision and Control, has been an associate editor of several journals and special guest editor for several special issues, and has served as a member of the Board of Governors and Vice President for Technical Affairs of the IEEE Control Systems Society. In 1988, he was made a Distinguished Member of the IEEE Control Systems Society. He has given plenary and keynote lectures at a number of major scientific meetings including the 20th IEEE Conference on Decision and Control, the 1991 IEEE International Conference on Systems Engineering, the 1991 SIAM Conference on Applied Linear Algebra, the 1992 Inaugural Workshop for the National Centre for Robust and Adaptive Systems, Canberra, Australia, the 1992 INRIA 25th Anniversary Symposium in Paris, the 1993 IEEE Symposium on Image and Multidimensional Signal Processing, Cannes, France, and the 1997 Wavelet Applications in Signal and Image Processing Conference. He was awarded the 1979 Alfred Noble Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE for a paper excerpted from his monograph.

A Statistical Method for Efficient Segmentation of MR Imagery *

J. Kaufhold[†], M. Schneider[‡], W. C. Karl[§], A. S. Willsky[‡]

Abstract

Magnetic resonance imaging (MRI) has become a widely used research and clinical tool in the study of the human brain. The ability to robustly, accurately, and repeatably quantify morphological measures from such data is aided by the ability to accurately segment the MRI data set into homogeneous regions such as gray matter, white matter, and cerebro spinal fluid. The large amount of data associated with typical MRI scans makes completely manual segmentation prohibitive on a large scale. In this paper an efficient approach to the segmentation of such MRI imagery is presented. The approach uses an estimation-theoretic interpretation of the segmentation problem to develop a computationally efficient, statistically-based recursive technique for its solution. Being statistically based, the method also provides associated measures of uncertainty of the resulting estimates, which are extremely important both for evaluation of the estimates as well as their combination with other sources of information.

1 Introduction

Magnetic resonance imaging (MRI) has become a widely used research and clinical tool in the study of the human brain. The size and shape (i.e. the morphology) of the various structures of the brain which can be obtained from such imagery correlate with various developmental differences, disease states, and injuries. In [1] independent measures of brain volumes, shapes, and positions are readily correlated with behavioral as well as physiological measures. For instance, [2] suggests that Rett syndrome, a progressive disorder associated with regression of psychomotor development and slow brain growth to age one, could be diagnosed by precise volumetric analysis of cortical and nuclear structures over time. Also, in [3], it was shown that the female brain is approximately 10% smaller than a male brain but the disproportionate variation in the volumes of male and female brain regions and substructures reported was not consistent with a simple allometric explanation for the sexual dimorphism. Shenton et. al [4] measured post mortem left temporal lobe volumes in schizophrenics. Certain anatomical differences from normals in these volumes correlated with schizophrenia. In [5], clinically relevant frontal brain lesion volumes were measured morphologically

*This work was supported by the National Institutes of Health under Grant NINDS 1 R01 NS34189, by the Air Force Office of Scientific Research under Grant F49620-96-1-0028, by ONR under Grant N00014-91-J-1004, by a National Science Foundation Graduate Research Fellowship, and a Whitaker Foundation Graduate Research Fellowship.

[†]Corresponding Author. Department of Biomedical Engineering, Boston University, 44 Cummington St., Boston, MA 02215, USA, johnk@bu.edu, (617) 353-9295, FAX:(617) 353-8437.

[‡]Stochastic Systems Group, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, USA

[§]Departments of Electrical and Computer Engineering and Biomedical Engineering, Boston University, 44 Cummington St., Boston, MA 02215, USA.

by MR imaging and validated by measuring CT blood flow in the same volume. These are only a few examples which strongly suggest the ability to robustly, accurately, and repeatably quantify such morphological measurements is extremely important.

Accurate and reproducible quantification depends on accurate segmentation and classification of the MRI brain data set into constituent functional parts. In particular, this overall goal is aided by the segmentation of the raw MRI data into homogeneous regions such as gray matter, white matter, and cerebro spinal fluid (CSF). These regions and their boundaries have typically been delineated through a labor intensive procedure by specially trained technicians, at great expense in time, money, and effort and with the greater variability that comes with human intervention [6]. The incorporation of automation in the segmentation process has been shown to reduce variability across technicians and across morphological structure measurements [7]. Such results have prompted greater interest in increasing the automation of this component of the analysis process, which is the long term focus of this work.

One promising approach that has been taken to automating such segmentation problems in the computer vision literature is the use of a variational or cost-functional framework [8–14]. In these approaches to segmentation an energy functional captures desired properties of the resulting segmentation, such as smoothness within homogeneous regions, preservation of boundaries between homogeneous regions, etc. The minimum energy the functional can attain given the observed image is chosen as the segmentation. The advantage of such methods is the ease with which various desirable effects can be directly incorporated into the energy functional and reflected in the resulting segmentation. Unfortunately, obtaining the minimum of these energy functions leads to large and computationally taxing optimization problems. In addition, while these deterministic methods can provide estimates of the underlying components and their boundaries, they do not provide information about the uncertainty in these estimates. Such measures of uncertainty are important both for direct evaluation of the estimates themselves, as well as providing the information necessary to fuse these estimates with other sources of knowledge, such as probabilistic atlases, images from other modalities, etc. Further, while the addition of many terms capturing various effects is conceptually simple, the specification of the corresponding weights or importance given to these terms can be challenging.

Alternative statistical formulations of such segmentation problems based on Markov random field (MRF) prior models [15–17], while also providing visually desirable solutions, also lead to computationally taxing optimization problems. Indeed, the resulting optimization problems are very similar to the variationally derived ones. Furthermore, while these Bayesian approaches conceptually provide a framework for the calculation of measures of uncertainty (i.e. estimation error variances), in practice, such measures are virtually never obtained due to the great computational cost involved. Thus current methods for the automation of MRI segmentation are computationally costly and, further, do not provide rational measures of uncertainty.

To overcome the above limitations of existing methods, in this paper we present a statistically-based recursive approach to the segmentation of MRI imagery which is computationally efficient and provides measures of uncertainty of the resulting estimates. The approach is based on an interpretation of the segmentation problem as an equivalent recursive estimation problem. This enables the use of efficient, near-optimal recursive filtering methods for its solution. As a result, rather than solving the large overall segmentation problem all at once, the problem is decomposed into a sequence of smaller estimation problems, yielding computational efficiency in addition to uncertainty measures.

The paper is organized as follows. In Section 2 we present a summary of the variationally-based image segmentation on which we build. We then present an interpretation of the segmentation

problem as an estimation problem in Section 3. In Section 4 we show how this problem may be solved recursively and thus efficiently. In Section 5 we show examples of brain segmentations using our approach. We then present our conclusions in Section 6.

2 A Variational Formulation of the Segmentation Problem

2.1 Problem Statement

We base our approach to the segmentation of MRI imagery on a variational formulation of the segmentation problem presented in [8,9,12]. This formulation simultaneously produces a segmented image estimate \hat{f} and edge map estimate \hat{s} as the minimizers of the following energy functional:

$$E(f, s) = \int_{\Omega} \underbrace{\mu(g - f)^2}_{\text{Data Fidelity}} + \underbrace{\lambda(1 - s)^2 |\nabla f|^2}_{\text{Smoothness}} + \underbrace{\nu \left(\rho |\nabla s|^2 + \frac{1}{\rho} s^2 \right)}_{\text{Edge Penalty}} dx dy \quad (1)$$

where g denotes the observed image data, f denotes the piecewise smooth approximating field, s denotes the corresponding continuous edge strength image, Ω is the image domain and μ , λ , ν , and ρ are scalar weights specifying the relative importance of the terms. The edge image s can be viewed as an indication of edge or boundary strength at each location in the observed image, and ranges in value between 0 and 1. There are three types of terms in (1) – a data fidelity term, a smoothness term, and edge penalty terms. These terms ensure that the pair \hat{f} , \hat{s} that minimizes (1) has the properties that \hat{f} is close to the data g , that \hat{f} is nearly flat except where s is close to 1 (indicating the presence of an edge), that the edge field \hat{s} itself is smooth, and that the number of edges (i.e. the size of \hat{s}) is not too great, preventing over-segmentation. Note that as $\rho \rightarrow 0$ it is shown in [12] that the solution of (1) approaches the binary edge formulation of [10].

In practice, (1) is solved by alternatively fixing s and minimizing with respect to f and then fixing f and minimizing with respect to s , as shown in Figure 1. This corresponds to alternatively estimating the values of homogeneous image regions assuming that the location of the boundaries between them are known and then estimating the location of the boundaries assuming the values within the smooth, homogeneous regions are known. In particular, with the edge field s considered fixed, the following energy functional is minimized for the segmented image values f :

$$E_s(f) = \int_{\Omega} \mu(g - f)^2 + \lambda(1 - s)^2 |\nabla f|^2 d\Omega \quad (2)$$

Note, in this step a penalty, $|\nabla f|^2$, enforces smoothness on the resulting estimated field f *except* where the edge field s is close to 1.

For the other step, the image f is considered fixed and the minimum of the functional with respect to s is sought. This step can be shown to be equivalent to minimizing the following energy functional for s [18]:

$$E_f(s) = \int_{\Omega} \left(\lambda |\nabla f|^2 + \frac{\nu}{\rho} \right) \left[\frac{\lambda |\nabla f|^2}{\lambda |\nabla f|^2 + \frac{\nu}{\rho}} - s \right]^2 + \nu \rho |\nabla s|^2 d\Omega \quad (3)$$

The second term is a penalty that will impose smoothness on the resulting edge field everywhere, while the first term contains an equivalent derived “observation” of the edge field based on a normalized gradient of the image f . In particular, note that this observation, $\lambda |\nabla f|^2 / (\lambda |\nabla f|^2 + \frac{\nu}{\rho})$, has values in the range $[0, 1]$ and will be close to 1 where the gradient of f is large.

2.2 Discretization

In practice, we work on a discrete grid of sample points and approximate the gradient operator with a finite difference scheme. In particular, we assume that both f and s are sampled on the same $n \times n$ row-column grid containing a total of n^2 points. It will prove convenient to define the n -vectors $\mathbf{f}(j)$ and $\mathbf{s}(j)$, obtained by stacking the sampled image values $(f)_{ij}$ and sampled edge field values $(s)_{ij}$ in column j , respectively, as illustrated for the f field in Figure 2. We also define \mathbf{f} and \mathbf{s} as the overall n^2 -vectors of these sampled points, obtained by stacking the $\mathbf{f}(j)$ and $\mathbf{s}(j)$, respectively. Similarly, we define $\mathbf{g}(j)$ to be the n -vector obtained by stacking the sampled observation values $(g)_{ij}$ in column j and define $\mathbf{h}(j)$ to be the n -vector obtained by stacking the sampled values of the s -field “observation” $\lambda|\nabla f|^2 / (\lambda|\nabla f|^2 + \nu/\rho)$ in column j . In particular, the latter samples are given by $\mathbf{h}(j)_i = ([f]_{ij} - [f]_{i,j-1}]^2 + [f]_{ij} - [f]_{i-1,j}]^2 / ([f]_{ij} - [f]_{i,j-1}]^2 + [f]_{ij} - [f]_{i-1,j}]^2 + \frac{\nu}{\lambda\rho})$ and serve as a normalized gradient observation, lying in the range $[0, 1]$.

With this notation, a discrete version of the field energy term $E_s(\mathbf{f})$ given in (2) may be obtained as¹:

$$E_s(\mathbf{f}) = \sum_{j=1}^n \|\mathbf{g}(j) - \mathbf{f}(j)\|_{\mu I}^2 + \|D_r \mathbf{f}(j)\|_{V(j)}^2 + \|\mathbf{f}(j) - \mathbf{f}(j-1)\|_{V(j)}^2 \quad (4)$$

where $\|\mathbf{x}\|_M^2 = \mathbf{x}^T M \mathbf{x}$ and D_r is a first-order difference operator along the rows of each column given by:

$$D_r = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}. \quad (5)$$

The weighting matrices $V(j)$ are diagonal and capture the spatially varying, edge-dependent weights on column j of the gradient term in (2). In particular, $V(j)_{ii} = \lambda[1 - (s)_{ij}]^2$.

Similar to (4), a discrete version of the edge energy term $E_f(\mathbf{s})$ given in (3) may be obtained as:

$$E_f(\mathbf{s}) = \sum_{j=1}^n \|\mathbf{h}(j) - \mathbf{s}(j)\|_{W(j)}^2 + \|D_r \mathbf{s}(j)\|_{\nu \rho I}^2 + \|\mathbf{s}(j) - \mathbf{s}(j-1)\|_{\nu \rho I}^2 \quad (6)$$

where D_r is the row-derivative operator defined in (5) and the weighting matrices $W(j)$ are diagonal and capture the spatially varying weights on column j of the data term in (3). In particular, $W(j)_{ii} = \lambda[(f)_{ij} - (f)_{i,j-1}]^2 + \lambda[(f)_{ij} - (f)_{i-1,j}]^2 + \frac{\nu}{\rho}$.

The solution to the overall segmentation problem is then obtained by alternately minimizing each of the two quadratic energies $E_s(\mathbf{f})$ in (4) and $E_f(\mathbf{s})$ in (6), as shown in Figure 1. Each of these minimizations is a large, computationally intensive, optimization problem, which must be performed repeatedly to obtain a segmentation. Direct solution of the system of equations defining the minimum of (4) or (6) at any stage of the algorithm would involve $O(n^6)$ calculations and, since n is typically in the range of 10^2 – 10^3 , would be prohibitively expensive. In practice, therefore, the solution is approximated to an arbitrary degree of accuracy via an iterative method, such as conjugate gradient [19] or multigrid [20]. Even with such methods, solution of the large set of equations is computationally challenging. Further, only the *estimate* of the image field $\hat{\mathbf{f}}$ or edge process $\hat{\mathbf{s}}$ is provided, with no measure of reliability.

To avoid these difficulties we use an interpretation of the problems (4) and (6) in an estimation theoretic context [18]. Such an interpretation of this segmentation problem allows a variety of

¹For clarity of presentation we ignore edge effects in the discrete formulations of (4),(5). Such edge effects only increase the notational complexity of the development.

statistically-motivated approaches to be taken. For example, in [18] a novel multiresolution estimation approach is used. In this work, however, we interpret the underlying spatial estimation problem as an equivalent *dynamic* estimation problem. Based on this interpretation we are able to formulate an efficient, statistically-based recursive solution. This statistically-based approach provides not only estimates of the fields themselves, but also corresponding measures of uncertainty. For simplicity we focus on the problem of processing two-dimensional MRI slices – i.e. images. The development to follow, however, carries over naturally to the case of three (or even higher) dimensional fields (i.e. volumes), where the computational gains should be even greater. Such results are the focus of present research.

3 An Estimation-Theoretic Interpretation

We now provide an estimation-theoretic interpretation of the quadratic minimization problems (4) and (6) as equivalent recursive estimation problems. To this end, note that (4) and (6) have the following common form:

$$E(\mathbf{x}) = \sum_{i=1}^n \|\mathbf{d}(j) - \mathbf{x}(j)\|_{R_1^{-1}(j)}^2 + \|D_r \mathbf{x}(j)\|_{R_2^{-1}(j)}^2 + \|\mathbf{x}(j) - \mathbf{x}(j-1)\|_{Q^{-1}(j)}^2 \quad (7)$$

where the weighting matrices $R_1(j)$, $R_2(j)$, and $Q(j)$ are diagonal, and the vector $\mathbf{x}(j)$ contains the elements from column j of the overall unknown desired image vector $\mathbf{x} = [\mathbf{x}^T(1) | \dots | \mathbf{x}^T(n)]^T$.

Now, it is straightforward to show [21–23] that the value of $\hat{\mathbf{x}}$ or, equivalently, $\hat{\mathbf{x}}(j)$ for each j , that minimizes the quadratic energy (7) is the same as the maximum-likelihood estimate of $\mathbf{x}(k)$ for each k based on the following set of observations and constraints:

$$\mathbf{x}(j) = \mathbf{x}(j-1) + \mathbf{q}(j-1), \quad \mathbf{q}(j) \sim \mathcal{N}(0, Q(j)) \quad (8)$$

$$\mathbf{y}(j) = C(j) \mathbf{x}(j) + \mathbf{r}(j), \quad \mathbf{r}(j) \sim \mathcal{N}(0, R(j)) \quad (9)$$

for $j = 1, \dots, n$, where $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, P)$ denotes a Gaussian random vector with mean \mathbf{m} and covariance P , the initial condition is given by $\mathbf{x}(0) = 0$ and $Q(0) = \lim_{\alpha \rightarrow \infty} \alpha I$, and the variables $\mathbf{y}(j)$, $C(j)$ and $R(j)$ are defined as:

$$\mathbf{y}(j) = \begin{bmatrix} \mathbf{d}(j) \\ \mathbf{0} \end{bmatrix}, \quad C(j) = \begin{bmatrix} I \\ D_r \end{bmatrix}, \quad R(j) = \begin{bmatrix} R_1(j) & \mathbf{0} \\ \mathbf{0} & R_2(j) \end{bmatrix} \quad (10)$$

Given the form we have written (8)–(9) in we can see that another, equivalent interpretation of these equations specifies a dynamic estimation problem for the $\mathbf{x}(j)$. In particular, the components $\hat{\mathbf{x}}(j)$ of the minimizer $\hat{\mathbf{x}}$ of (7) are *precisely the same* as the series of smoothed estimates of $\mathbf{x}(j)$ based on the dynamic equation (8) and the observation (9) for $1 \leq j \leq n$. The advantage of the formulation (8)–(9) is that we may use efficient, recursive Kalman filtering based techniques for its solution. In effect, by recursively processing $\mathbf{x}(j)$ we are recursively estimating the columns of the segmented image. In effect, we trade solving a single large problem (estimating the entire image) for solving a series of smaller problems (estimating the values in a column). In addition, these estimation-theoretic methods provide not only estimates of the field, but also corresponding measures of uncertainty. Finally, the statistical smoothing view of (8)–(9) suggests a statistical interpretation of the parameters of the original formulation, which can provide a rational guide to their selection in practice.

4 A Recursive Filter-Based Solution

In this section we show how to efficiently solve (8),(9). First note that what is needed is the optimal estimate $\hat{\mathbf{x}}(j)$ for each j based on *all* the data $\mathbf{y}(k)$, $k = 1, \dots, n$. This is known as the optimal *smoothed* estimate [24]. There are a number of efficient ways to generate this estimate, as described in [25–27]. In this work we use the Mayne-Fraser two-filter form of the optimal smoother [25, 26]. In this approach, the smoothed estimates and associated error variances are obtained by combining the outputs of two independent Kalman filters based on (8),(9) – one starting at column $j = 1$ and running forward (i.e. left to right) and the other starting at column $j = n$ and running backward (i.e. right to left), as depicted in Figure 3. Since each filter is independent, they may be run in parallel for greater throughput. The separate forward and backward estimates are then combined by weighting them by the inverse of their respective error variances, as detailed in [25, 26]. The key, then, is to find an efficient method for performing the independent filtering steps, which we focus on next. In particular, we focus on the forward filtering step. The backward step is equivalent and the weighted averaging operation straightforward.

4.1 An Information Form Filter

To perform the filtering based on (8),(9) we will use an efficient version of the *information form* of the Kalman filter developed in [22]. In particular, an optimal information-form filtering algorithm for the system (8)–(10) is provided by the following recursive algorithm [22]:

Information Filter Algorithm:

Prediction:

$$\bar{L}(j) = Q^{-1}(j) - Q^{-1}(j) \left(Q^{-1}(j) + \tilde{L}(j-1) \right)^{-1} Q^{-1}(j), \quad (11)$$

$$\bar{\mathbf{x}}(j) = \tilde{\mathbf{x}}(j-1), \quad (12)$$

$$\bar{\mathbf{z}}(j) = \bar{L}(j)\bar{\mathbf{x}}(j) \quad (13)$$

Update:

$$\tilde{L}(j) = \bar{L}(j) + R_1^{-1}(j) + D_r^T R_2^{-1}(j) D_r, \quad (14)$$

$$\tilde{\mathbf{z}}(j) = \bar{\mathbf{z}}(j) + R_1^{-1} \mathbf{d}(j) \quad (15)$$

$$\tilde{L}(j)\tilde{\mathbf{x}}(j) = \tilde{\mathbf{z}}(j) \quad (16)$$

where D_r is the derivative operator defined in (5) and note that $R_1(j)$, $R_2(j)$, and $Q(j)$ are diagonal matrices. The matrices $L(j)$ appearing in the above algorithm are the *information matrices* corresponding to the estimate, and are equal to the inverse of the estimation error covariance for the filtered estimate of column j . The filtered solution to (8)–(10) provided by the use of (11)–(16) is exact and, when optimally combined with the corresponding backward filtered estimates provides an estimate which is the same as the minimizer of (7). However, the above algorithm and interpretation break the problem into a series of smaller recursive steps and also provide measures of uncertainty for the estimates through the associated estimation error variances, which are the diagonal elements of $L^{-1}(j)$.

For the sake of computational efficiency we go even further and approximate this exact solution. In particular, as done in [22], we use a two term series expansion of the matrix inverse in parentheses and replace the exact expression (11) for the predicted information matrix with:

$$\bar{L}(j) \approx Q^{-1}(j) - Q^{-1}(j) \left(\Lambda^{-1}(j) - \Lambda^{-1}(j)\Omega(j)\Lambda^{-1}(j) \right) Q^{-1}(j) \quad (17)$$

where $\Lambda(j)$ is the diagonal part of $Q^{-1}(j) + \tilde{L}(j-1)$ and $\Omega(j)$ is the remaining, off-diagonal part. This simple approximation to (11) yields results close to the exact solution of the variational formulation in most cases. If greater fidelity to the original variational formulation is required additional terms may certainly be added to the above approximate expansion, at the cost of more computation, larger models, and greater storage requirements. The fidelity of such approximations will depend on quantities such as the specific choice of parameters for the original variational problem along with the structure of the image. Such relationships are a topic of current investigation.

Finally, note that calculation of the updated estimate in (16) requires the solution of a system of equations. While this system of equations is much smaller than that appearing in the original problem, it can still be a computationally intensive task. Fortunately, the structure of this segmentation problem coupled with the approximation (17) combine to ensure that $\tilde{L}(j)$ will remain sparse and banded (in particular, tridiagonal). In this case, the inversion operation implied in (16) can itself be efficiently performed via iterative methods, such as multigrid [20] or preconditioned conjugate gradient [19]. In our present implementation we use the conjugate gradient method preconditioned with the diagonal of $\tilde{L}(j)$.

4.2 The Smoothed Estimate

Once the separate forward and backward filtered estimates are found using (8),(9) together with the efficient variant of (11)–(16), the individual estimates are then combined to obtain the resulting smoothed estimate $\hat{\mathbf{x}}(j)$ together with its uncertainty. Let $\hat{\mathbf{x}}_f(j)$ and $\hat{\mathbf{x}}_b(j)$ denote the output of the forward and backward filtering operations, respectively, and let $L_f(j)$ and $L_b(j)$ denote the information matrices corresponding to these filtered estimates. Then the overall smoothed estimate at column j , $\hat{\mathbf{x}}(j)$, is obtained as the solution of:

$$[L_f(j) + L_b(j)] \hat{\mathbf{x}}(j) = L_f(j) \hat{\mathbf{x}}_f(j) + L_b(j) \hat{\mathbf{x}}_b(j) \quad (18)$$

which we again solve using the diagonally preconditioned conjugate gradient technique. The corresponding error variances are given by the diagonal elements of $(L_f(j) + L_b(j))^{-1}$, which we approximate by the reciprocal of the corresponding diagonal elements of $L_f(j) + L_b(j)$. These error variances give us a measure of the uncertainty of the corresponding estimates.

5 Examples

In this section, we present numerical examples of segmented MRI imagery using the techniques we have described. We use the global segmentation algorithm shown in Figure 1 with the energy terms $E_s(\mathbf{f})$ and $E_f(\mathbf{s})$ described in (4) and (6), respectively. We initialize the \mathbf{s} field estimate to zeros and terminate the iteration when the percent change in $E(f, s)$ falls below a threshold. For the experiments shown here we used a 2% change as the indication of convergence. This structure is common to all the examples. What will differ is how the subproblems of minimization of $E_s(\mathbf{f})$ or $E_f(\mathbf{s})$ are performed. In particular, we compare straightforward solution of (4),(6) to the proposed recursive estimation-theoretic based technique. In both cases note that the formulation of (1) assumes that the value of the edge field s lies in the range $[0, 1]$ while our solutions to the subproblem (6) contain no such constraint. To ensure the global iterations are well defined we simply truncate the edge function to the range $[0, 1 - \epsilon]$ for some small ϵ at each iteration if necessary. In practice, the solution to the unconstrained problem almost always lie in the required range anyway.

Comparison of estimates obtained with direct and recursive approaches

Here we compare images segmented using the algorithm of Figure 1 and exact solution of the normal equations to those obtained using the proposed efficient recursive technique. For these experiments we used parameter values of $\mu = 1$, $\lambda = 3$, $\nu = 1200$, and $\rho = \frac{1}{2}$. In Figure 4 we show the original proton density weighted MR image. In Figure 5, the two segmentations are shown. On the left is the solution obtained through direct minimization of the energy functionals (4) and (6) via exact solution of the corresponding normal equations or, equivalently, Euler equations. On the right is the corresponding solution for both \mathbf{f} and \mathbf{s} obtained by using the near optimal recursive approach. The images show that the recursive technique produces results that are almost identical to the direct solutions. To emphasize this point, in Figure 6 we show a histogram of the differences between the field estimates obtained by the two techniques. These histograms show the relative frequency (i.e. the histogram area equals one) of the corresponding errors as a percent of the corresponding full scale value. As can be seen all the values are clustered around zero. Both segmentations appear to do well in capturing the main boundaries between gray and white matter despite the presence of a significant gradient in absolute intensity across the image. In Figure 8 we show the result of thinning the thresholded edge field obtained in Figure 5.

Uncertainty Estimates

While the segmentations obtained by direct energy minimization and our recursive approach are quite similar, since our Kalman-filter-based technique is derived from an estimation-theoretic interpretation of the segmentation problem, we also generate associated error variance information, and thus measures of segmentation uncertainty. Such measures are important not only for the direct evaluation of the estimates themselves, but also for the fusion of such estimates with other sources of information, which we demonstrate later. While many approaches have been proposed for the segmentation of such MRI imagery, few have addressed this issue of error statistics. In Figure 9 we show the error standard deviation fields corresponding to the estimates in Figure 5. On the left is the standard deviation field corresponding to the smoothed field while on the right is the standard deviation field corresponding to the edge field. As can be seen, the error standard deviations for the field itself increase near edges while those of the edge field decrease near edges, as we would expect. In areas where the edges are weak, the corresponding error field indicates low reliability. These error measures are obtained as a by product of the processing in the recursive technique, and thus are obtained for free.

Computational Cost

We compare the computational cost of solving the segmentation problem represented in Figure 1 both through direct minimization of the energies and by our recursive method. The direct method finds the minimum of the energies (4),(6) at each iteration by explicitly solving the resulting system of so called normal or Euler equations. In practice, such large systems of equations are usually solved via iterative techniques, such as preconditioned conjugate gradient. Thus, for our comparison here we find the minimum of the energies (4),(6) at each stage using the conjugate gradient technique preconditioned with the diagonal of the corresponding normal equations [19]. Our recursive approach has already been described.

First we consider only the field estimates. In this case, both techniques took 4 global iterations to converge. Our recursive approach required approximately 20% fewer computations per iteration compared to the direct minimization approach. These solutions were obtained in a few minutes on

a SPARC 20 using unoptimized MATLAB code.

Next we compare the cost of calculating *both* the estimates of \mathbf{f} and \mathbf{s} together with the corresponding error measures (i.e. the error variances). For the direct approach, this calculation essentially corresponds to finding the diagonal elements of the *inverse* of the large matrix arising in the normal equations of the optimization problem, and since it requires inverting the matrix, is extremely costly. In contrast, for our recursive approach these error measures are obtained as a by product of the processing and thus are effectively free. In particular, in computing both these quantities the direct approach required a factor of approximately 10^4 more computations than our recursive method.

Fusion of Multichannel Data

Here we demonstrate the use of the error statistics provided by our technique for the fusion of edge information from multichannel data, in particular from registered T_1 , T_2 and PD imagery. Our goal here is to demonstrate the potential usefulness of such uncertainty information through a prototype fusion problem.

There are at least two ways our approach can accomplish this fusion of edge information. The first method is based on the direct use of the three channels in a single overall variational formulation, as proposed in [8]. In this case all three channels are simultaneously used to directly obtain a *single* edge field. This single edge map is used as the basis for the generation of a piecewise smooth estimate of the intensity of each separate channel. Our approach to this method also provides the associated uncertainty measure for the aggregate edge map and individual channel estimates, and thus provides one rational means to obtain such overall measures for multichannel data. In particular, note that the measures obtained for each channel will reflect the effects of the information in all the channels. Since this approach is a direct extension of that in [8] we will not pursue it here.

An alternative approach to the fusion of the edge or segmentation information from such multichannel data is to perform a segmentation on each channel separately and then use the resulting uncertainty measures obtained for each edge image to subsequently fuse the individual edge maps. In effect, this approach treats the separate edge field estimates from each channel as independent observations of the edge field with uncertainty provided by the associated error variance measure. Such an approach serves as a paradigm of a variety of similar fusion problems, in which it is desired to combine information obtained from a variety of sources. In particular, suppose at a given image pixel location (i, j) , each channel provides a corresponding edge field estimate at that pixel, denoted by s_{T_1} , s_{T_2} , s_{PD} , respectively. In addition, suppose that $\sigma_{T_1}^2$, $\sigma_{T_2}^2$, σ_{PD}^2 are the corresponding error variances associated with that pixel value as, for example, provided by our technique. If these edge estimates are uncorrelated from pixel to pixel and from channel to channel, then the optimal linear estimate of the overall edge field value at that pixel is provided by the quantity:

$$s = \frac{\sigma_{T_1}^{-2}s_{T_1} + \sigma_{T_2}^{-2}s_{T_2} + \sigma_{PD}^{-2}s_{PD}}{\sigma_{T_1}^{-2} + \sigma_{T_2}^{-2} + \sigma_{PD}^{-2}} \quad (19)$$

and the associated overall error variance σ^2 for the combined edge field estimate at that pixel is given by:

$$\sigma^2 = \frac{1}{\sigma_{T_1}^{-2} + \sigma_{T_2}^{-2} + \sigma_{PD}^{-2}} \quad (20)$$

As stated above, this assumes that the edge estimates obtained for each channel correspond to uncorrelated random variables. While this will not be the case, in general, the formulas (19) and

(20) still provide a rational way to combine such data. Note that if other sources of edge information are available with associated uncertainty measures, they may also be included in this framework.

Figure 10 shows T1 and T2 weighted MRI images that correspond to the PD weighted image in Figure 4. We generated separate edge estimates for each of these images using our technique. For the T1 image we used the parameters $\mu = 1$, $\lambda = 3$, $\rho = \frac{1}{2}$ and $\nu = 1000$, while for the T2 image we used the parameters $\mu = 1$, $\lambda = 3$, $\rho = \frac{1}{2}$ and $\nu = 1000$. In Figure 11 the resulting estimated edge fields for the T1 and T2 images are shown. The edge field for the PD weighted image was shown on the right in Figure 5. We then combined these individual edge estimates and corresponding uncertainty measures using the formulae in (19) and (20). The resulting fused edge image is shown on the left of Figure 12. Note that it contains edge information that does not exist in the edge image of any single channel. For instance, in the fused edge estimate, the boundary indicated by the arrow has been reinforced, indicating the incorporation of collateral edge information, since the same boundary in the PD edge estimate is incomplete. Note also how the estimation error, shown on the right of Figure 12 decreases as more information is added to the fused estimate of the edge field and more confidence in its value is developed.

6 Conclusions

In this paper we have presented a recursive approach to the segmentation of MRI imagery. This approach was based on an estimation theoretic interpretation of the segmentation problem and while efficient, provides solutions indistinguishable from direct solution of an equivalent variational problem. In addition to its efficiency, being statistically based the method also provides associated measures of uncertainty of the resulting estimates. Such measures are critical not only for evaluation of the segmentation but also for subsequent stages of the processing. An illustration of the use of such information was demonstrated through the fusion of edge information obtained from registered multichannel (PD, T1, and T2) imagery. This example demonstrated how the collateral information present in such data can be rationally combined when such measures of uncertainty are available. We are in the process of extending this approach to the 3-D case, which while providing greater computational challenges, should also provide greater gains. A systematic study of the robustness of this approach in a clinical setting together with the usefulness and interpretation of the associated uncertainty measures that are generated is also currently being undertaken.

References

- [1] V. S. Caviness, P. Filipek, and D. Kennedy. Magnetic resonance technology in human brain science: A blueprint for a program based upon morphometry. *Brain Development*, 11:1–13, 1989.
- [2] V. S. Caviness, P. Filipek, and D. Kennedy. Quantitative magnetic resonance imaging and studies of degenerative diseases of the developing human brain. *Brain Development*, 14 Suppl:S80–5, 1992.
- [3] P. Filipek, C. Richelme, D. Kennedy, J. Rademacher, D. Pitcher, S. Zidel, and V. S. Caviness. Morphometric analysis of the brain in developmental language disorders and autism. *Annals of Neurology*, 32:475, 1992.

- [4] M. Shenton, R. Kikinis, F. Jolesz, S. Pollak, M. LeMay, C. Wible, H. Hokama, J. Martin, D. Metcalf, M. Coleman, and et. al. Abnormalities of the left temporal lobe and thought disorder in schizophrenia. *New England Journal of Medicine*, 327(9):604–12, 1992.
- [5] Maeder and et al. Volumes of chromis traumatic frontal brain lesions measured by MR imaging and cbf tomography. *Acta Radiologica*, 32, 1991.
- [6] D. Kennedy, P. Filipek, and V. Caviness. Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 8(1):1–7, 1989.
- [7] R. Kikinis, M. Shenton, G. Gerig, J. Martin, M. Anderson, D. Metcalf, C. Guttman, R. McCarley, W. Lorensen, H. Cline, and et. al. Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging. *Journal of Magnetic Resonance Imaging*, 2(6):619–29, 1992. Department of Radiology, Harvard Medical School, Brigham and Womens Hospital. English Journal-article.
- [8] H. H. Pien and J. M. Gauch. Variational segmentation of multi-channel MRI images. In *IEEE International Conf. on Image Processing*, Austin, Texas, November 1994.
- [9] J. Shah, H. Pien, and J. Gauch. Recovery of surfaces with discontinuities by fusing shading and range data within a variational framework. *IEEE Trans. Imag. Proc.*, 5(8), August 1996.
- [10] D. Mumford and J. Shah. Boundary detection by minimizing functionals, I. In *Proc. IEEE Conf. on Comp. Vis. and Patt. Recog.*, pages 22–26, 1985.
- [11] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. Technical Report CICS-P-88, Center for Intelligent Control Systems, MIT, 1988.
- [12] J. Shah. Segmentation by nonlinear diffusion, II. In *Proc. IEEE Conf. on Comp. Vis. and Patt. Recog.*, pages 644–647. IEEE, 1992.
- [13] A. Blake and A. Zisserman. *Visual Reconstructions*. MIT Press,, Cambridge, MA, 1987.
- [14] P. Perona and J. Malik. Scale-space edge detection using anisotropic diffusion. *IEEE Trans. Patt. Anal. & Mach. Intell.*, 12(7), July 1990.
- [15] H. Derin, H. Elliot, R. Cristi, and D. Geman. Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields. *IEEE Trans. Patt. Anal. & Mach. Intell.*, 6:707–720, 1984.
- [16] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. & Mach. Intell.*, 6:721–741, 1984.
- [17] R. Azencott and B. Chalmond. Nuclear magnetic resonance imagery restoration by Markov modelization and annealing. volume 864, page 555.
- [18] M. Schneider, P. Fieguth, W. C. Karl, and A. S. Willsky. Multiscale statistical methods for the segmentation of signals and images. *IEEE Trans. Imag. Proc.*, 1998. In review.
- [19] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Press, 1989.

- [20] W. Briggs. *A Multigrid Tutorial*. Society for Industrial and Applied Mathematics, Philadelphia PA, 1987.
- [21] Frank L. Lewis. *Optimal Estimation*. John Wiley & Sons, New York, 1986.
- [22] T. M. Chin, W. C. Karl, and A. S. Willsky. Sequential filtering for multi-frame visual reconstruction. *Signal Processing*, 28(3):311–333, September 1992. Special issue on Multidimensional Signal Processing.
- [23] Richard Szeliski. *Bayesian Modeling of Uncertainty in Low-level Vision*. Kluwer Academic Publishers, Norwell, Massachusetts, 1989.
- [24] H. Van Trees. *Detection, Estimation, and Modulation Theory*. John Wiley and Sons, New York, NY, 1968.
- [25] A. Gelb. *Applied Optimal Estimation*. MIT Press, Cambridge, MA, 1974.
- [26] D. Fraser and J. Potter. The optimum linear smoother as a combination of two optimum linear filters. *IEEE Transactions on Automatic Control*, 7(8):387–90, August 1969.
- [27] H. E. Rauch, F. Tung, and C. T. Striebel. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445–50, aug 1965.

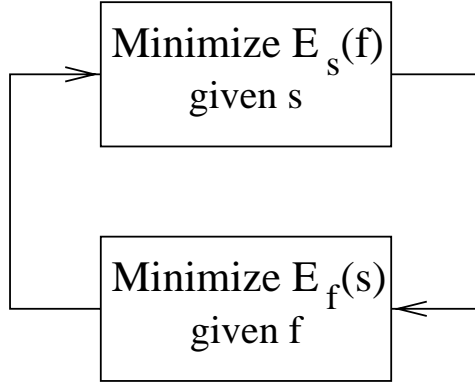


Figure 1: Segmentation Algorithm

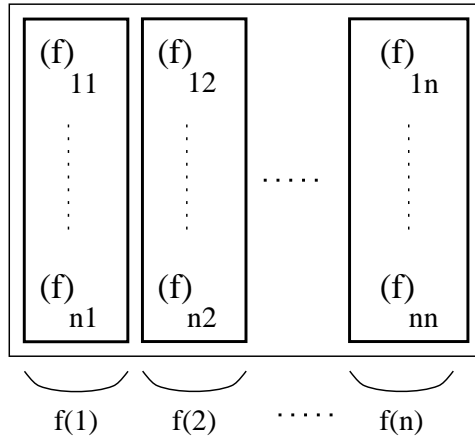


Figure 2: Sampled values.

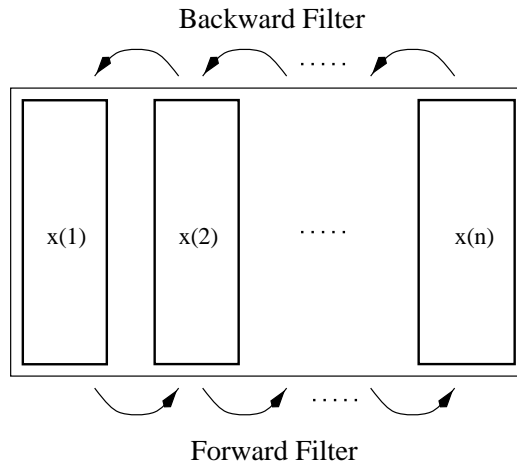


Figure 3: Recursive segmentation via recursive smoothing.

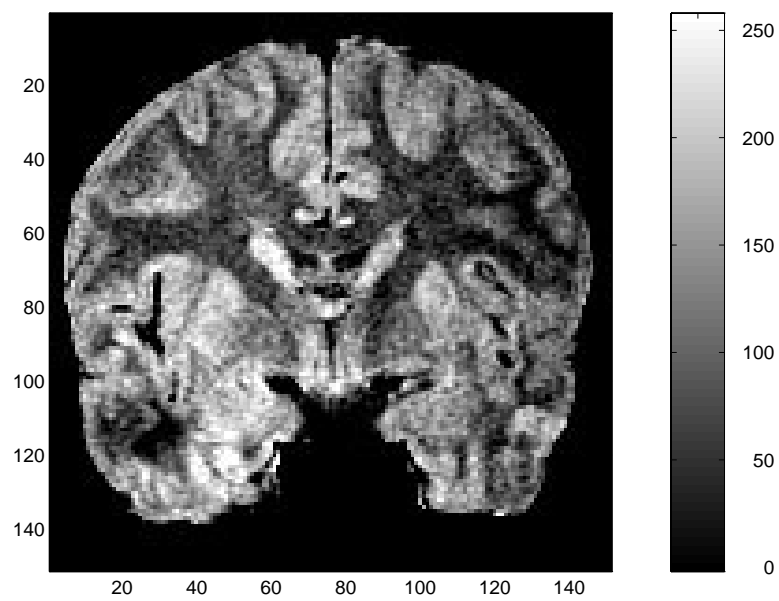


Figure 4: Original PD weighted image

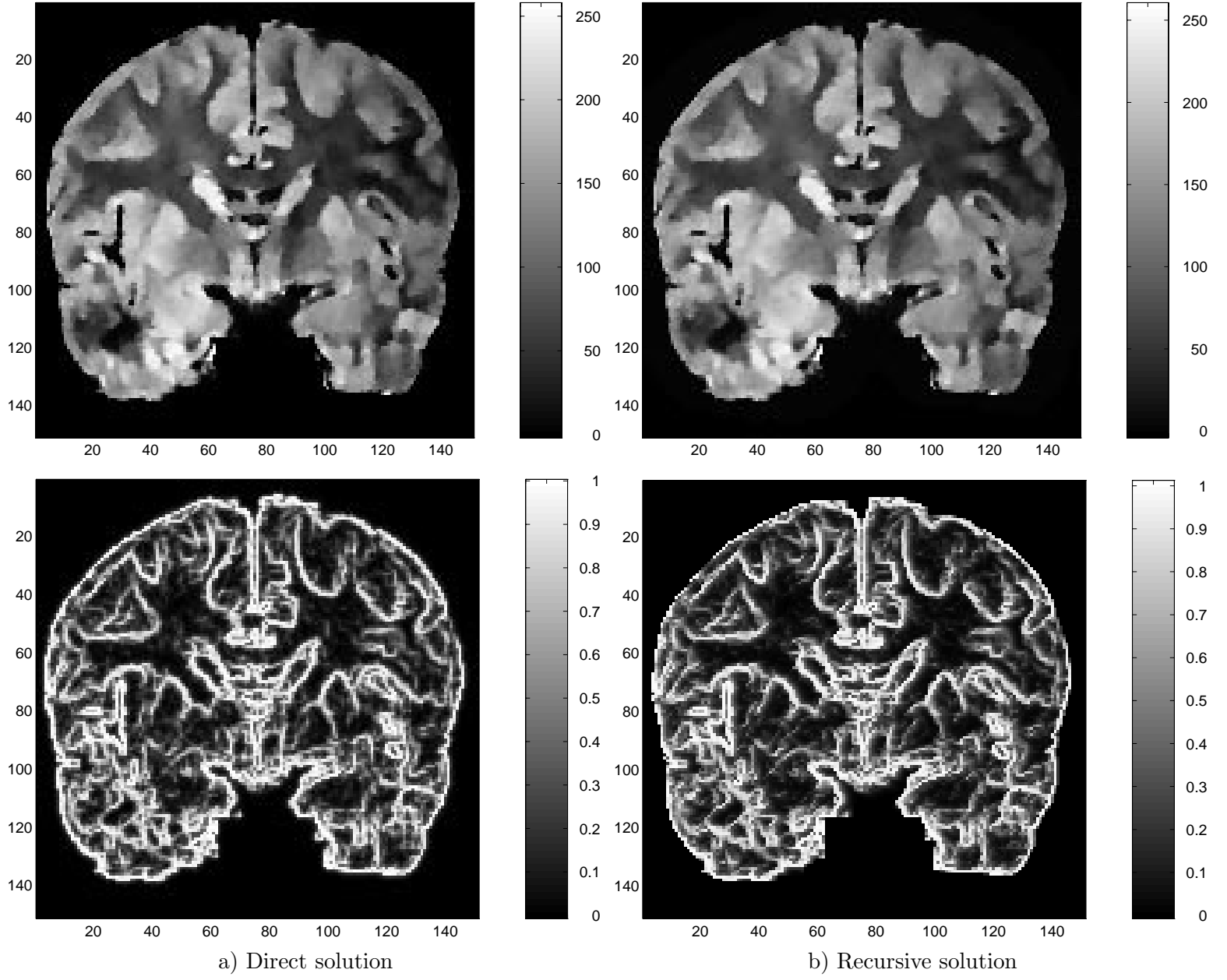


Figure 5: Comparison of answer obtained by direct solution of the normal equations and our recursive estimation-based approach. The top figures are the \mathbf{f} field estimates and the bottom figures are the corresponding \mathbf{s} field estimates.

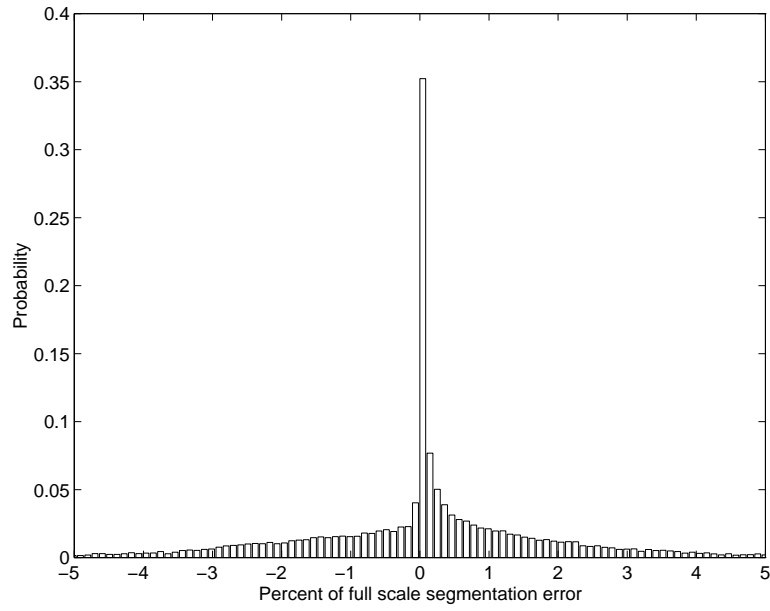


Figure 6: Histogram of difference between field estimates based on direct solution of the normal equations and our recursive estimation-based approach.

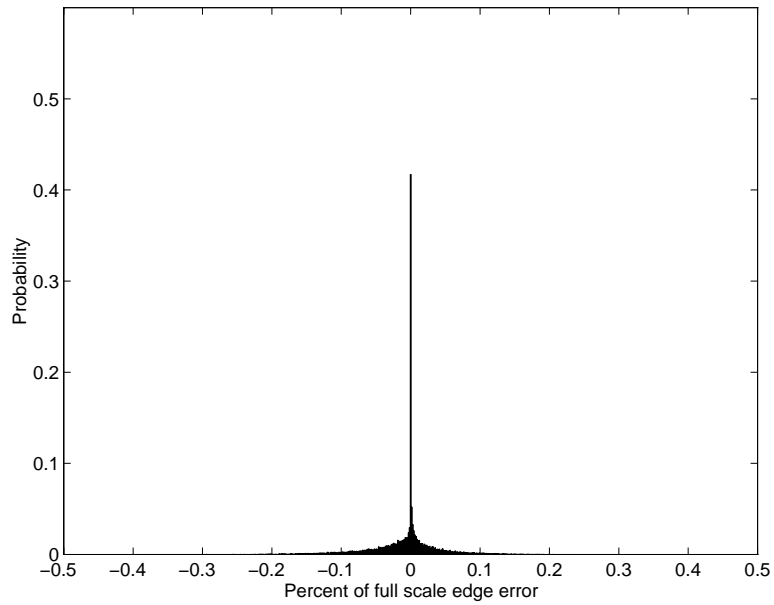


Figure 7: Histogram of difference between edge estimates based on direct solution of the normal equations and our recursive estimation-based approach.

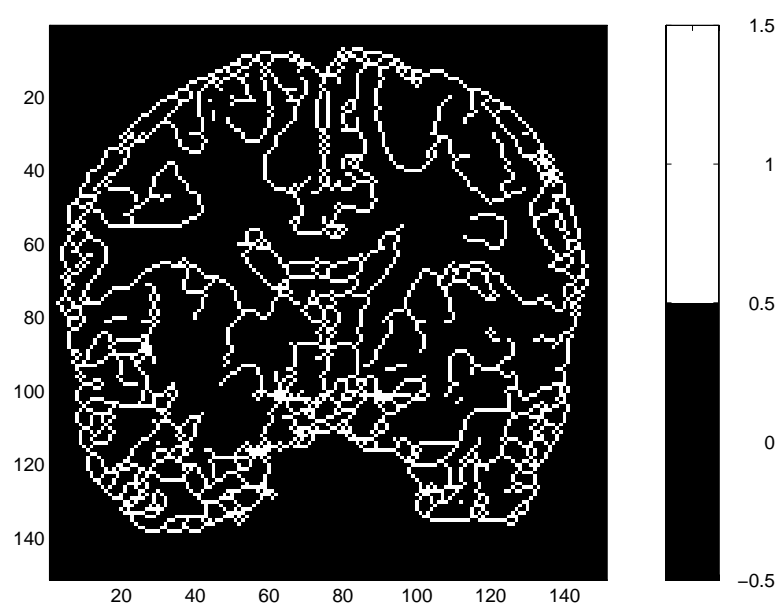


Figure 8: Thinned edge field

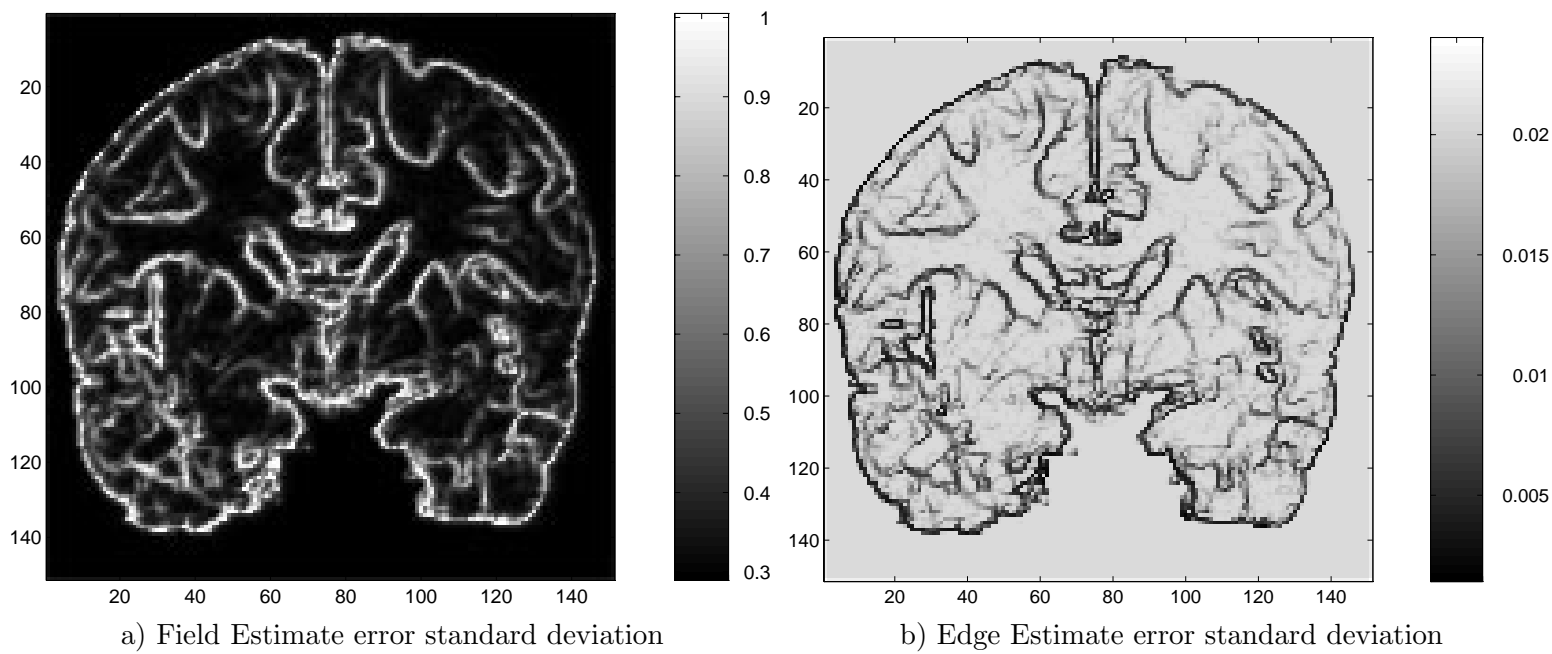


Figure 9: Field and Edge Estimate Error Standard Deviations

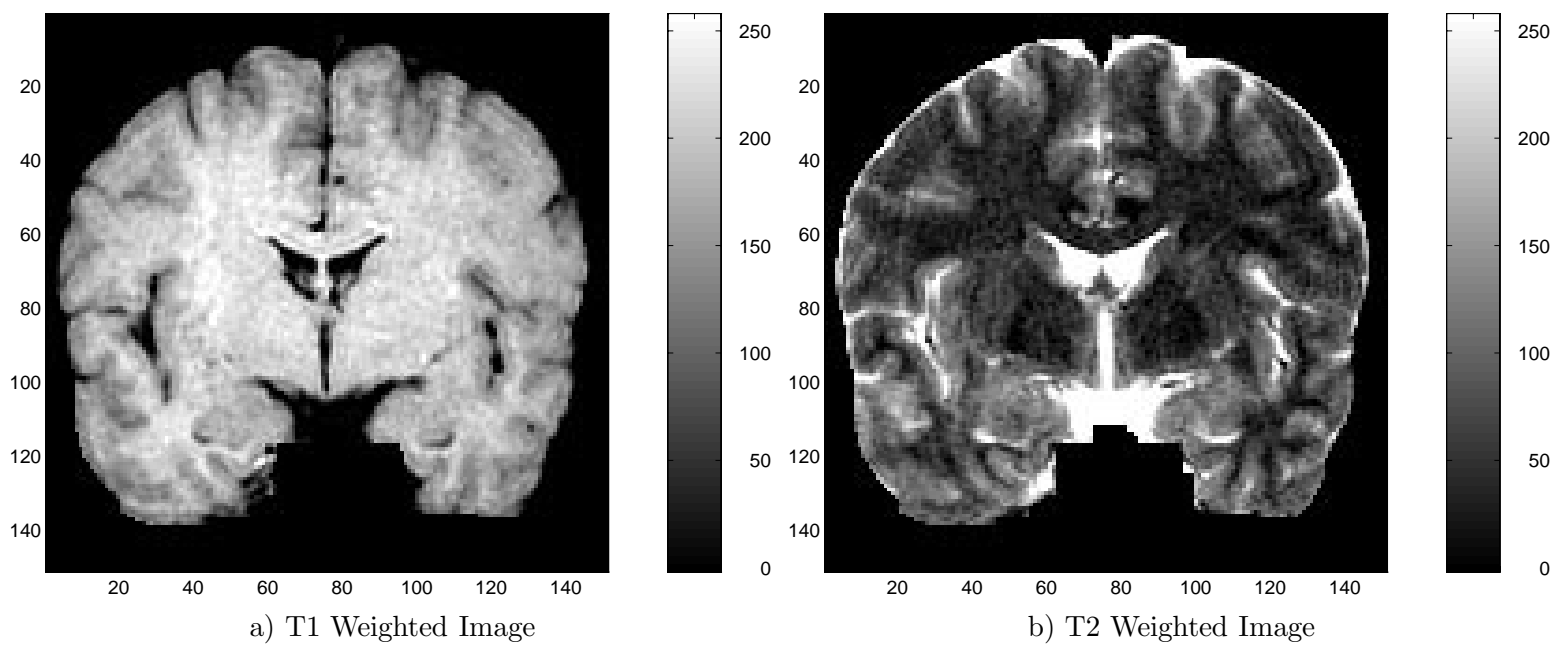


Figure 10: T1 and T2 weighted images.

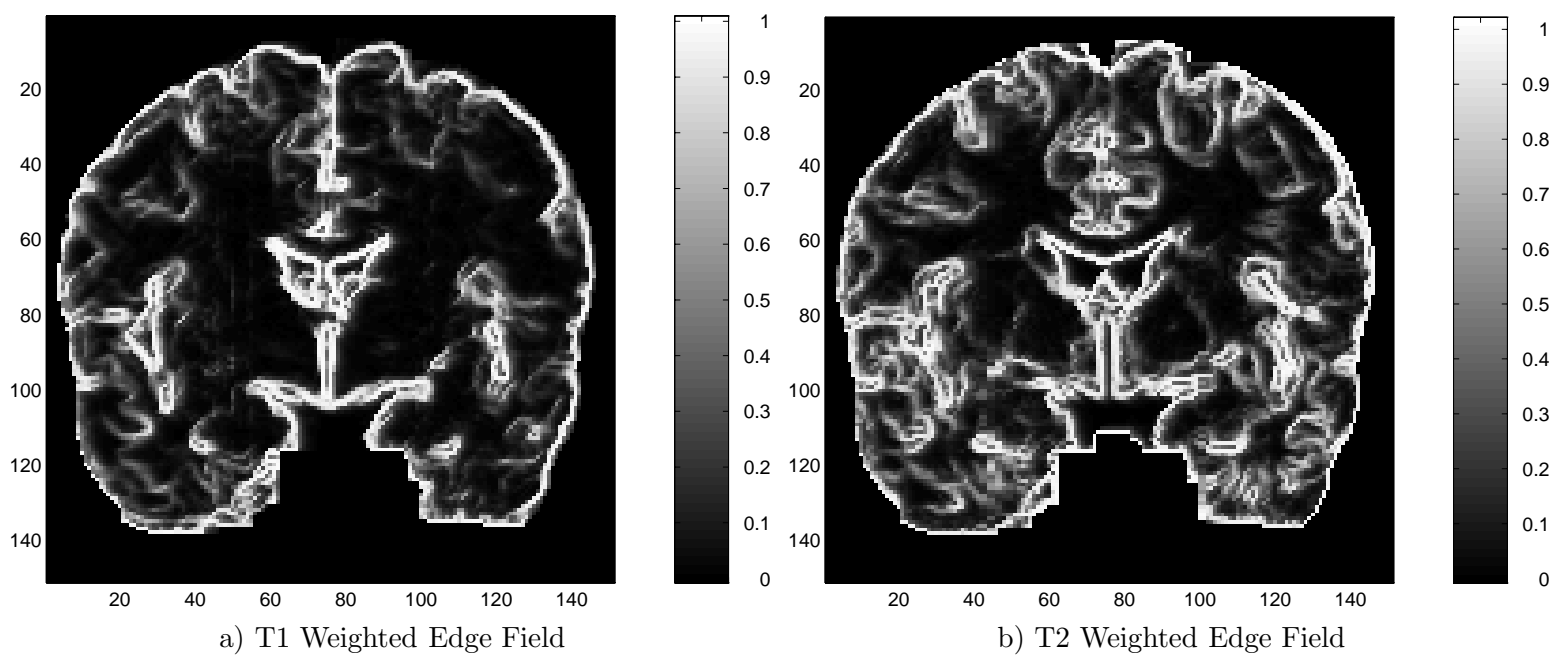


Figure 11: T1 and T2 weighted edge estimates.

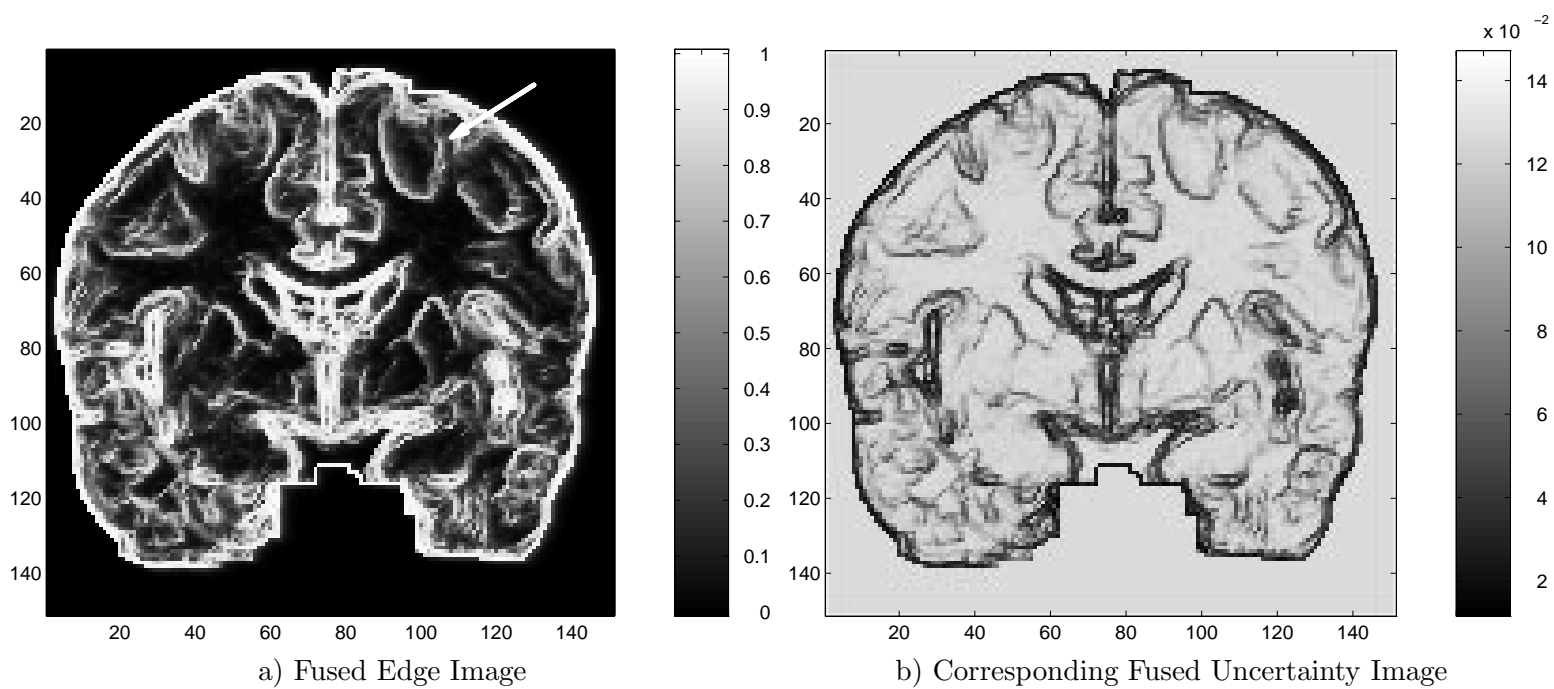


Figure 12: Fusion of edge estimates and standard deviation of estimation error

A Nested Recursive Approach to MAP Estimation based on Gauss-Markov Random Fields *

J. Kaufhold

BME Dept., Boston University
44 Cummington St, Boston MA

W. C. Karl

ECE and BME Depts., Boston University
8 St Mary's Street, Boston MA

D. A. Castañón

ECE Dept., Boston University
8 St Mary's Street, Boston MA

Abstract

Many large multidimensional space-time signal processing and data inversion applications (e.g. deconvolution) require some form of regularization to extract meaningful information. A popular approach to regularizing such problems in a statistical context is via a Gauss-Markov Random Field (GMRF) prior model in a Maximum A Posteriori (MAP) estimation framework. While providing good reconstructions, the high dimensionality of these problems can lead to prohibitive computational constraints which limit their practical applicability, particularly in real or near-real time applications. It has recently been shown that GMRF models possess a particular recursive structure. Conversely, complementary work in suboptimal filtering has been based on reduced order GMRF modeling. In this work, we combine these two results to present a suboptimal filter design which repeatedly takes advantage of this recursive GMRF structure to subdivide a large problem into a series of smaller, more tractable problems. In this way we present a method for approximate, model-based, recursive solution to such high dimensional problems based on their inherent recursive structure.

1 Introduction

Many large multidimensional space-time signal processing and data inversion applications require a method of regularization to extract meaningful information. These applications include large static spatial deconvolution problems, such as noise suppression in 3D diffusion-weighted (vector-valued) magnetic resonance imagery (MRI) and mine detection in EM ground sensor array data, as well as space-

time problems, such as arise in physical oceanography, space-time inverse scattering problems, visual field reconstruction and functional MRI (fMRI). A popular approach to regularizing such problems in a statistical context is via a Gauss-Markov Random Field (GMRF) prior model in a Maximum A Posteriori (MAP) estimation framework. While such approaches provide good reconstructions, the difficulty in their direct application to large spatial or space-time problems is that their exact solution leads to impractically large and complex optimization problems due to high dimensionality. This limits the practical applicability and power of such GMRF-based formulations, particularly in real or near-real time applications. In this work we combine existing results on the recursive structure of GMRFs with reduced order model-based recursive filtering techniques to present a rational basis for approximate, model-based, recursive approaches to efficiently solve such problems.

2 Problem Statement

For simplicity, we will consider MAP estimation problems of the following form:

$$y = x + w, \quad w \sim \mathcal{N}(0, R) \quad (1)$$

$$x \sim \mathcal{N}(0, M) \quad (2)$$

where M corresponds to the covariance of a GMRF and R is the diagonal covariance of the observation noise, w . In particular, M^{-1} has a sparse banded structure reflecting the nearest neighbor interactions of the lexicographically ordered data [1]. The problem in (1) and (2) is representative of many estimation problems. For example, M is often chosen to capture a “smoothness prior” in which case $M^{-1} \approx D^T D$, where D is a derivative operator. The challenge posed by the applications we are considering is due to size and dimensionality. For example, consider the space-time problem posed by enhancing a 3D fMRI data set. In this problem, we would need

*This work was supported by a Whitaker Foundation Graduate Research Fellowship, the National Institutes of Health under Grant NINDS 1 R01 NS34189, by the Air Force Office of Scientific Research under Grant F49620-96-1-0028, and by the Army Research Office under Grant ARO DAAG55-97-1-0013.

to estimate over 10^9 elements (*e.g.* typical dimensions are 256 columns \times 256 rows \times 60 slices \times 60 time points). Direct solution or inversion of corresponding normal equations is currently prohibitive if not impossible.

3 Dynamic Reformulation

It is known that the *inverse* of the covariance matrix of a GMRF, M^{-1} in equation (2), termed the “potential matrix” in [3] and more commonly known as the “information matrix” [4], is sparse and banded with a particular structure. Recently, it was shown that the information matrix, M^{-1} of a GMRF possesses a certain highly interesting recursive structure [3]. Specifically, it was shown that the structure of the information matrix can be exploited to define an equivalent implicit autoregressive model on the subelements of the field of the form:

$$y_k = x_k + w_k, \quad w_k \sim N(0, R_k) \quad (3)$$

$$x_{k+1} = A_k x_k + v_k, \quad v_k \sim N(0, Q_k) \quad (4)$$

where the subelements x_k and y_k now define a “slice” (*e.g.* the rows or columns of an image) of the original field, and R_k and Q_k are diagonal covariance matrices with compatible partitions. For appropriate choices of R_k , A_k , and Q_k , It can be shown that that equations (3) and (4) yield the same overall covariance structure as the GMRF defined by M in the original problem. This equivalent dynamic model is derived through a Cholesky factorization of the corresponding information matrix, M^{-1} .

To understand this, suppose we have a symmetric positive definite block tridiagonal matrix, $M^{-1} = U$ of the following form:

$$U = \begin{bmatrix} U_{11} & U_{12} & & & \\ U_{12}^T & U_{22} & U_{23} & & \\ & U_{23}^T & U_{33} & & \\ & & & \ddots & \\ & & & & U_{(n-1)n}^T & U_{nn} \end{bmatrix}$$

where U is composed of block tridiagonal subblocks, U_{ii} on the block diagonal and diagonal subblocks, U_{ii+1} on the upper block off diagonal. Then we can uniquely decompose it as $M^{-1} = U = G^T G$ where G is the Cholesky factor of U and is of the form:

$$G = \begin{bmatrix} G_{11} & G_{12} & & & \\ & G_{22} & G_{23} & & \\ & & G_{33} & & \\ & & & \ddots & \\ & & & & G_{(n-1)n} & G_{nn} \end{bmatrix}. \quad (5)$$

The blocks of G are related to the blocks of U as follows: G_{11} is the Cholesky factor of U_{11} . $G_{ij} = G_{ii}^{-1} U_{ij}$ for all off-diagonal blocks where $i = j - 1$, and G_{jj} is the Cholesky factor of $U_{jj} - G_{ij}^T G_{ij}$ for all other diagonal blocks.

Equations (3) and (4) can be derived from equation (1) and the Cholesky factor of M^{-1} in equation (5). First, defining equation (3) entails decomposing the diagonal covariance, R , and the observation vector, y , of equation (1) into subelements indexed to be compatible with the partitioning the equation (4). Now equation (4) is defined using the Cholesky factor in equation (5). In particular, $Q_{n-i+1}^{-1} = G_{ii}^T G_{ii}$ and $A_{n-i+1} = (G_{ii}^T G_{ii})^{-1} G_{ij}$ where $j = i+1$. This fact allows us to recast (1) and (2) as an equivalent *dynamic* estimation problem of the form given in equations (3) and (4). Note that this equivalence is exact, with no approximations. Conceptually, this dynamic model, obtained via Cholesky factorization, enables the optimal processing of the field with efficient techniques such as Kalman filter-based smoothing (*e.g.* Rauch, Tung, Striebel; Mayne-Fraser) algorithms.

While this approach indeed transforms the original large problem into a sequence of smaller problems and while providing storage advantages, the overall amount of computation is still the same as that obtained if we exploited the banded MRF structure from the outset in the solution of the original complete normal equations. As we have argued, for the large practical problems we are considering, this amount of computation is still too large. Thus, to reduce the computational burden, some form of suboptimal approach is necessary. We discuss next a particular suboptimal approach to recursive filtering based on reduced order GMRF modeling of the estimation error field [1] which is particularly well-matched to our proposed solution.

4 GMRF-based Kalman Filter

In [1], an efficient suboptimal approach to Kalman filtering for large problems, such as in (3) and (4), is presented. In this development, a suboptimal variant of the information form of the Kalman Filter for (3) and (4), consisting of the following prediction and update steps is implemented:

PREDICTION STEP

$$D \triangleq \text{diag}(A_k^T Q_k^{-1} A_k + \hat{L}_k) \quad (6)$$

$$\Omega \triangleq (A_k^T Q_k^{-1} A_k + \hat{L}_k) - D \quad (7)$$

$$K_{k+1} = D^{-1} - D^{-1} \Omega D^{-1} \dots \quad (8)$$

$$\bar{L}_{k+1} = Q_k^{-1} - Q_k^{-1} A_k K_{k+1} A_k^T Q_k^{-1} \quad (9)$$

$$\bar{x}_{k+1} = A_k \hat{x}_k \quad (10)$$

$$\bar{z}_{k+1} = \bar{L}_{k+1} \bar{x}_{k+1} \quad (11)$$

UPDATE STEP

$$\hat{L}_{k+1} = \bar{L}_{k+1} + R_{k+1}^{-1} \quad (12)$$

$$\hat{z}_{k+1} = \bar{z}_{k+1} + R_{k+1}^{-1} y_{k+1} \quad (13)$$

$$\hat{L}_{k+1} \hat{x}_{k+1} = \hat{z}_{k+1} \quad (14)$$

where, in the filtering equations above, as applied to (3) and (4), K_{k+1} is a series expansion to $(A_k^T Q_k^{-1} A_k + \hat{L}_k)^{-1}$, \bar{L}_{k+1} is the predicted information matrix, \bar{x}_{k+1} is the predicted state, \hat{L}_{k+1} is the updated information matrix, and \hat{x}_{k+1} is the updated state as described in [1]. The suboptimal filter structure of [1] is based on the idea of imposing a reduced order GMRF-type model of the error field, as reflected in the structure of \bar{L}_{k+1} , the predicted information matrix in (9). In particular, \bar{L}_{k+1} is constrained to have a sparse, banded MRF structure with the number of bands reflecting an imposed reduced-order GMRF neighborhood. Notice that such a structure, once imposed on \bar{L}_{k+1} , is maintained throughout (10)-(14) and imparted to \hat{L}_{k+1} . In summary, at each iteration, \hat{L} and \bar{L} are constrained to have an MRF-like structure.

5 Nested Solution

The main computational bottleneck in the approximate implementation of (6)-(14) can be traced to solving the implicit equation (14) for the updated state \hat{x}_{k+1} . In [1], (14) was solved iteratively by Gauss Seidel Successive Over Relaxation (SOR), exploiting the sparse structure of \hat{L}_{k+1} . Note, however, that (14) can be viewed as *another* static spatial estimation problem, which is of the following form:

$$y_{k+1} = \hat{x}_{k+1} + \hat{v}_{k+1}, \quad \hat{v}_{k+1} \sim \mathcal{N}(0, R_{k+1}) \quad (15)$$

$$\hat{x}_{k+1} \sim \mathcal{N}(0, \bar{L}_{k+1}^{-1}) \quad (16)$$

where the prior model, \bar{L}_{k+1}^{-1} again corresponds to a covariance of a GMRF. This is true precisely because of the GMRF approximation imposed in (9). It is *not true* in the exact Kalman filtering equations where \bar{L}_{k+1} is full in general. The problem in (15) and (16) is of the same form as the overall problem stated in equations (1) and (2). Therefore, we can apply the same technique we applied to solve the original problem, by performing a Cholesky factor of \bar{L} to again repose this subproblem as a dynamic estimation problem amenable to Kalman smoothing. In

this way, we *nest* the recursive solution to the overall high-dimensional problem, casting it as a series of progressively lower dimensional problems. This is the original contribution of this work.

6 Approximations

6.1 Incomplete Cholesky

While conceptually, a Cholesky factorization of M^{-1} in (2) provides the dynamics necessary to define an equivalent recursive smoothing problem, performing such a factorization is computationally prohibitive. The computational cost of computing an exact Cholesky factor of the block tridiagonal matrices we are considering following the definition given in Section 3 is $O(nn_b^3)$ where n is the number of blocks on the block diagonal, and n_b is the dimension of each block. For instance, for the fMRI example discussed previously, n may correspond to 60 time points, and n_b would be the number of voxels in each volume. In addition to the prohibitive computational cost of computing the Cholesky factor, fill-in occurs between the diagonal and off diagonal bands, destroying the structure of the problem so carefully preserved for efficiency in the suboptimal Kalman filter design in (6) through (14). Thus, in keeping with the reduced order GMRF modeling philosophy, we instead find an incomplete Cholesky factor, G_{inc} [2], such that $G_{\text{inc}}^T G_{\text{inc}} \approx \bar{L}$. In particular, the incomplete Cholesky factorization we use maintains the structure of the nearest-neighbor approximation scheme set up in the information form Kalman filtering equations.

A range of incomplete Cholesky factorization algorithms appear in the computational linear algebra literature for preconditioning sparse linear systems. A problem with most incomplete Cholesky factorization algorithms is that their computational cost, although smaller than $O(nn_b^3)$, is not linear in the block size. For the problem sizes we are considering, we seek an incomplete Cholesky factorization with computational cost linear in the block size.

The prohibitive computation in this process described in Section 3 can be traced to the inverse required for computing every off-diagonal block of the factor; this inverse is also the operation which leads to fill-in. For diagonally dominant diagonal blocks, G_{ii} , this inverse may be well-approximated by the inverse of each block's diagonal, which has computational cost $O(nn_b)$ rather than $O(nn_b^3)$. This diagonal approximation to the inverse also imposes the overall sparsity pattern of the U matrix onto its factor, G . For applications we have examined, this approxima-

tion is usually reasonable.

6.2 Dynamic and Observation Equations

Although the exact dynamic reformulation of the GMRF MAP estimation problem in equations (1) and (2) is given in Section 3, and the reformulation is analogous for the incomplete Cholesky factor we compute, certain operations in the reformulation destroy the structure of the subproblems. In particular, the inverse in the definition of each A_k destroys its GMRF structure, making it full in general. Because Q_k^{-1} is constructed from the multiplication of two nondiagonal matrices, it is no longer diagonal in general. As has been the modeling philosophy to alleviate similar issues in the rest of this paper, to maintain the required sparsity patterns for Q_k^{-1} and A_k , we make sparse approximations to each. For Q_k^{-1} , we retain only its diagonal elements. For A_k , we make the same two-term series expansion approximation we make to K_{k+1} in equation (8) of the Kalman filter, which again, is a rational approximation to the inverse which preserves the sparsity pattern of the underlying GMRF model. However, even this approximation to A_k is not completely adequate for the filter structure as it is presented in equations (6) through (14). Specifically, the bandwidth of \bar{L}_{k+1} will grow in equation (9). To remedy this issue, the \bar{L}_{k+1} matrix is truncated at every step through the filter, preserving its structure to reflect the original GMRF structure of the overall problem.

7 Examples

We have implemented our nested recursive approach to MAP estimation of Markov Random Fields on a test volume. The small example problem is an edge-preserving smoothing operation. Three simultaneous cross-sections of the underlying volume, a sphere of ones on a background of zeros, are shown in a slice display format in Figure 1. This underlying volume is the state, x , we are trying to estimate in equations (1) and (2). This sphere is observed in white Gaussian additive noise with diagonal covariance, or $r \sim (0, R = 1/\mu I)$ in equation (1). The observation is $y = x + r$, shown in Figure (2). The prior model for the volume in equation (2) is a 3D GMRF prior where M^{-1} is a discrete approximation to the gradient operator weighted by a function of the given edge process shown in Figure (3), such that the 3D GMRF prior has a large covariance at the pixels indicated by the edges, and a small covariance at all

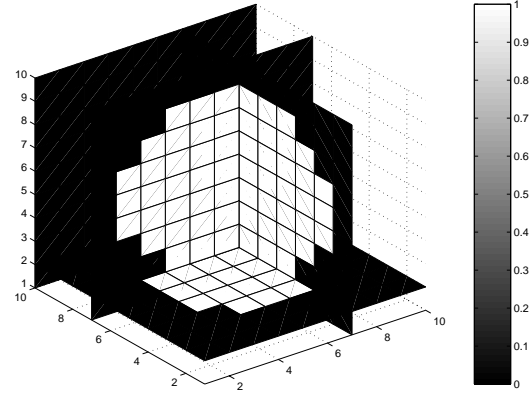


Figure 1: Original Volume

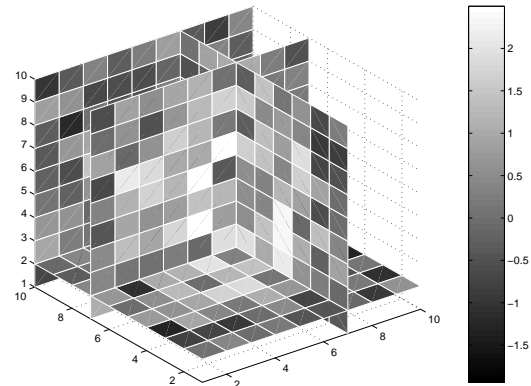


Figure 2: Observed Volume

other pixels. In Figure 4, we show an example of our nested recursive filtering method applied to processing the observed volume shown in Figure 2 given the normal equations corresponding to the GMRF model discussed above.

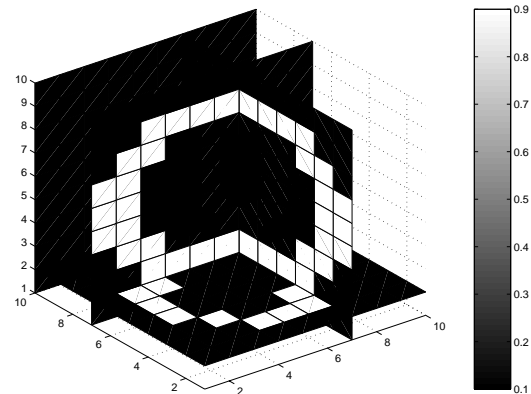


Figure 3: Given Edge Locations

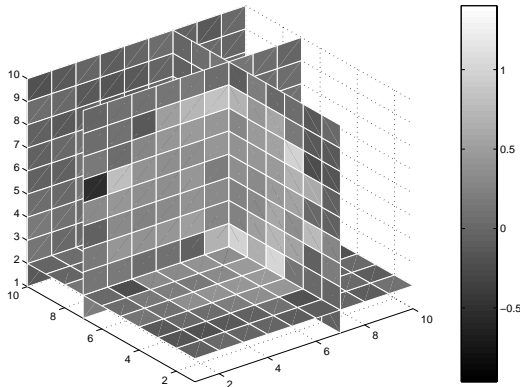


Figure 4: Restored Volume (by nested recursion)

8 Discussion

In the edge-preserving smoothing example above, we illustrate 3 concepts. Most importantly, it demonstrates Kalman filtering as a means for volumetric data processing. Specifically, it illustrates the idea of splitting the GMRF neighborhood structure in 3D into an equivalent neighborhood structure in 2D along with a prior term which depends on the previous slice of data. This splitting is powerful, and can be used for multidimensional data where the state dimension may be much larger. Secondly, for such an approach, the storage space for the sparse matrices required to define the dynamic and observation equations in the Kalman filtering grows linearly with the number of pixels in the volume. Finally, the computation is $O(n)$ rather than $O(n^3)$, where n is the total number of elements in the field. For computing simultaneously the approximate MAP estimate of the state as well as the approximate MAP estimation error. These are the primary motivations for the suboptimal nested recursive filtering approach.

The specific performance of our method depends on a host of factors, including the degree of diagonal dominance of the information matrices, and the order of the GMRF being processed. These factors are consequences of the overall parameterization of the GMRF. Thus, the parameterization of the GMRF impacts the approximation error of the proposed nested recursive method. Specifically, for GMRFs enforcing strong regularization (i.e. for a large local correlation strength and large neighborhood sizes), approximation error will be larger. The relationship between smoothing in the GMRF definition and approximation error is of practical importance and is a focus of current investigation.

9 Conclusion

We have combined a recursive interpretation of GMRF's with an approximate information form Kalman filter based precisely on reduced order GMRF spatial models to develop a nested and efficient recursive approach to solution of GMRF regularized spatial estimation problems. In particular, we exploit the specific structure of the overall problem to reduce a large optimization problem into a series of dynamic equations amenable to solution by Kalman filtering techniques. We choose a specific information form Kalman filtering technique which allows us to exploit the sparse banded structure of the information matrix to make rational approximations to the prediction and update step. These approximations preserve the sparse banded GMRF structure of the original problem. Preserving this structure allows us to reformulate subproblems arising in the information form Kalman filter as a series of yet smaller dynamic equations. This *nested* recursive structure of the overall problem is elegant and also serves to make the solution more efficient. For higher dimensional problems, preliminary calculations suggest that this method will save more computation than traditional techniques for solving spatial estimation problems.

References

- [1] T. M. Chin, W. C. Karl and A. S. Willsky. Sequential Filtering for Multi-Frame Visual Reconstruction. *IEEE Trans. Sig. Proc., Special Issue on Multidimensional Signal Processing*, 28(3):311–333, Sept 1992.
- [2] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Press, 1989.
- [3] J. Moura and N. Balram. Recursive Structure of Noncausal Gauss-Markov Random Fields. *IEEE Trans. on Information Theory*, 38(2):334–354, March 1992.
- [4] P. S. Maybeck. *Stochastic models, estimation and control*. Academic Press, New York, 1979.

Diffraction tomography for multi-monostatic ground penetrating radar imaging

Ross W Deming[†] and Anthony J Devaney

Center for Electromagnetics Research, Department of Electrical and Computer Engineering,
Northeastern University, Boston, MA 02115, USA

Received 31 May 1996

Abstract. A generalized diffraction tomographic (DT) algorithm is derived for subsurface imaging from multifrequency multi-monostatic ground penetrating radar (GPR) data. The algorithm is based on the Born approximation for vector electromagnetic scattering and incorporates realistic nearfield models for the receiving and transmitting antennas. The forward scattering model is inverted analytically using the regularized pseudoinverse operator to yield an algorithm for imaging the underground region based on scattered field measurements at a set of receiving antennas. Whereas the usual inversion algorithms of DT require a lossless background medium and ideal point sources and receivers, the algorithm described here allows an attenuating background and arbitrary transmitting and receiving antennas. The algorithm places no restrictions on the radar frequency, and can thus include shallow imaging applications where the wavelengths are on the same order as the depth of buried objects of interest. Versions of the algorithm are given for both the three dimensional and the 2.5-dimensional cases. Results are given of computer simulations designed to test the algorithm.

1. Introduction

In this paper we present a new approach for subsurface imaging from multifrequency multi-monostatic data measurements obtained using ground penetrating radar (GPR) in a reflection geometry. The method employs a linear scattering model for electromagnetic wavefields based on the Born approximation, which is inverted analytically to yield an image of the subsurface based on scattered field measurements. Thus, our imaging algorithm is related to the well established method of *diffraction tomography* (DT), which is used in various forms for such applications as optical inverse scattering [15], medical ultrasonic imaging [6, 14], and geophysical imaging [7, 8, 20, 26–29]. However, in contrast with many applications of DT [6, 7], in GPR imaging background losses are significant, and evanescent wavefield components are important since radar wavelengths are often times on the same order as the depth and size of underground objects of interest [22]. Therefore in this development, unlike the usual treatments of DT, we incorporate soil attenuation into the mathematical inversions, and include evanescent components to help combat the restrictions on image resolution imposed by the relatively low frequencies used. Moreover, we employ realistic nearfield models for the transmitting and receiving antennas, in contrast with most treatments of DT that employ ideal point sources and receivers. Our method can be used with bandlimited

[†] Presently employed at Nichols Research Corporation, 70 Westview Street, Kilnbrook IV, Lexington, MA 02173, USA. This publication represents research carried out by the author in the doctoral program at Northeastern University.

pulses, and can include a constant spatial offset between the transmitter and receiver in each monostatic experiment.

Several authors have addressed, within the context of DT, the problem of inverse scattering using multi-monostatic field measurements. In [19] exact inversion formulae, within the Born approximation, are derived using broadband multi-monostatic measurements conducted on planar, spherical, and cylindrical surfaces. The authors use an ideal point source and receiver approximation, and assume the distance to scattering objects is much greater than a wavelength. This treatment requires transmitted pulses which are not bandlimited although the authors suggest Wiener filtering as a means to circumvent this restriction. An iterative, time domain, approach to the multi-monostatic imaging problem is suggested in [25] based on the Born approximation. The background medium is assumed to be lossless and nondispersive, and the ideal point source and receiver approximation is used. In [27] DT algorithms for GPR imaging are derived and tested on experimental data. The Born approximation is employed, as well as the approximations that the background is lossless and the distance between antennas and scatterers is much greater than a wavelength. A target detection algorithm is also given which can incorporate an attenuating background. Efficient DT imaging methods are described in [29] using multifrequency multi-monostatic data for both constant and vertically varying backgrounds. Inversion formulae are given using both the Born and the physical optics approximations, assuming a lossless background medium and ideal point sources and receivers. With these algorithms the final image is formed by coherent superposition of subimages from each frequency, whereas we show in this paper that an optimum multifrequency algorithm must exploit the correlation between the data at different frequencies.

The approach employed in this paper consists of several steps. We first define a vector electromagnetic forward scattering model based on the Born approximation, which incorporates nearfield characteristics of the transmitting and receiving antennas via the scattering matrix models developed by Kerns [13]. This model then yields a coupled set of integral equations relating the data at each excitation frequency to the sought after ‘object function’ defining the complex index of refraction profile of the subsurface. After casting the integral equations in a mathematical operator formalism, we are then able to find the minimum L^2 -norm solution to the object function by deriving and applying the regularized pseudoinverse operator [2, 17, 21]. The method is found to be similar to *filtered backpropagation algorithms* of DT described in the literature [4, 6, 7], and like these algorithms it is quantitative, and is designed to yield a solution which is mathematically consistent with the measured data. The inversion is fully analytic, and therefore relatively efficient to compute.

In section 2 we derive the forward scattering model in detail, then in section 3 we derive the regularized inversion of this model for the case when the scatterer is three dimensional and monostatic experiments are conducted over the xy plane defining the ground surface. In section 4 we modify the three-dimensional method to the 2.5-dimensional case where the scatterer is invariant in the \hat{y} -direction and experiments are conducted along the \hat{x} -axis. The paper includes computed results from synthetic data in section 5, as well as a discussion of future directions for the methodology of this paper in section 6.

2. Forward model for electromagnetic scattering

By probing the earth with electromagnetic wavefields we wish to estimate the electrical permittivity distribution in the underground region $z < 0$ from scattered field measurements at the surface $z = 0$ (where the orientation of the coordinate axes is illustrated in figure 1).

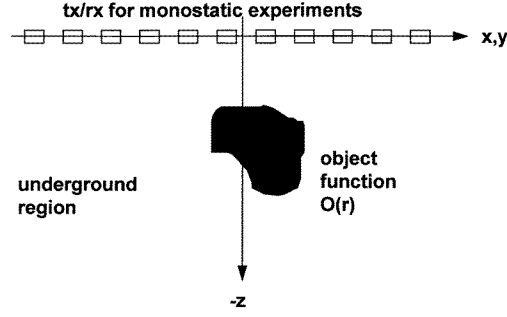


Figure 1. Geometry for multi-monostatic radar imaging.

The incident (probing) fields are generated by a ground penetrating radar (GPR) system operating in either a pulsed time domain or step frequency mode. Our scattering model is developed here in the frequency domain, related to the time domain through the standard Fourier and inverse Fourier transform definitions.

The GPR survey is assumed to consist of a number of monostatic experiments, each corresponding to a different location of the transmitting/receiving antenna on the ground surface, and each incorporating data collected over a band of frequencies ω . In each experiment the scattered field results from the interaction of the incident field with inhomogeneities in the subsurface, described by the *object function*

$$O(\mathbf{r}, \omega) = 1 - \frac{\epsilon(\mathbf{r}, \omega)}{\epsilon_0(\omega)}.$$

Here, $\mathbf{r} = (x, y, z)$ is the three-dimensional spatial coordinate and $\epsilon(\mathbf{r}, \omega) = \epsilon'(\mathbf{r}, \omega) + i\sigma(\mathbf{r}, \omega)/\omega$ is the complex permittivity in the underground, i.e. the soil with the embedded inhomogeneities. The quantity $\epsilon'(\mathbf{r}, \omega)$ is the real dielectric constant and $\sigma(\mathbf{r}, \omega)$ is the conductivity, while $\epsilon_0(\omega)$ is the complex permittivity of the homogeneous background soil medium. It is assumed that the magnetic permeability in the underground is equal to μ_0 in a vacuum. Throughout the discussion, MKS units are used.

The Fourier amplitude of the electric field vector generated in any given experiment satisfies the well known Lippmann Schwinger equation [3, 11, 16]

$$\begin{aligned} \mathbf{E}(\mathbf{r}, \omega) &= \mathbf{E}_{\text{inc}}(\mathbf{r}, \omega) + \mathbf{E}_{\text{scat}}(\mathbf{r}, \omega) \\ &= \mathbf{E}_{\text{inc}}(\mathbf{r}, \omega) - k_0^2(\omega) \int d^3 r' \bar{\mathbf{G}}(\mathbf{r} - \mathbf{r}', \omega) \cdot \mathbf{E}(\mathbf{r}', \omega) O(\mathbf{r}', \omega) \end{aligned} \quad (1)$$

where $\mathbf{E}_{\text{inc}}(\mathbf{r}, \omega)$ is the incident field for the experiment, $\mathbf{E}_{\text{scat}}(\mathbf{r}, \omega)$ is the scattered field component of the electric field vector, $k_0(\omega) = \omega\sqrt{\epsilon_0(\omega)\mu_0}$ is the complex wavenumber of the homogeneous background soil medium, and the Green's dyadic $\bar{\mathbf{G}}(\mathbf{r}, \omega)$ satisfies

$$\nabla \times \nabla \times \bar{\mathbf{G}}(\mathbf{r}, \omega) - k_0^2(\omega) \bar{\mathbf{G}}(\mathbf{r}, \omega) = \bar{\mathbf{I}}\delta(\mathbf{r}).$$

In the above equation the scattering effects of the air–soil surface interface are ignored but could, in principle, be incorporated in the Green's dyadic.

The Born approximation to the electric field is the leading term in the Liouville Neumann expansion of the solution of the above integral equation and results in the following expression for the scattered field component of the electric field vector [7, 16]:

$$\mathbf{E}_{\text{scat}}(\mathbf{r}, \omega) = -k_0^2(\omega) \int d^3 r' \bar{\mathbf{G}}(\mathbf{r} - \mathbf{r}', \omega) \cdot \mathbf{E}_{\text{inc}}(\mathbf{r}', \omega) O(\mathbf{r}', \omega). \quad (2)$$

It is evident that the Born approximation results in the replacement of the total field $\mathbf{E}(\mathbf{r}', \omega)$ within the \mathbf{r}' integration by the incident field $\mathbf{E}_{\text{inc}}(\mathbf{r}', \omega)$. Thus, the Born approximation is valid for applications where multiple scattering effects can be neglected. The Born approximation is convenient for us because it allows a linear relation between $O(\mathbf{r}, \omega)$ and $\mathbf{E}_{\text{scat}}(\mathbf{r}, \omega)$, which will generally be easier to invert than the nonlinear equation (1), where $\mathbf{E}_{\text{scat}}(\mathbf{r}, \omega)$ appears implicitly both on the left-hand side and within the \mathbf{r}' integration. It has been shown [24] that DT using the Born approximation is adequate for quantitatively reconstructing objects with roughly

$$\left(0.8 \leq \left| \frac{\epsilon(\mathbf{r}, \omega)}{\epsilon_0(\omega)} \right| \leq 1.2 \right)$$

or a 20% contrast, when the object size is on the order of a wavelength. This size-to-wavelength ratio is not unrealistic for GPR imaging, therefore the Born approximation should not be overly restrictive. Generally, as the object size decreases the allowable contrast increases. Thus, our algorithm is suitable not only for imaging larger weak scatterers, but also for imaging smaller diameter strong scatterers such as metal pipes. The Born approximation is often used, either implicitly or explicitly, for inverse problems in geophysics [7, 20, 26, 27] and it is assumed to be valid throughout this discussion.

To incorporate the characteristics of the transmitting and receiving antennas into the scattering equation, we first transform equation (2) into the spatial frequency domain, and then use Kerns' antenna scattering matrix formulation [13] to model the near-field interactions between antennas and scatterers. To convert equation (2) to the spatial frequency domain we employ the *plane wave expansion* for the Green's dyadic [3]

$$\tilde{\mathbf{G}}(\mathbf{r} - \mathbf{r}', \omega) = \frac{i}{8\pi^2} \int_{-\infty}^{\infty} \frac{d^2 \mathbf{K}}{\gamma(\mathbf{K}, \omega)} \left[\tilde{\mathbf{I}} + \frac{\mathbf{k}^+(\omega) \mathbf{k}^+(\omega)}{k_0^2(\omega)} \right] e^{i\mathbf{k}^+(\omega) \cdot (\mathbf{r} - \mathbf{r}')} \quad (3)$$

where we have assumed $z > z'$ and where $\mathbf{K} = K_x \hat{\mathbf{x}} + K_y \hat{\mathbf{y}}$ is the spatial frequency variable, $\mathbf{k}^+(\omega) = \mathbf{K} + \gamma(\mathbf{K}, \omega) \hat{\mathbf{z}}$ is the wave vector for each planewave $e^{i\mathbf{k}^+(\omega) \cdot (\mathbf{r} - \mathbf{r}')}$ in the expansion, and $\gamma(\mathbf{K}, \omega) = \pm \sqrt{k_0^2(\omega) - \mathbf{K} \cdot \mathbf{K}}$ with the sign chosen to render $\Im(\gamma) \geq 0$. For the lossless case (real k_0) the above plane wave expansion includes both *homogeneous* (propagating) plane waves ($|\mathbf{K}| \leq k_0$) having purely real wave vectors $\mathbf{k}^+(\omega)$ and *evanescent* (non-propagating) plane waves ($|\mathbf{K}| > k_0$) whose wave vectors are complex and which decay exponentially with increasing $(z - z')$. In the lossy case treated here (complex k_0) all plane waves will have complex wave vectors, however those corresponding to the higher spatial frequency range $|\mathbf{K}| > \Re[k_0]$ will decay quickest with increasing $(z - z')$. Thus, in the lossy case the term 'evanescent' is often applied to planewaves corresponding to the range $|\mathbf{K}| > \Re[k_0]$. In most formulations of DT the evanescent waves are discarded [7] since it is tacitly assumed that the measurements are conducted more than several wavelengths from the scatterers, i.e. $(z - z') \gg \frac{2\pi}{\Im[k_0]}$. In GPR imaging we should not discard evanescent components since objects of interest may be near the surface and wavelengths are relatively long (for adequate penetration), and therefore evanescent waves may contain valuable information [22].

We substitute equation (3) into equation (2) and evaluate the resulting expression at $z = 0$ to find the scattered electric field at the ground surface. We then convert this expression to the spatial frequency domain by Fourier transforming relative to the $\mathbf{X} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}}$ coordinate to obtain

$$\tilde{\mathbf{E}}_{\text{scat}}(\mathbf{K}, \omega) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} d^2 \mathbf{X} e^{-i\mathbf{K} \cdot \mathbf{X}} [\mathbf{E}_{\text{scat}}(\mathbf{r}, \omega)]_{z=0}$$

$$= \frac{-ik_0^2(\omega)}{8\pi^2\gamma(\mathbf{K}, \omega)} \int_{z' < 0} d^3r' e^{-i\mathbf{k}^+(\omega) \cdot \mathbf{r}'} \left[\bar{\mathbf{I}} + \frac{\mathbf{k}^+(\omega)\mathbf{k}^+(\omega)}{k_0^2(\omega)} \right] \cdot \mathbf{E}_{\text{inc}}(\mathbf{r}', \omega) O(\mathbf{r}', \omega). \quad (4)$$

Using Kerns' scattering matrix formulation, a transmitting antenna centred at a position $\mathbf{X}_j \equiv x_j \hat{\mathbf{x}} + y_j \hat{\mathbf{y}}$ on the ground surface and driven by matched terminal voltage $C(\omega)$ will give rise to the following plane wave expansion for the incident electric field propagating in the negative $\hat{\mathbf{z}}$ -direction: [13, 23]

$$\mathbf{E}_{\text{inc}}(\mathbf{r}, \omega) = C(\omega) \int_{-\infty}^{\infty} d^2K_0 e^{-i\mathbf{K}_0 \cdot \mathbf{X}_j} S_{10}(\mathbf{K}_0, \omega) e^{i\mathbf{k}_0^-(\omega) \cdot \mathbf{r}} \quad (5)$$

where $\mathbf{K}_0 = K_{0x} \hat{\mathbf{x}} + K_{0y} \hat{\mathbf{y}}$ and $\mathbf{k}_0^-(\omega) = \mathbf{K}_0 - \gamma(\mathbf{K}_0, \omega) \hat{\mathbf{z}}$. Similarly, the matched terminal voltage at a receiver centred at \mathbf{X}_j (a monostatic experiment) is given by [13, 23]

$$V(\omega; \mathbf{X}_j) = \int_{-\infty}^{\infty} d^2K e^{i\mathbf{K} \cdot \mathbf{X}_j} S_{01}(\mathbf{K}, \omega) \cdot \tilde{\mathbf{E}}_{\text{scat}}(\mathbf{K}, \omega). \quad (6)$$

In equations (5) and (6), S_{10} and S_{01} are off-diagonal scattering matrix coefficients for the antennas, with each vector component corresponding to a Cartesian electric field polarization. It should be emphasized that equations (5) and (6) are valid in the nearfield of the antennas, thus, we need not make the point source approximation typical to other inversion algorithms of diffraction tomography. If the transmitting and receiving antennas are reciprocal, as would be the case for a typical monostatic radar system, then [13, 23]

$$Y_0(\omega) S_{01}(\mathbf{K}, \omega) = \frac{\gamma(\mathbf{K}, \omega)}{\omega \mu_0} S_{10}(-\mathbf{K}, \omega) \quad (7)$$

where Y_0 is the antenna terminal admittance.

Combining equations (4)–(7), the measured voltage $V(\omega)$ for a single monostatic experiment with the transmitter/receiver centred at \mathbf{X}_j is:

$$\begin{aligned} V(\omega; \mathbf{X}_j) &= P(\omega) \int_{-\infty}^{\infty} d^2K \int_{-\infty}^{\infty} d^2K_0 e^{-i(\mathbf{K}_0 - \mathbf{K}) \cdot \mathbf{X}_j} B(\mathbf{K}, \mathbf{K}_0; \omega) \\ &\times \int_{-\infty}^0 dz' e^{-i[\gamma(\mathbf{K}, \omega) + \gamma(\mathbf{K}_0, \omega)]z'} \int_{-\infty}^{\infty} d^2X' e^{-i(\mathbf{K} - \mathbf{K}_0) \cdot \mathbf{X}'} O(\mathbf{r}') \end{aligned} \quad (8)$$

where

$$B(\mathbf{K}, \mathbf{K}_0; \omega) \equiv S_{10}(-\mathbf{K}, \omega) \cdot \left[\bar{\mathbf{I}} + \frac{\mathbf{k}^+(\omega)\mathbf{k}^+(\omega)}{k_0^2(\omega)} \right] \cdot S_{10}(\mathbf{K}_0, \omega)$$

and

$$P(\omega) = \frac{-iC(\omega)k_0^2(\omega)}{8\pi^2\omega Y_0(\omega)\mu_0}.$$

If the receiver in each experiment is spatially offset from the transmitter by a constant amount \mathbf{X}_0 , then we must multiply $B(\mathbf{K}, \mathbf{K}_0; \omega)$ given above by the factor $e^{i\mathbf{K} \cdot \mathbf{X}_0}$.

Notice that in equation (8) the object function is expressed as a frequency independent (dispersionless) quantity $O(\mathbf{r}, \omega) = O(\mathbf{r})$. This assumption, although not necessary, allows us to couple the measured data at each frequency, thus, incorporating more information into the mathematical inversions and leading to ‘better’ solutions for the object function. There are alternatives to this approach that will work using the methodology of this paper, including: (i) solve for the frequency dependent object function $O(\mathbf{r}, \omega)$ independently at each single radar frequency; (ii) treat the object function as the product of a known frequency dependent factor $\Theta(\omega)$ and an unknown frequency independent factor $O(\mathbf{r})$,

then incorporate the known function $\Theta(\omega)$ into $P(\omega)$ given above, subsequently solving for the unknown $O(\mathbf{r})$ in the standard manner discussed below. In any case, we note that the assumption of a dispersionless object function in GPR is common (e.g. [27]) and models many important applications over relatively large frequency bands.

A typical GPR survey consists of multiple experiments, each corresponding to a different position \mathbf{X}_j for the transmitting/receiving antennas, thus we have a coupled set of equations, each in the form of equation (8). The objective of the present work is to estimate the object function $O(\mathbf{r})$ by inverting the coupled equations using the regularized pseudoinverse formula as described below.

3. Data inversion for three-dimensional surveys

In this section we construct a mathematical operator formalism for the coupled set of equations (8), where each equation corresponds to a different transmitter/receiver position. This mathematical framework allows us to use well known linear inversion methods to analytically solve for the object function in terms of the measured data.

We consider a three-dimensional survey where a series of monostatic experiments are performed over the ground surface at evenly spaced locations corresponding to points on a two-dimensional grid. For each experiment we store data from N different excitation frequencies ω_n . If the grid spacing is small enough to satisfy the Nyquist sampling criterion for the voltage measurements, we can treat the transmitter/receiver position as a continuous variable $\mathbf{X} = (x, y)$ (below we discuss the spatial sampling requirements in further detail). Thus, from equation (8) the measured voltage as a function of frequency and transmitter/receiver position is

$$V(\omega_n; \mathbf{X}) = P(\omega_n) \int_{-\infty}^{\infty} d^2 K \int_{-\infty}^{\infty} d^2 K_0 e^{-i(\mathbf{K}_0 - \mathbf{K}) \cdot \mathbf{X}} B(\mathbf{K}, \mathbf{K}_0; \omega_n) \\ \times \int_{-\infty}^0 dz' e^{-i[\gamma(\mathbf{K}, \omega_n) + \gamma(\mathbf{K}_0, \omega_n)]z'} \int_{-\infty}^{\infty} d^2 X' e^{-i(\mathbf{K} - \mathbf{K}_0) \cdot \mathbf{X}'} O(\mathbf{r}').$$

We make the change of variable $\bar{\mathbf{K}} = \mathbf{K} - \mathbf{K}_0$, and drop the bar notation on $\bar{\mathbf{K}}$ to yield

$$V(\omega_n; \mathbf{X}) = \int_{-\infty}^{\infty} d^2 K e^{i\mathbf{K} \cdot \mathbf{X}} P(\omega_n) \int_{-\infty}^{\infty} d^2 K_0 B(\mathbf{K} + \mathbf{K}_0, \mathbf{K}_0; \omega_n) \\ \times \int_{-\infty}^0 dz' e^{-i[\gamma(\mathbf{K} + \mathbf{K}_0, \omega_n) + \gamma(\mathbf{K}_0, \omega_n)]z'} \int_{-\infty}^{\infty} d^2 X' e^{-i\mathbf{K} \cdot \mathbf{X}'} O(\mathbf{r}'). \quad (9)$$

We now spatially Fourier transform equation (9) with respect to the \mathbf{X} variable to obtain

$$\tilde{V}(\omega_n; \mathbf{K}) \equiv \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} d^2 X e^{-i\mathbf{K} \cdot \mathbf{X}} V(\omega_n; \mathbf{X}) \\ = P(\omega_n) \int_{-\infty}^{\infty} d^2 K_0 B(\mathbf{K} + \mathbf{K}_0, \mathbf{K}_0; \omega_n) \\ \times \int_{-\infty}^0 dz' e^{-i[\gamma(\mathbf{K} + \mathbf{K}_0, \omega_n) + \gamma(\mathbf{K}_0, \omega_n)]z'} \int_{-\infty}^{\infty} d^2 X' e^{-i\mathbf{K} \cdot \mathbf{X}'} O(\mathbf{r}'). \quad (10)$$

It is important to define Hilbert spaces for the various quantities entering into equation (10) and to express this equation in compact operator formalism. Thus, we introduce the vector space U of object functions $O(\mathbf{r})$ and the vector space Y of transformed measured voltages $\tilde{V}(\omega_n; \mathbf{K})$. We will employ the standard L^2 -inner products in both spaces and will assume the elements of each space to have finite L^2 -norms, i.e.

- $O(\mathbf{r}) \in U$, where U is the space of square integrable functions on $-\infty < (x, y) < \infty$, $-\infty < z < 0$;
- $\tilde{V}(\omega_n; \mathbf{K}) \in Y$, where Y is the direct product space of square integrable functions on $-\infty < (K_x, K_y) < \infty$ with the finite-dimensional vector space Y_0 of functions of the discrete variable ω_n .

With these definitions we can write equation (10) in the compact form

$$\tilde{V}(\omega_n; \mathbf{K}) = H O(\omega_n; \mathbf{K}) \quad (11)$$

where H is a linear operator which maps U into Y .

Our goal is to find the object function $O(\mathbf{r}')$ which satisfies the set of equations (11). Since we use only a finite number of excitation frequencies, the set of equations (11) are underdetermined so that there are an infinite number of object functions which will produce the data. The minimum L^2 -norm solution is approximately found by applying the regularized pseudoinverse operator [1, 2, 12, 17] to the data, i.e.

$$\hat{O}_\beta(\mathbf{r}) = H^\dagger [H H^\dagger + \beta I]^{-1} \tilde{V}(\mathbf{r}) \quad (12)$$

where H^\dagger is the Hermitian adjoint of H . In the limit as β goes to zero, $\hat{O}_\beta(\mathbf{r})$ is equal to the minimum L^2 -norm solution to equation (11)[†]. The number β is known as the Tikhonov–Phillips regularization parameter [21, 17], and is necessary to stabilize the inversion against noise from measurements or numerical truncations. The regularization parameter is also necessary in the event that $(H H^\dagger)^{-1}$ does not exist. Generally there is a tradeoff between selecting β small enough such that $\hat{O}_\beta(\mathbf{r})$ approximately satisfies the data yet large enough that the inversion is stable, and typically β is selected by trial and error. We show below that equation (12) is in the form of a filtered backpropagation operation, where $[H H^\dagger + \beta I]^{-1}$ is the filtering operator and the H^\dagger is a coherent sum over frequencies of the filtered, backpropagated, data.

Before proceeding with the inversion algorithm, we should make several comments about our solution strategy. First, in GPR imaging there will be a significant amount of noise associated both with the receiver electronics and scattering from ground clutter, and this noise may profoundly affect computed solutions. One can modify equation (11) to include this additive noise component n explicitly as follows:

$$\tilde{V} = H O + n.$$

The minimum variance (Wiener filtering) solution [2, 12] to this equation is

$$\hat{O}_w = R_V H^\dagger [H R_V H^\dagger + R_n]^{-1} \tilde{V} \quad (13)$$

where \tilde{V} and n are assumed to be zero-mean random processes having covariances R_V and R_n , respectively. If \tilde{V} and n are Gaussian white processes with variances σ_V^2 and σ_n^2 , respectively (the Gaussian white assumption may be sensible if no other information is available), then equation (13) reduces to [2, 12]

$$\hat{O}_w = H^\dagger [H H^\dagger + \sigma_n^2 / \sigma_V^2]^{-1} \tilde{V}. \quad (14)$$

Note that equations (12) and (14) are identical for $\beta = \sigma_n^2 / \sigma_V^2$, and thus we expect that the optimum value for β in Tikhonov–Phillips regularization is roughly proportional to

[†] In practice one does not have access to an infinite number of monostatic measurements, rather the received voltage is sampled at discrete spatial intervals over a finite area of the xy plane. Representing this sampled voltage as $\tilde{V} = R F^{-1} \tilde{V} = R F^{-1} H O$, where R is the sampling operator and F^{-1} is the inverse Fourier transform, we can show that $\hat{O}_{\beta=0}$ from equation (12) is also the minimum L^2 norm solution to $\tilde{V} = R F^{-1} H O$ if the linear operator $(R F^{-1})$ can be inverted. In practice, $(R F^{-1})$ can be inverted if the Nyquist sampling criterion is satisfied and the spatial extent of the sampling area is large relative to the support region dimensions (depth and horizontal) of the object function.

the noise variance. Note that there are other robust alternatives to Tikhonov–Phillips regularization and Wiener filtering, for example the reduced rank pseudoinverse operator [1, 10]. We choose the method of Tikhonov–Phillips regularization because it allows us to derive analytical formulae.

Let us now analytically expand the operator notation in equation (12) to obtain explicitly the pseudoinverse solution. Given the vector space definitions, the Hermitian adjoint H^\dagger of H maps the space Y onto the space U so that

$$\langle H O_\beta, \tilde{V} \rangle_Y = \langle O_\beta, H^\dagger \tilde{V} \rangle_U.$$

From this inner product relation the action of H^\dagger is found to be

$$\begin{aligned} H^\dagger \hat{\tilde{V}}(\mathbf{r}') \equiv \hat{O}_\beta(\mathbf{r}') &= \sum_{n=1}^N P^*(\omega_n) \int_{-\infty}^{\infty} d^2 K' \int_{-\infty}^{\infty} d^2 K'_0 B^*(\mathbf{K}' + \mathbf{K}'_0, \mathbf{K}'_0; \omega_n) \\ &\times e^{i[\gamma^*(\mathbf{K}' + \mathbf{K}'_0, \omega_n) + \gamma^*(\mathbf{K}'_0, \omega_n)]z'} e^{i\mathbf{K}' \cdot \mathbf{X}'} \hat{\tilde{V}}(\omega_n; \mathbf{K}') \end{aligned} \quad (15)$$

where the $(*)$ denotes the complex conjugate, $\mathbf{r}' = (\mathbf{X}', z')$, and where we have added a ‘hat’ to \tilde{V} for later notational purposes. In typical filtered backpropagation algorithms [4, 6, 7] the backpropagation operation is the adjoint of the forward propagation. Similarly, each term in the summation in equation (15) is the adjoint of the forward scattering operator at a single excitation frequency. Because of this analogy, we refer to each term in the summation as a backpropagation, although this is not true in the usual sense.

Let us now consider the operation $\hat{\tilde{V}} = [HH^\dagger + \beta I]^{-1} \tilde{V}$, which corresponds to the filtering operation in standard DT filtered backpropagation algorithms, where $\hat{\tilde{V}}$ is the filtered data and \tilde{V} is the raw, unfiltered, data. From equations (10) and (15), the inverse of the filtering operation is

$$\begin{aligned} \tilde{V}(\omega_m; \mathbf{K}) &= [HH^\dagger + \beta I] \hat{\tilde{V}}(\omega_m; \mathbf{K}) \\ &= \sum_{n=1}^N P(\omega_m) P^*(\omega_n) \int_{-\infty}^{\infty} d^2 K' \hat{\tilde{V}}(\omega_n; \mathbf{K}') \\ &\times \int_{-\infty}^{\infty} d^2 K_0 \int_{-\infty}^{\infty} d^2 K'_0 B(\mathbf{K} + \mathbf{K}_0, \mathbf{K}_0; \omega_m) B^*(\mathbf{K}' + \mathbf{K}'_0, \mathbf{K}'_0; \omega_n) \\ &\times \int_{-\infty}^0 dz' e^{-i[\gamma(\mathbf{K} + \mathbf{K}_0, \omega_m) + \gamma(\mathbf{K}_0, \omega_m) - \gamma^*(\mathbf{K}' + \mathbf{K}'_0, \omega_n) - \gamma^*(\mathbf{K}'_0, \omega_n)]z'} \\ &\times \int_{-\infty}^{\infty} d^2 X' e^{-i(\mathbf{K} - \mathbf{K}') \cdot \mathbf{X}'} \\ &+ \beta \hat{\tilde{V}}(\omega_m; \mathbf{K}). \end{aligned}$$

The integral over \mathbf{X}' reduces to a Dirac delta function, and the integral over z' can be solved since $\Im(\gamma) > 0$, therefore

$$\tilde{V}(\omega_m; \mathbf{K}) = \sum_{n=1}^N R_{mn}(\mathbf{K}) \hat{\tilde{V}}(\omega_n; \mathbf{K}) + \beta \hat{\tilde{V}}(\omega_m; \mathbf{K})$$

where

$$\begin{aligned} R_{mn}(\mathbf{K}) &= \int_{-\infty}^{\infty} d^2 K_0 \int_{-\infty}^{\infty} d^2 K'_0 \\ &\times \frac{4i\pi^2 P(\omega_m) P^*(\omega_n) B(\mathbf{K} + \mathbf{K}_0, \mathbf{K}_0; \omega_m) B^*(\mathbf{K} + \mathbf{K}'_0, \mathbf{K}'_0; \omega_n)}{[\gamma(\mathbf{K} + \mathbf{K}_0, \omega_m) + \gamma(\mathbf{K}_0, \omega_m) - \gamma^*(\mathbf{K} + \mathbf{K}'_0, \omega_n) - \gamma^*(\mathbf{K}'_0, \omega_n)]} \end{aligned} \quad (16)$$

For each value of \mathbf{K} this set of equations can be written in the matrix notation

$$\begin{bmatrix} \tilde{V}(\omega_1; \mathbf{K}) \\ \tilde{V}(\omega_2; \mathbf{K}) \\ \vdots \\ \tilde{V}(\omega_N; \mathbf{K}) \end{bmatrix} = \begin{bmatrix} R_{11}(\mathbf{K}) + \beta & R_{12}(\mathbf{K}) & \cdots & R_{1N}(\mathbf{K}) \\ R_{21}(\mathbf{K}) & R_{22}(\mathbf{K}) + \beta & \cdots & R_{2N}(\mathbf{K}) \\ \vdots & \vdots & \ddots & \vdots \\ R_{N1}(\mathbf{K}) & R_{N2}(\mathbf{K}) & \cdots & R_{NN}(\mathbf{K}) + \beta \end{bmatrix} \begin{bmatrix} \hat{\tilde{V}}(\omega_1; \mathbf{K}) \\ \hat{\tilde{V}}(\omega_2; \mathbf{K}) \\ \vdots \\ \hat{\tilde{V}}(\omega_N; \mathbf{K}) \end{bmatrix} \quad (17)$$

which can be inverted by standard methods of linear algebra to yield the filtered data $\hat{\tilde{V}}$ in terms of the raw, unfiltered data \tilde{V} . The regularized pseudoinverse solution $\hat{O}_\beta(\mathbf{r})$ can then be found by operating on the filtered data by the backpropagation operator H^\dagger as given in equation (15).

As an example of the object function reconstruction, let us consider the case where there are two radar frequencies ω_1 and ω_2 . Inverting the 2×2 version of equation (17), we have the following solutions for the filtered data:

$$\hat{\tilde{V}}(\omega_1; \mathbf{K}) = \frac{(R_{22}(\mathbf{K}) + \beta)}{\Delta(\mathbf{K})} \tilde{V}(\omega_1; \mathbf{K}) - \frac{R_{12}(\mathbf{K})}{\Delta(\mathbf{K})} \tilde{V}(\omega_2; \mathbf{K})$$

and

$$\hat{\tilde{V}}(\omega_2; \mathbf{K}) = -\frac{R_{21}(\mathbf{K})}{\Delta(\mathbf{K})} \tilde{V}(\omega_1; \mathbf{K}) + \frac{(R_{11}(\mathbf{K}) + \beta)}{\Delta(\mathbf{K})} \tilde{V}(\omega_2; \mathbf{K})$$

where

$$\Delta(\mathbf{K}) = [R_{11}(\mathbf{K}) + \beta][R_{22}(\mathbf{K}) + \beta] - R_{21}(\mathbf{K})R_{12}(\mathbf{K}).$$

According to equation (12), we then reconstruct the object function $\hat{O}_\beta(\mathbf{r})$ by backpropagating the filtered data using equation (15) with $N = 2$.

4. Data inversion for 2.5-dimensional surveys

The 2.5-dimensional case corresponds to assuming a two-dimensional object function $O(x, z)$, invariant in the \hat{y} -direction, but fully three-dimensional incident and scattered wavefields. We can treat the 2.5-dimensional case in a manner completely analogous to the three-dimensional analysis of section 3. Whereas for the three-dimensional case we conducted a series of monostatic experiments over a two-dimensional grid on the xy plane, here we conduct experiments at intervals along a one-dimensional line corresponding to the \hat{x} -axis. From the scattered field measurements we wish to reconstruct the object function $O(x, z)$. The receiver voltage for each experiment is related to the object function by equation (8). We emphasize that the antennas are three dimensional, such that the incident and scattered electric fields vary in three dimensions and have three Cartesian polarizations.

For each experiment we store data from N different excitation frequencies ω_n . We assume the interval spacing along the \hat{x} -axis is small enough to satisfy the Nyquist sampling criterion for the voltage measurements, thus we can treat the measured voltage $V(\omega_n; x)$ as a function of the continuous variable x (the discussion of the sampling requirements from the footnote in section 3 is valid here also). The data are defined as the spatial Fourier transform relative to the x coordinate of the measured voltage. By methods similar to the three-dimensional analysis, these data are found to be

$$\tilde{V}(\omega_n; K_x) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} dx e^{-iK_x x} V(\omega_n; x)$$

$$\begin{aligned}
&= 2\pi P(\omega_n) \int_{-\infty}^{\infty} d^2 K_0 B(K_x \hat{\mathbf{x}} + \mathbf{K}_0, \mathbf{K}_0; \omega_n) \int_{-\infty}^0 dz' e^{-i[\gamma(K_x \hat{\mathbf{x}} + \mathbf{K}_0, \omega_n) + \gamma(\mathbf{K}_0, \omega_n)]z'} \\
&\quad \times \int_{-\infty}^{\infty} dx' e^{-iK_x x'} O(x', z') = A O(\omega_n; K_x).
\end{aligned} \tag{18}$$

Here we use the Hilbert space definitions;

- $O(x, z) \in W$, where W is the space of square integrable functions on $-\infty < x < \infty$, $-\infty < z < 0$;

- $\tilde{V}(\omega_n; K_x) \in Z$, where Z is the direct product space of square integrable functions on $-\infty < K_x < \infty$ with the finite-dimensional vector space Y_0 of functions of the discrete variable ω_n ;

- A is a linear operator which maps W into Z .

We seek the regularized pseudoinverse solution for the object function

$$\begin{aligned}
\hat{O}_\beta(x, z) &= A^\dagger [AA^\dagger + \beta I]^{-1} \tilde{V}(x, z) \\
&= A^\dagger \hat{\tilde{V}}(x, z).
\end{aligned} \tag{19}$$

Given the vector space definitions, the adjoint of A is specified by the action

$$\begin{aligned}
A^\dagger \hat{\tilde{V}}(x', z') &\equiv \hat{O}_\beta(x', z') = \sum_{n=1}^N 2\pi P^*(\omega_n) \int_{-\infty}^{\infty} dK'_x \int_{-\infty}^0 dz' d^2 K'_0 B^*(K'_x \hat{\mathbf{x}} + \mathbf{K}'_0, \mathbf{K}'_0; \omega_n) \\
&\quad \times e^{i[\gamma^*(K'_x \hat{\mathbf{x}} + \mathbf{K}'_0, \omega_n) + \gamma^*(\mathbf{K}'_0, \omega_n)]z'} e^{iK'_x x'} \hat{\tilde{V}}(\omega_n; K'_x).
\end{aligned} \tag{20}$$

The filtered data $\hat{\tilde{V}}$ and the raw, unfiltered, data \tilde{V} are related by the filtering relation $\hat{\tilde{V}} = [AA^\dagger + \beta I]^{-1} \tilde{V}$, which is expanded by combining equations (18) and (20) and evaluating the x' and z' integrals as in section 3. In this manner we obtain the following relation between the filtered and unfiltered data:

$$\tilde{V}(\omega_m; K_x) = \sum_{n=1}^N Q_{mn}(K_x) \hat{\tilde{V}}(\omega_n; K_x) + \beta \hat{\tilde{V}}(\omega_m; K_x) \tag{21}$$

where

$$\begin{aligned}
Q_{mn}(K_x) &= \int_{-\infty}^{\infty} d^2 K_0 \int_{-\infty}^0 dz' d^2 K'_0 \\
&\quad \times \frac{8i\pi^3 P(\omega_m) P^*(\omega_n) B(K_x \hat{\mathbf{x}} + \mathbf{K}_0, \mathbf{K}_0; \omega_m) B^*(K_x \hat{\mathbf{x}} + \mathbf{K}'_0, \mathbf{K}'_0; \omega_n)}{[\gamma(K_x \hat{\mathbf{x}} + \mathbf{K}_0, \omega_m) + \gamma(\mathbf{K}_0, \omega_m) - \gamma^*(K_x \hat{\mathbf{x}} + \mathbf{K}'_0, \omega_n) - \gamma^*(\mathbf{K}'_0, \omega_n)]}.
\end{aligned}$$

For each value of K_x , equation (21) can be written in the matrix notation

$$\begin{aligned}
\begin{Bmatrix} \tilde{V}(\omega_1; K_x) \\ \tilde{V}(\omega_2; K_x) \\ \vdots \\ \tilde{V}(\omega_N; K_x) \end{Bmatrix} &= \begin{Bmatrix} Q_{11}(K_x) + \beta & Q_{12}(K_x) & \cdots & Q_{1N}(K_x) \\ Q_{21}(K_x) & Q_{22}(K_x) + \beta & \cdots & Q_{2N}(K_x) \\ \vdots & \vdots & \ddots & \vdots \\ Q_{N1}(K_x) & Q_{N2}(K_x) & \cdots & Q_{NN}(K_x) + \beta \end{Bmatrix} \begin{Bmatrix} \hat{\tilde{V}}(\omega_1; K_x) \\ \hat{\tilde{V}}(\omega_2; K_x) \\ \vdots \\ \hat{\tilde{V}}(\omega_N; K_x) \end{Bmatrix}
\end{aligned} \tag{22}$$

which can be inverted by standard methods of linear algebra to yield the filtered data $\hat{\tilde{V}}$ in terms of the raw, unfiltered data \tilde{V} . The regularized pseudoinverse solution $\hat{O}_\beta(\mathbf{r})$ can then be found by operating on the filtered data by the operator A^\dagger as given in equation (20).

5. Computed results from synthetic data

We present here the results of computer simulations designed to test the reconstruction algorithm. For simplicity, we treat the two-dimensional problem where the transmitted electric field has only a $\hat{\mathbf{y}}$ polarization, and the antennas and scatterers do not vary in the $\hat{\mathbf{y}}$ -direction[†]. Thus, $\mathbf{K} = K\hat{\mathbf{x}}$, $\mathbf{K}_0 = K_0\hat{\mathbf{x}}$, $\mathbf{X} = x\hat{\mathbf{x}}$, and $S_{10}(\mathbf{K}, \omega) = S_{10}(K, \omega)\hat{\mathbf{y}}$. Using this two-dimensional model, we can simplify all results in sections 2 and 3 by converting vector quantities to scalars, reducing by one the-dimensions of all integrations over \mathbf{K} , \mathbf{K}_0 , and \mathbf{r} , and changing some constant factors.

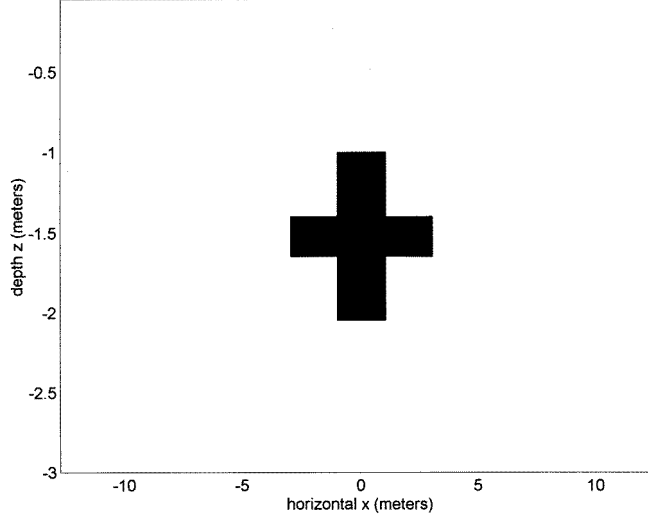


Figure 2. Original object function for frequency domain simulations. $O(x, z) = 0.2$ within the cross region at all frequencies.

More explicitly, the forward model for the two-dimensional case can be shown to be

$$\begin{aligned} \tilde{V}(\omega_n; K_x) &= 2\pi P(\omega_n) \int_{-\infty}^{\infty} dK_0 B(K_x + K_0, K_0; \omega_n) \int_{-\infty}^0 dz' e^{-i[\gamma(K_x + K_0, \omega_n) + \gamma(K_0, \omega_n)]z'} \\ &\quad \times \int_{-\infty}^{\infty} dx' e^{-iK_x x'} O(x', z'). \end{aligned} \quad (23)$$

As we discussed in the previous sections, the object function is reconstructed by a filtered backpropagation algorithm. The filtering operation consists of inverting the matrix equation (22) for each value of K_x where, for the two-dimensional case, the matrix elements can be shown to be

$$\begin{aligned} Q_{mn}(K_x) &= \int_{-\infty}^{\infty} dK_0 \int_{-\infty}^{\infty} dK'_0 \\ &\quad \times \frac{8i\pi^3 P(\omega_m) P^*(\omega_n) B(K_x + K_0, K_0; \omega_m) B^*(K_x + K'_0, K'_0; \omega_n)}{[\gamma(K_x + K_0, \omega_m) + \gamma(K_0, \omega_m) - \gamma^*(K_x + K'_0, \omega_n) - \gamma^*(K'_0, \omega_n)]}. \end{aligned} \quad (24)$$

[†] In practice, this model will be valid for one-dimensional linear antennas, and scatterers that vary slowly in the $\hat{\mathbf{y}}$ -direction.

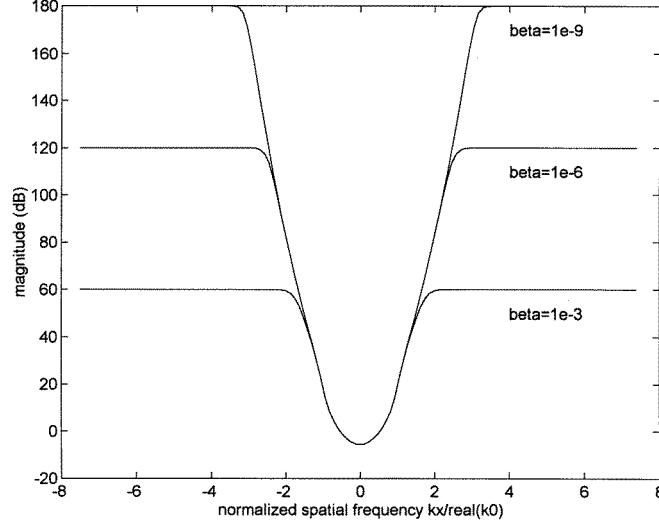


Figure 3. Magnitude of the filter versus spatial frequency K_x and $\beta = 10^{-3}, 10^{-6}, 10^{-9}$ for a single frequency simulation with $\lambda = 1.5$ m.

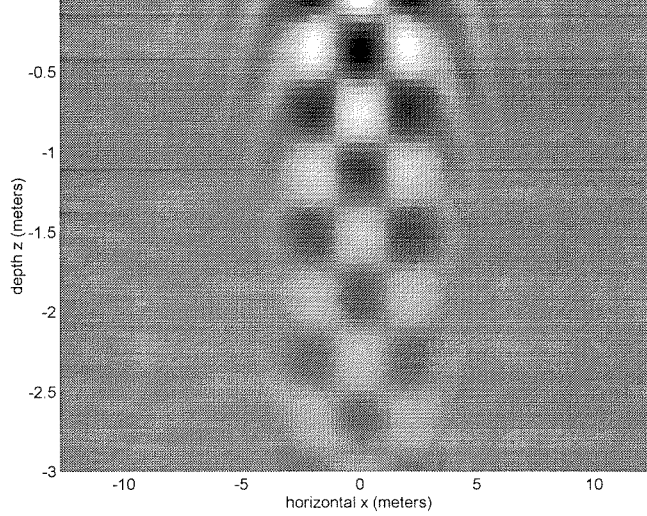


Figure 4. Pseudoinverse reconstruction based on a single excitation frequency corresponding to $\lambda = 1.5$ m. The penetration depth in the soil is 3 m.

Once the data is filtered, the object function is reconstructed by the following two-dimensional backpropagation, analogous to equations (15) and (20):

$$\hat{O}_\beta(x', z') = \sum_{n=1}^N 2\pi P^*(\omega_n) \int_{-\infty}^{\infty} dK'_x \int_{-\infty}^{\infty} dK'_0 B^*(K'_x + K'_0, K'_0; \omega_n) \times e^{i[\gamma^*(K'_x + K'_0, \omega_n) + \gamma^*(K'_0, \omega_n)]z'} e^{iK'_x x'} \hat{V}(\omega_n; K'_x). \quad (25)$$

Figure 2 shows a purely real two-dimensional object function $O(x, z)$ which is embedded in a soil background having a penetration depth of 3 m at all frequencies. The

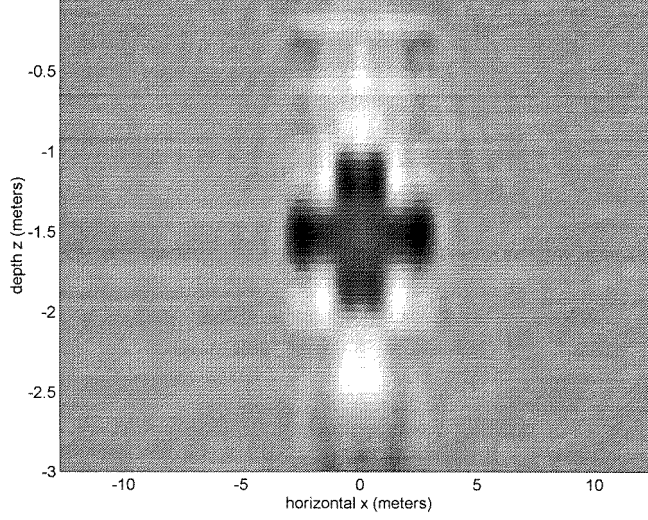


Figure 5. Pseudoinverse reconstruction based on nine different excitation frequencies corresponding to $0.375 \text{ m} < \lambda < 3.375 \text{ m}$. The penetration depth in the soil is 3 m for all frequencies.

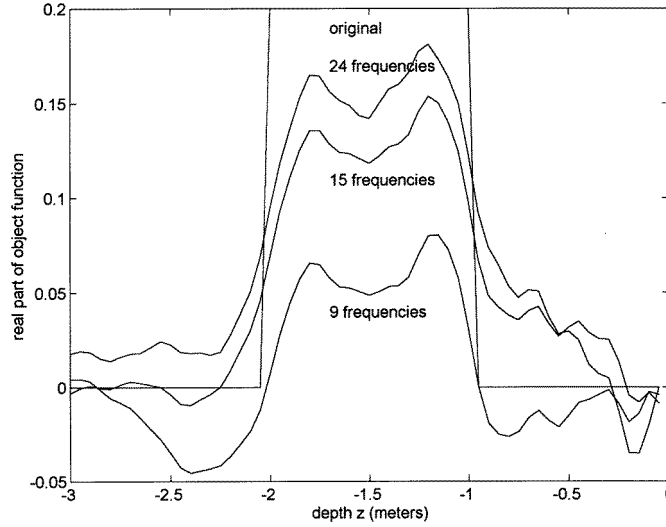


Figure 6. Cross sections at $x = 0$ for pseudoinverse reconstructions computed from 9, 15, and 24 frequencies corresponding to $0.375 \text{ m} < \lambda < 3.375 \text{ m}$.

value of the object function is $O(x, z) = 0.2$ within the cross region at all frequencies, therefore the dielectric contrast is $\frac{\epsilon(r, \omega)}{\epsilon_0(\omega)} = 0.8$. The antennas are characterized as being fairly directional, with (two-dimensional scalar) transmitting coefficients given by[†]

$$S_{10}(K, \omega) = \frac{e^{-a(\omega)K^2}}{\gamma(K, \omega)} \quad (26)$$

[†] For comparison, it can be shown from the planewave expansion of the Green's dyadic in equation (3) that a thin linear antenna, exciting an omni-directional field, will have the transmitting coefficient $S_{10}(K, \omega) = 1/\gamma(K, \omega)$. The exponential numerator in equation (26) makes the antenna directional.

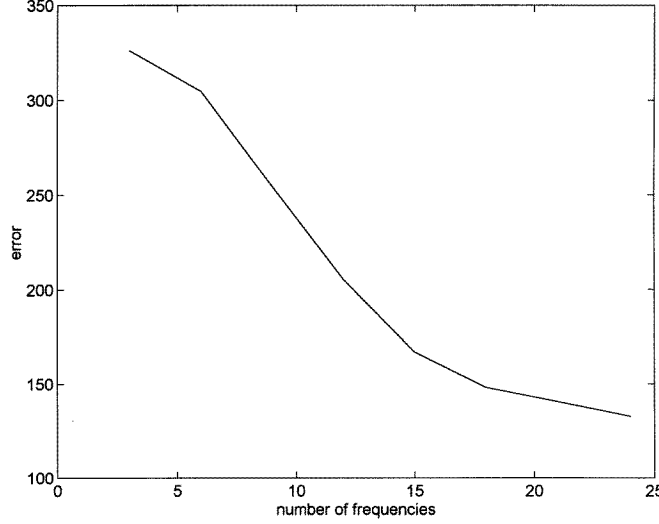


Figure 7. Error = $\int d^2r |O(r) - \hat{O}_\beta(r)|^2$ versus number of frequencies used in the reconstruction. Frequencies correspond to the range $0.375 \text{ m} < \lambda < 3.375 \text{ m}$.

where $a(\omega)$ is a constant chosen such that $e^{-a(\omega)[\Re(k_0)]^2} \approx \frac{1}{10}$. For each simulation, the synthetic data are obtained by numerically applying equation (23) to the object of figure 2.

We initially compute a reconstruction based on a single radar frequency corresponding to the wavelength $\lambda = 1.5 \text{ m}$. To accomplish this task we first filter the data according to the equation

$$\hat{V}(\omega_1; K_x) = \frac{\tilde{V}(\omega_1; K_x)}{Q_{11}(K_x) + \beta} \quad (27)$$

(with $Q_{11}(K_x)$ given in equation (24)) which is the inverse of the 1×1 version of equations (22). The minimum norm reconstruction is then obtained by backpropagation as defined by equation (25) with $N = 1$. The magnitude of the filter $1/(Q_{11}(K_x) + \beta)$ is plotted versus the spatial frequency K_x in figure 3 for several different values of the regularization parameter β . It is evident that this filter tends to amplify the high spatial frequency components of the data relative to the low frequency components. Without regularization, high frequency components from additive noise will be amplified without bound, thus leading to unstable inversions. By using a non-zero β this amplification is limited, thus stabilizing the inversion. The real part of the minimum norm reconstruction is shown in figure 4, where $\beta = 1e - 9$ was chosen by trial and error. The criterion used in the selection of β was that both the original object of figure 2 and the reconstruction of figure 4 should yield nearly identical data since they are both solutions to equation (23). Note from figure 4 that the outline of the original object function is evident, however, there is considerable periodic ambiguity in the image.

Next we compute a reconstruction based on nine frequencies corresponding to $0.375 \text{ m} < \lambda < 3.375 \text{ m}$ using the $N = 9$ versions of equations (25) and (22). The regularization parameter is set to $\beta = 1e - 9$, as before. It can be noted from the real part of the reconstruction, shown in figure 5, that by using multiple frequencies the periodic ambiguity observed in figure 4 is diminished, and we obtain a fairly good representation of the original scattering object. In figure 6 we plot the cross section of figure 5 along the line $x = 0$, as well as cross sections from reconstructions using 15 and 21 frequencies

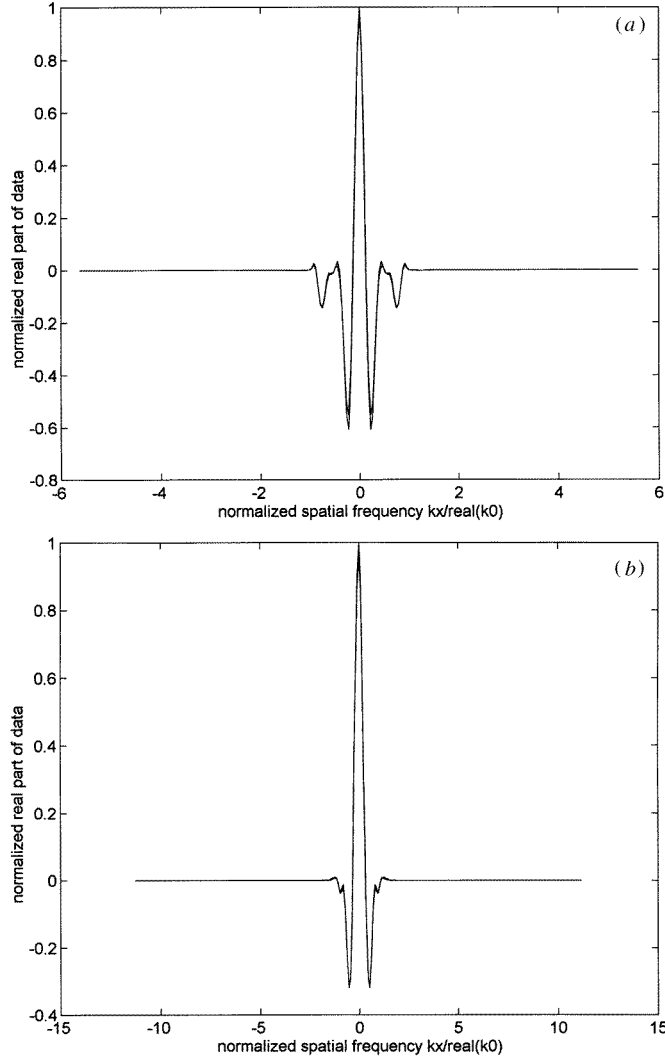


Figure 8. (a) Real part of the data $\tilde{V}(\omega_n; K_x)$ corresponding to $\lambda = 1.125$ m as computed from equation (23) on both the original object of figure 2 and the reconstruction of figure 5 computed with nine frequencies. (b) The same plot as (a), for data from the frequency corresponding to $\lambda = 2.25$ m.

corresponding to the range $0.375 \text{ m} < \lambda < 3.375 \text{ m}$. It is evident that increasing the number of frequencies leads to a significant quantitative improvement. This concept is also supported by figure 7, which shows decreasing error $= \int d^2r |O(\mathbf{r}) - \hat{O}_\beta(\mathbf{r})|^2$ with increasing numbers of frequencies corresponding to the range $0.375 \text{ m} < \lambda < 3.375 \text{ m}$.

As a test that the minimum norm reconstruction is consistent with the data, and therefore satisfies equation (23), we implement the forward scattering relation of equation (23) on both the original object function of figure 2 and the reconstructed object function of figure 5 computed from nine frequencies. The real part of the data $\tilde{V}(\omega_n; K_x)$ is plotted in figures 8(a) and (b) for both the original and reconstructed object functions, at two of the nine frequencies corresponding to $\lambda = 1.125$ m and $\lambda = 2.25$ m (to avoid redundancy, data

from all the frequencies are not plotted). Based on the excellent match obtained between data from the original and reconstructed objects at all frequencies (the plots are almost completely overlapping in the scale shown), we conclude that the reconstruction in figure 5 is consistent with the data, as expected.

6. Discussion

In this paper we derived a direct, non-iterative, inversion formula for multi-monostatic GPR data which was tested successfully on synthetic radar data. To solve this problem we cast the forward scattering relation in a mathematical operator framework, allowing computationally efficient, analytical expressions to be found for both the adjoint and pseudoinverse operators. This type of approach has been used with similar success in other tomographic imaging applications [4], and we anticipate future uses as well.

We mentioned above that significant noise will be inevitable in realistic GPR imaging applications. In preliminary studies, we have verified that our algorithm is robust in the presence of both additive noise at receivers, and noise due to scattering from ground clutter. In these noisy simulations, the results were quite sensitive to the value of the regularization parameter β , as expected. In a future publication we plan to present the results of these noisy simulations, as well as results using experimental GPR data.

Acknowledgments

Research supported by DOE grant DE-FG02-93ER14392.

References

- [1] Bertero M, DeMol C and Viano G A 1980 The stability of inverse problems *Inverse Scattering Problems in Optics: Topics in Current Physics* vol 20, ed H P Baltes (Berlin: Springer) pp 161–214
- [2] Bertero M 1988 Linear inverse and ill-posed problems *Adv. Electron Electron. Phys.* vol 75 ed P W Hawkes (New York: Academic) pp 1–121
- [3] Chew W C 1990 *Waves and Fields in Inhomogeneous Media* (New York: IEEE Press)
- [4] Deming R W and Devaney A J 1995 A filtered backpropagation algorithm for GPR *J. Environmental Eng. Geophys.* **0** 113–23
- [5] Devaney A J and Sherman G C 1973 Plane-wave representations for scalar wave fields *SIAM Rev.* **15** 765–86
- [6] Devaney A J 1983 Inverse source and scattering problems in ultrasonics *IEEE Trans. Sonics Ultrasonics* **SU-30** 355–64
- [7] Devaney A J 1984 Geophysical diffraction tomography *IEEE Trans. Geosci. Remote Sensing* **GE-22** 3–13
- [8] Devaney A J and Zhang D 1991 Geophysical diffraction tomography in a layered background *Wave Motion* **14** 243–65
- [9] Goodman J W 1968 *Introduction to Fourier Optics* (New York: McGraw-Hill)
- [10] Hager W W 1988 *Applied Numerical Linear Algebra* (Englewood Cliffs, NJ: Prentice-Hall)
- [11] Ishimaru A 1991 *Electromagnetic Wave Propagation, Radiation, and Scattering* (Englewood Cliffs, NJ: Prentice-Hall)
- [12] Jain A K 1986 *Fundamentals of Digital Image Processing* (Englewood Cliffs, NJ: Prentice-Hall)
- [13] Kerns D M 1981 *Plane-Wave Scattering-Matrix Theory of Antennas and Antenna-Antenna Interactions* (National Bureau of Standards Monograph) (Washington, DC: US Govt Printing Office) p 162
- [14] Ladas K T and Devaney A J 1993 Application of an ART algorithm in an experimental study of ultrasonic diffraction tomography *Ultrasonic Imaging* **15** 48–58
- [15] Maleki M H and Devaney A J 1993 Phase-retrieval and intensity-only reconstruction algorithms for optical diffraction tomography *J. Opt. Soc. Am. A* **10** 1–7
- [16] Morse P M and Feshbach H 1953 *Methods of Mathematical Physics* vols 1 and 2 (New York: McGraw-Hill)
- [17] Natterer F 1986 *The Mathematics of Computerized Tomography* (New York: Wiley)
- [18] Noble B and Daniel J W 1988 *Applied Linear Algebra* (Englewood Cliffs, NJ: Prentice Hall)

- [19] Norton S J and Linzer M 1981 Ultrasonic reflectivity imaging in three dimensions: exact inverse scattering solutions for plane, cylindrical, and spherical apertures *IEEE Trans. Biomed. Eng.* **BME-28** 202–20
- [20] Pratt R G and Worthington M H 1988 The application of diffraction tomography to cross-hole seismic data *Geophysics* **53** 1284–94
- [21] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes in FORTRAN: the Art of Scientific Computing* 2nd edn (Cambridge: Cambridge University Press)
- [22] Schatzberg A and Devaney A J 1992 Super-resolution in diffraction tomography *Inverse Problems* **8** 149–64
- [23] Schatzberg A, Devaney A J and Deming R W 1995 Maximum likelihood estimation of target location in acoustic and electromagnetic imaging *65th Ann. Meeting of the Society of Exploration Geophysicists (Houston)* Expanded Abstracts
- [24] Slaney M, Kak A C and Larsen L E 1984 Limitations of imaging with first order diffraction tomography *IEEE Trans. Microwave Theory Tech.* **MTT-32** 860–73
- [25] Tarantola A 1984 Linearized inversion of seismic reflection data *Geophys. Prosp.* **32** 998–1015
- [26] Witten A J and Molyneux J E 1992 Geophysical diffraction tomography: validity and implementation *Geophysical Inversion* ed J B Bednar, L R Lines, R H Stolt and A B Weglein (New York: SIAM) pp 354–69
- [27] Witten A J, Molyneux J E and Nyquist J E 1994 Ground penetrating radar tomography: algorithms and case studies *IEEE Trans. Geosci. Remote Sensing* **32** 461–7
- [28] Wu R and Toksoz M N 1987 Diffraction tomography and multisource holography applied to seismic imaging *Geophys.* **52** 11–25
- [29] Wu R, Araujo F V and Huang L 1994 Multifrequency backscattering tomography for constant and vertically varying backgrounds *Int. J. Imaging Systems Tech.* **5** 7–21

Monte Carlo Simulations for Clutter Statistics in Minefields: AP-Mine-Like-Target Buried Near a Dielectric Object Beneath 2-D Random Rough Ground Surfaces

Magda El-Shenawee¹ and Carey Rappaport²

¹Department of Electrical Engineering
University of Arkansas
Fayetteville, Arkansas 72701
Tel: (501)-575-6582, Fax: (501)-575-7967
magda@uark.edu

²Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
Tel: 617-373-2064, Fax: 617-373-8627
rappaport@neu.edu

Abstract

A rigorous three-dimensional (3-D) electromagnetic model is developed to analyze the scattering from anti-personnel (AP) nonmetallic mine-like target when it is buried near a clutter object under two-dimensional (2-D) random rough surfaces. The Steepest Descent Fast Multipole Method (SDFMM) is implemented to solve for the unknown electric and magnetic surface currents on the ground surface, on the target and on the clutter object. A comprehensive numerical investigation of two clutter sources; the ground roughness and the nearby benign object, is presented based on using more than 800 random rough surface realizations which could not be achieved without using fast algorithms such as the SDFMM. The statistics of the scattered near-electric fields are computed using the Monte Carlo simulations for both polarizations. For the parameters used here, the results show that the average and the standard deviation of the target signature represent 5-7% and 3-3.5% of the total scattered signal, respectively, while they represent 16-20% and 7-12% of the signal for the clutter object, respectively. This study indicates the high possibility of a false alarm during the detection process when the target is located nearby a realistic object such as a piece of a tree root.

Key words: Multiple buried objects, subsurface sensing, rough surface scattering, SDFMM, Monte Carlo simulations, clutter statistics.

I. INTRODUCTION

Electromagnetic sensing of targets buried in soil with two-dimensional random rough surface necessitates the investigation of clutter sources that may cause false alarms in the detection process. In real minefields, there are a variety of clutters that can easily obscure the signal of the target such as: the roughness of the ground, nearby benign objects (e.g., tree roots, rocks, etc.), spots of concentrated inhomogeneity in the soil, etc. Detecting small anti-personnel nonmetallic mines is very difficult as it is, but proximity of clutter objects obscures target detection considerably.

The statistics of the scattered waves from the random rough ground were previously investigated in the literature either analytically, [1]-[4], or computationally using the Monte Carlo simulations, [5]-[10]. The challenging part of conducting the Monte Carlo simulations is to run a 3-D computer code hundreds of times when the rough ground has large electrical size and contains multiple penetrable objects buried under the interface. In [1]-[4], integral expressions were developed to calculate the incoherent radar cross section (RCS) of the rough ground (no buried objects). These integrations were basically obtained by considering the ground heights and slopes as random variables producing analytic closed forms in some cases (e.g. at high frequency). On the other hand, conducting the Monte Carlo numerical simulations were used to obtain the incoherent RCS of the rough surface without buried objects in [5]-[8], to obtain the RCS and/or the angular correlation function of a PEC sphere buried under the rough interface [9], and to obtain the statistical average of the scattered near-fields when a penetrable object was buried under the interface in [10].

In our previous work [10] and [11], we implemented the Steepest Descent Fast Multipole Method (SDFMM) [6], [12] to investigate the effect of the ground roughness on the signature of a single target buried under the ground. The $O(N)$ computational complexity of the SDFMM accelerated the calculations of the unknown surface currents on the shallow mine-like object buried beneath a moderately random rough ground especially in conducting the Monte Carlo simulations. Similarly, to study the influence of the ground roughness (with no buried targets) on the scattered signals for different minefields, the SDFMM was used to run the Monte Carlo simulations from Puerto-Rican and Bosnian soils [13].

To the best of our knowledge, no work is published for investigating the statistics of multiple objects buried under the 2-D random rough ground, which is the objective of this paper. In [14], rigorous generalized electromagnetic formulations were developed for scattering from multiple objects buried under the rough ground, where the SDFMM was implemented to speed up the computations. A parametric investigation is conducted in [14] to study the effect on the scattered signature due to the objects proximity, ground roughness, object's material, shape, location, etc. A variety of object shapes were studied in [14], e.g., one object is spheroid and the second object is ellipsoid, spheroid, disk, or horizontal cylinder.

In this paper, we are using the Monte Carlo simulations to conduct a statistical study of mine-like dielectric target's signature versus the clutter signature due to both the ground roughness and a benign object as shown in Fig. 1. Even though, we are emphasizing here only on these two types of clutter, however, there are additional types of clutter that may significantly affect the target signature such as the soil inhomogeneities and the multilayered nature of the ground, which are not accounted for in this work. According to the United States Department of Defense data (US DoD), European agencies and others, it is a fact that there are hundreds of different

types, shapes and manufacturers of the AP-plastic mines. This is one of the reasons that these mines are difficult to detect. In this work, we are not modeling a specific type of these mines but instead modeling a plastic-mine-like target that has a shape very close to an oblate spheroid. Several examples of spheroid-shape types can be obtained from the US DoD CD-ROM Humanitarian Demining Equipment database.

In Section II the formulations of the electromagnetic model developed in [14] are summarized, numerical results are shown in Section III and conclusions are drawn in Section IV.

II. PROBLEM FORMULATIONS

The rigorous electromagnetic model developed in [14] is employed in this work to conduct the Monte Carlo simulations for two objects buried under the 2-D random rough ground as shown in Fig. 1. The inhomogeneous scatterer is composed of four different regions; air R_1 , ground R_2 , target R_3 and clutter object R_4 where the relative permittivity and permeability are (ϵ_1, μ_1) , (ϵ_2, μ_2) , (ϵ_3, μ_3) and (ϵ_4, μ_4) , respectively, as shown in Fig. 2a. The unknown equivalent electric and magnetic surface currents are (\bar{J}_1, \bar{M}_1) on S_1 , (\bar{J}_3, \bar{M}_3) on S_2 , and (\bar{J}_5, \bar{M}_5) on S_3 . The final set of surface integral equations on S_1 , S_2 , and S_3 in Fig. 2a are given by [14]:

$$\bar{E}^{inc}(\bar{r})\big|_{\text{tang.}} = \left[(L_1 + L_2)\bar{J}_1 - (K_1 + K_2)\bar{M}_1 - L_3\bar{J}_3 + K_3\bar{M}_3 - L_4\bar{J}_5 + K_4\bar{M}_5 \right]_{\text{tang.}}, \bar{r} \in S_1 \quad (1a)$$

$$\bar{H}^{inc}(\bar{r})\big|_{\text{tang.}} = \left[(K_1 + K_2)\bar{J}_1 + \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{M}_1 - K_3\bar{J}_3 - \frac{L_3}{\eta_2^2} \bar{M}_3 - K_4\bar{J}_5 - \frac{L_4}{\eta_2^2} \bar{M}_5 \right]_{\text{tang.}}, \bar{r} \in S_1 \quad (1b)$$

$$0 = \left[-L_2\bar{J}_1 + K_2\bar{M}_1 + (L_3 + L_5)\bar{J}_3 - (K_3 + K_5)\bar{M}_3 + L_4\bar{J}_5 - K_4\bar{M}_5 \right]_{\text{tang.}}, \bar{r} \in S_2 \quad (1c)$$

$$0 = \left[-K_2 \bar{J}_1 - \frac{L_2}{\eta_2^2} \bar{M}_1 + (K_3 + K_5) \bar{J}_3 + \left(\frac{L_3}{\eta_2^2} + \frac{L_5}{\eta_3^2} \right) \bar{M}_3 + K_4 \bar{J}_5 + \frac{L_4}{\eta_2^2} \bar{M}_5 \right]_{\text{tang.}}, \bar{r} \in S_2 \quad (1d)$$

$$0 = \left[-L_2 \bar{J}_1 + K_2 \bar{M}_1 + L_3 \bar{J}_3 - K_3 \bar{M}_3 + (L_4 + L_6) \bar{J}_5 - (K_4 + K_6) \bar{M}_5 \right]_{\text{tang.}}, \bar{r} \in S_3 \quad (1e)$$

$$0 = \left[-K_2 \bar{J}_1 + \frac{L_2}{\eta_2^2} \bar{M}_1 + K_3 \bar{J}_3 + \frac{L_3}{\eta_2^2} \bar{M}_3 + (K_4 + K_6) \bar{J}_5 + \left(\frac{L_4}{\eta_2^2} + \frac{L_6}{\eta_4^2} \right) \bar{M}_5 \right]_{\text{tang.}}, \bar{r} \in S_3 \quad (1f)$$

in which the intrinsic impedance is $\eta_i = \sqrt{\mu_i / \epsilon_i}$ in region R_i with $i = 1, 2, \dots, 4$ and L_j and K_j ,

$j = 1, 2, \dots, 6$ are the integro-differential operators given in Appendix A. The tangential

component of the incident electric and magnetic fields on the rough surface S_1 are given by

$\bar{E}^{inc}(\bar{r})|_{\text{tang}}$ and $\bar{H}^{inc}(\bar{r})|_{\text{tang}}$. The surfaces S_1, S_2 and S_3 are discretized into triangular patches

where the unknown equivalent electric and magnetic currents in (1) are approximated by using

the Rao, Wilton and Glisson (RWG) vector basis functions $\bar{j}(\bar{r})$ [15], [16] as:

$$\bar{J}_i(\bar{r}) = \sum_{n=1}^{N_k} I_{in} \bar{j}_{kn}(\bar{r}), \quad \bar{M}_i(\bar{r}) = \eta_1 \sum_{n=1}^{N_k} I_{(i+1)n} \bar{j}_{kn}(\bar{r}), \quad \bar{r} \in S_k, \quad \text{for } i = 1, 3 \text{ and } 5 \quad (2)$$

and $k = (i + 1) / 2$. Upon substituting (2) in (1), and testing using the same vector basis functions,

the linear system of equations is obtained [14]:

$$\bar{\bar{Z}} \bar{\bar{I}} = \bar{\bar{V}} \Rightarrow \begin{pmatrix} \bar{Z}_{11} & \bar{Z}_{12} & \bar{Z}_{13} \\ \bar{Z}_{21} & \bar{Z}_{22} & \bar{Z}_{23} \\ \bar{Z}_{31} & \bar{Z}_{32} & \bar{Z}_{33} \end{pmatrix} \begin{pmatrix} I^1 \\ I^2 \\ I^3 \end{pmatrix} = \begin{pmatrix} \bar{V}_1 \\ 0 \\ 0 \end{pmatrix} \quad (3)$$

The total matrix $\bar{\bar{Z}}$ has order of $2(N_1 + N_2 + N_3) \times 2(N_1 + N_2 + N_3)$, where the number of

unknowns on the ground, on the target and on the clutter object are $2N_1$, $2N_2$ and $2N_3$,

respectively. It is necessary to emphasize that this rigorous electromagnetic model considers all

scattering and interaction scenarios; the self-interactions of the ground (\bar{Z}_{11}), target (\bar{Z}_{22}) and

clutter object (\bar{Z}_{33}), the target with the ground and with clutter object (\bar{Z}_{12} , \bar{Z}_{23}) and the clutter object with the ground \bar{Z}_{13} , etc. The expressions of these submatrices are given in Appendix B. The tested tangential incident electric field \bar{E}^{inc} and the tested normalized magnetic field $\eta_1 \bar{H}^{inc}$ on the exterior surface of the ground are expressed as part of vector \bar{V} in (3) while the vector \bar{I} represents the unknown current coefficients on all involved surfaces.

The emphasis of this study is to investigate the statistics of the target signature versus the clutter signature. The proposed strategy is to use the same sample of random rough surface realizations four times as follows: (i) once to compute the scattering from the rough ground alone (no buried objects); (ii) once to compute the scattering from the rough ground when both the target and clutter object are buried beneath the interface; (iii) once to compute the scattering from the rough ground when only the target is buried beneath the interface; and (iv) once to compute the scattering from the rough ground when only the clutter object is buried beneath the interface. The scattered electric near-fields are obtained in this procedure at single incident angle, single frequency, for specific polarization (e.g., vertical or horizontal), etc. This shows that conducting the Monte Carlo simulations is computationally expensive requiring fast methods such as the SDFMM. In this case, the computational complexity of the problem will be of $O(K = 2N_1 + 2N_2 + 2N_3)$ as described earlier. Once the electric and magnetic surface currents are obtained, then the scattered fields above the ground and in the near-zone can be calculated using the surface integrations given by [17]:

$$\bar{E}^A(x, y, z) = -\frac{1}{4\pi(i\omega\epsilon_0)} \nabla \times \iint_{S'} (\hat{R} \times \bar{J}_1(x', y', z')) \frac{1 + ik_0 R}{R^2} \exp(-ik_0 R) dS' \quad (4a)$$

$$\bar{E}^F(x, y, z) = -\frac{1}{4\pi} \nabla \times \iint_{S'} \bar{M}_1(x', y', z') \frac{\exp(-ik_0 R)}{R} dS' \quad (4b)$$

where the total electric field is $\bar{E}^A + \bar{E}^F$, the superscripts A and F refer to the vector potentials \bar{A} and \bar{F} , the point source (x', y', z') is located on the ground while the observation point (x, y, z) is located above the ground, \hat{R} and R are the unit vector and the distance between the observation and the source points, k_0 , ϵ_0 and μ_0 are the wave number, the permittivity, and permeability of the free space, respectively, and dS' is the differential surface element on the rough ground. Every source point (x', y', z') in (4) is the centroid of every triangular patch on the ground at which the surface currents are approximately evaluated [15].

III. NUMERICAL RESULTS

The principle of computing the target signature at multiple views was employed in [9] and [11] by rotating the incident beam around the target, i.e., changing the azimuth angle of the incident waves. As presented in [14], when the target is buried nearby a second object, a significant electromagnetic interference between the two objects is observed in the scattered fields at single azimuth angle. In this work, the Monte Carlo simulations will be conducted using the incident beam view (azimuth angle) that largely indicates the presence of a clutter object buried nearby the target.

Trying to simulate the ground penetrating radar (GPR) of center frequency 1GHz experiments reported in [18]-[19], the scattered electric fields at $\lambda_0/2$ above the rough ground mean are computed, where λ_0 is the free space wavelength. In addition, to simulate the GPR experimental track, the ground is modeled as a plane of dimensions $8\lambda_0 \times 8\lambda_0$. For the same reason, all simulations are conducted here at 1GHz; however, the current technique was successfully applied for the frequency band 800MHz-1.4GHz (not presented here). The rough ground surface is characterized with Gaussian statistics for the heights, assuming zero mean and

root mean square (rms) σ , and for the autocorrelation function with correlation length l_c . It was observed that the incident beam pattern of the GPR parabolic antenna reported in [18], [19] resembles a Gaussian beam. Thus for simplicity, the incident wave in the current simulations is assumed to be a Gaussian beam tapered toward the edges of the ground with half-beam width of $1.6\lambda_0$ centered at $x = y = 4\lambda_0$ [9], [20]. The incident angle of the beam, measured from the z -direction, is $\theta^i = 30^\circ$ while the azimuth angle ϕ^i will be described. As mentioned in Section I, the AP-mine-like target is modeled here as an oblate spheroid with dimensions $a = 0.15\lambda_0, b = 0.3\lambda_0$ (Fig. 2b) where its center is located at $x = 4.5\lambda_0, y = 3.5\lambda_0$ and $z = -0.4\lambda_0$ (lower right quadrant of the ground as shown in Fig. 3a). While, the clutter object is modeled as a horizontal cylinder (Fig. 2b) such as a piece of a tree root with radius $a = 0.15\lambda_0$ and height $h = 0.9\lambda_0$ with its axis tilted at angle 30° with the x -direction and located at $x = 4.01\lambda_0, y = 4.375\lambda_0$ and $z = -0.4\lambda_0$ measured from the axis mid-point (upper left quadrant of the ground as shown in Fig. 3a). This implies a $1.0\lambda_0$ distance separating the centers of the two objects. In [14], a parametric study was conducted for different ground roughness in which the clutter for $\sigma = 0.1\lambda_0$ was shown, as expected, to be stronger than that of $\sigma = 0.04\lambda_0$ [11]. Therefore, in this work, the rms height is assumed to be $\sigma = 0.1\lambda_0$ with correlation length $l_c = 0.5\lambda_0$. The relative dielectric constants of the soil, the AP-mine-like target, and the tree root object are assumed to be $\epsilon_r = 2.5 - j0.18$ (Bosnian soil with 3.8% moisture at 1GHz) [21], $\epsilon_r = 2.9 - j0.072$ (TNT) [22], and $\epsilon_r = 4.0$ (dry wood) [23], respectively. The near-electric fields are computed using (4) at point receivers located at $z = 0.5\lambda_0$ above the ground with resolution equal to $x = y = 0.04\lambda_0$. Several numerical experiments, depending on the dielectric

constants and on the ground roughness parameters, were conducted to choose the discretization distances on all surfaces. The adopted criterion is to obtain a solution that is independent of these distances. In this work, $0.08\lambda_0$ discretization distance on the rough ground is used leading to 60,000 ($2N_1$) current unknowns on its surface. For the two objects, approximately the same discretization distances are used leading to 600 ($2N_2$) on the AP-mine-like target and 600 ($2N_3$) on the tree root. The triangular mesh for the rough ground, the target (oblate spheroid) and the tree root (horizontal cylinder) is shown in Fig. 2b. To speedup the solution's convergence, a preconditioner consists of the diagonal self-elements of $\bar{\bar{Z}}$ in (3) is used. A relative residual error of 10^{-3} is used in the Transpose-Free-Quasi-Minimal-Residual (TFQMR) iterative solver [24] and $0.32\lambda_0 \times 0.32\lambda_0$ is assumed for the smallest Fast Multipole Method (FMM) block size. Moreover, the SDFMM is validated with the method of moments (MoM) as discussed in [14]. The complex subtraction process is used in this work to remove the clutter due to both the random rough ground and the benign object (tree root) leading to obtain a statistical study of the target signature compared with these sources of clutter [25]. The numerical results are divided into sub-sections as shown below.

A. Multiple views

The azimuth angle of the incident beam is varied as $\phi^i = 0^\circ, 30^\circ, 60^\circ, \dots, 360^\circ$ (i.e., the azimuth angle of the plane of incidence). The fields scattered from just the rough surface realization (no buried objects) is computed and subtracted from those scattered from the same rough surface realization when both the AP-mine-like target and the tree root are buried, and shown in Fig. 3a. The results for the incident horizontal polarization are considered in this case where the electric field is perpendicular to the plane of incidence. Only four results are presented

for $\phi^i = 0^\circ, 60^\circ, 120^\circ$, and 240° in Figs. 3b-e. The results show almost no indication to the presence of two separate objects except at $\phi^i = 120^\circ$ where the mine-like target signature can be distinguished from that of the tree root. At this azimuth angle, the plane of incidence is perpendicular to the tree root axis. The results of Figs. 3b-e are for single rough surface realization. The interesting view at $\phi^i = 120^\circ$ will be used in all Monte Carlo simulations considered in this work for both polarizations.

B. Monte Carlo simulations

A sample of 100 independent random rough surface realizations was used in this work. However, the fields scattered from the geometry shown in Fig. 1 and described above was computed 800 times; 400 times for the horizontal polarization and 400 times for the vertical polarization as described in cases (i-iv) in Section II. The statistical average of the electric fields, i.e., $\langle \bar{E} \rangle$, and the standard deviation (STD), i.e. $\sqrt{\langle |\bar{E}|^2 \rangle - \langle \bar{E} \rangle^2}$, are presented [10]. It is necessary to mention that in the horizontal polarization case, the incident electric field is parallel to the tree root axis, while in the vertical polarization case it is perpendicular to its axis.

(i) Horizontal polarization

The results of the cases (i-iv) are plotted versus the x - and y -directions and shown in Figs. 4a-d, respectively. The slight differences between the results in these figures clearly indicate that the rough ground is dominating the scattering scenario. Then, the results of case (i) are subtracted from those of case (ii) and the output is plotted in Fig. 5a where the average signature of both the clutter object and the target are shown. In addition, a clutter due to the interaction between the tree root and the rough ground is clearly observed in Fig. 5a. Similarly, the results of case (iii)

are subtracted from those of case (ii) and the output is plotted in Fig. 5b where the average signature of just the tree root is shown in addition to clutter due to interaction with the nearby target. Finally, the results of case (iv) are subtracted from those of case (ii) and the output is plotted in Fig. 5c where the average signature of just the target is shown in addition to interaction with the clutter object. It is important to mention that the subtraction process does not completely remove the effect of the rough ground or the nearby object; however, the remainder of these interactions depends on the ground roughness and the object proximity [14]. The results of Figs. 5-c show a remarkable difference between the magnitude of the clutter object average signature (max. of 0.037 V/m) and that of the target average signature (max. of 0.009 V/m). Moreover in Fig. 5a, the presence of the target causes the magnitude of the clutter object signature to slightly increase due to the interactions with the target as discussed in [14]. All these results, which are obtained using the same 100 rough surface realizations, show that a false alarm is highly possible due to the presence of a nearby clutter object (a piece of a tree root in this case).

(ii) Vertical polarization

Upon changing the incident fields to the vertical polarization, then cases (i-iv) are re-computed and plotted in Figs. 6a-d, respectively. In these figures, the maximum magnitude of the average scattered electric fields has dropped by almost 50%, even for the case of rough ground alone, compared with the horizontal polarization results in Fig. 5. This is because the incident electric fields at oblique angle are decomposed into two components; a component in the z-direction and a component parallel to the x-y plane (perpendicular to the tree root axis). While for the horizontally polarized incident fields, all the electric fields are parallel to the x-y plane (parallel to the tree root axis). Similarly, results of Figs. 6a-d show that the rough ground is dominating the scattering mechanism such that a slight difference between the results in Fig. 6a-d

can be observed. The subtraction process is repeated in this case and the average signature of both the clutter object and the target is shown in Fig. 7a, of just the clutter object is shown in Fig. 7b and of just the target is shown in Fig. 7c. The magnitude of the average target signature (max. of 0.0075 V/m) is smaller than that of the clutter object (max of 0.018 V/m), which agrees with the observation in the horizontal polarization case.

The standard deviation of the scattered fields is plotted in Fig. 8a for just the tree root, in Fig. 8b for just the target for the horizontal polarization. In these figures, the maximum standard deviation occurs almost at the location of the tree root as shown in Fig. 8a and of the target as shown in Fig. 8b. For the vertical polarization case, the standard deviation is plotted in Figs. 8c and 8d for the tree root and for the target, respectively. However, the results show that the maximum standard deviation of the clutter object is shifted from the location of the object. This could be attributed to the electromagnetic interference between the target and the clutter object [14]. In addition, the maximum magnitude of the standard deviation of the target signature is almost the same for both polarizations, however, for the tree root it is reduced by almost 50% in the vertical polarization. Therefore, the relative statistics are computed with respect to signature of the ground without buried objects (i.e. object signature/ground signature) as summarized in Table 1.

The total CPU time required for computing the scattered electric fields from one surface realization is less than 3 Hrs (after almost 200 TFQMR iterations) and the required computer memory is 850 MB. These computations were conducted using the Compaq Alpha Server (GS140 EV6) with 667MHz clock speed [14].

Table 1. Maximum relative statistics for the scattered electric fields

	Maximum relative average		Maximum relative STD	
	Clutter object	Target	Clutter object	Target
H-Pol.	20%	5%	12%	3%
V-Pol.	16%	6.5%	7.5%	3.5%

It is necessary to mention that the current work is not a signal processing technique that can be applied to the real measured data [26]. However, the current technique presents a powerful computational tool to analyze, study and statistically estimate some of the causes of false alarms encountered in the detection process.

IV. CONCLUSIONS

The statistics of the target, the rough ground and the clutter object signatures depend on several factors such as the orientation, dielectric constants, proximity between objects, burial depth, etc. For the parameters used in this work, the statistical study showed that the target signature appears much weaker than the clutter object signature even for a small realistic nearby buried object such as a piece of a tree root. Moreover, consistent with our previous work [10]-[11], the rough ground clutter dominates the scattering scenario necessitating its removal in the detection process. The conducted statistical study is based on running the 3-D scattering computer code more than 800 times which could not be achieved without using a fast algorithm such as the SDFMM.

ACKNOWLEDGMENTS

This research was sponsored from the Northeastern University's Demining MURI grant # DAA 0-55-97-0013. The SDFMM was originally developed at the UIUC by V. Jandhyala, E. Michielssen and W. Chew. This work was conducted using the ASCC at NU.

Appendix A

With representing the surface electric and magnetic currents \bar{J} and \bar{M} on S_1 , S_2 and S_3 by the vector \bar{X} , the integro-differential operators L_j and $K_j, j=1, 2, \dots, 6$, are [14], [16]:

$$L_{1,2}\bar{X} = \int_{S_1} \left\{ i\omega\mu_{1,2}\Phi_{1,2}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{1,2}}\nabla\nabla' \cdot \bar{X}(\bar{r}')\Phi_{1,2} \right\} ds', \quad K_{1,2}\bar{X} = \int_{S_1} \bar{X}(\bar{r}') \times \nabla\Phi_{1,2} ds' \quad (A1)$$

$$L_{3,5}\bar{X} = \int_{S_2} \left\{ i\omega\mu_{2,3}\Phi_{2,3}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{2,3}}\nabla\nabla' \cdot \bar{X}(\bar{r}')\Phi_{2,3} \right\} ds', \quad K_{3,5}\bar{X} = \int_{S_2} \bar{X}(\bar{r}') \times \nabla\Phi_{2,3} ds' \quad (A2)$$

$$L_{4,6}\bar{X} = \int_{S_3} \left\{ i\omega\mu_{2,4}\Phi_{2,4}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{2,4}}\nabla\nabla' \cdot \bar{X}(\bar{r}')\Phi_{2,4} \right\} ds', \quad K_{4,6}\bar{X} = \int_{S_3} \bar{X}(\bar{r}') \times \nabla\Phi_{2,4} ds' \quad (A3)$$

Appendix B

The elements of the submatrix \bar{Z}_{11} in (3) are given by [14], [16]

$$\bar{Z}_{11} = \begin{pmatrix} \langle \bar{j}_1, (L_1 + L_2)\bar{j}_1 \rangle_{S_1} & \langle \bar{j}_1, -\eta_1(K_1 + K_2)\bar{j}_1 \rangle_{S_1} \\ \langle \bar{j}_1, \eta_1(K_1 + K_2)\bar{j}_1 \rangle_{S_1} & \left\langle \bar{j}_1, \eta_1^2 \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{j}_1 \right\rangle_{S_1} \end{pmatrix} \quad (B1)$$

in which $\langle \bar{A}, \bar{B} \rangle_S$ denotes the complex inner product between vector functions \bar{A} and \bar{B} on a

surface S . The submatrices \bar{Z}_{12} , \bar{Z}_{13} , and \bar{Z}_{23} are given by

$$\bar{Z}_{12} = \begin{pmatrix} \langle \bar{j}_1, -L_3\bar{j}_2 \rangle_{S_1} & \langle \bar{j}_1, \eta_1 K_3 \bar{j}_2 \rangle_{S_1} \\ \langle \bar{j}_1, -\eta_1 K_3 \bar{j}_2 \rangle_{S_1} & \left\langle \bar{j}_1, -\eta_1^2 \left(\frac{L_3}{\eta_2^2} \right) \bar{j}_2 \right\rangle_{S_1} \end{pmatrix}, \quad \bar{Z}_{13} = \begin{pmatrix} \langle \bar{j}_1, -L_4\bar{j}_3 \rangle_{S_1} & \langle \bar{j}_1, \eta_1 K_4 \bar{j}_3 \rangle_{S_1} \\ \langle \bar{j}_1, -\eta_1 K_4 \bar{j}_3 \rangle_{S_1} & \left\langle \bar{j}_1, -\eta_1^2 \left(\frac{L_4}{\eta_2^2} \right) \bar{j}_3 \right\rangle_{S_1} \end{pmatrix} \quad (B2)$$

$$\bar{Z}_{23} = \begin{pmatrix} \langle \bar{j}_2, L_4\bar{j}_3 \rangle_{S_2} & \langle \bar{j}_2, -\eta_1 K_4 \bar{j}_3 \rangle_{S_2} \\ \langle \bar{j}_2, \eta_1 K_4 \bar{j}_3 \rangle_{S_2} & \left\langle \bar{j}_2, \eta_1^2 \frac{L_4}{\eta_2^2} \bar{j}_3 \right\rangle_{S_2} \end{pmatrix} \quad (B3)$$

Similar expressions can be obtained for all other submatrices in (3).

References

- [1] P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, New York, 1963.
- [2] E. Bahar and M. El-Shenawee, "Double scatter cross sections for two dimensional random rough surfaces that exhibit backscatter enhancement," *J. of Optical Society of America A*, vol. 18, no. 1, pp. 108-116, January 2001.
- [3] A. Ishimaru and J. S. Chen, "Scattering from very rough metallic and dielectric surfaces: a theory based on the modified Kirchhoff approximation," *Waves in Random Media*, pp. 21-34, January 1991.
- [4] E. Thorsos, "The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum," *J. Acoust. Soc. Am.* 83 (1), 1988.
- [5] R. L. Wagner, J. Song and W. C. Chew, "Monte Carlo simulation of electromagnetic scattering from two-dimensional random rough surfaces," *IEEE Trans. Anten. & Prop.*, vol. 45, no. 2, pp. 235-245, February 1997.
- [6] V. Jandhyala, E. Michielssen, B. Shanker and W.C. Chew, "A combined steepest descent-fast multipole algorithm for the fast analysis of three-dimensional scattering by rough surfaces," *IEEE Trans. Geosci. & Remote Sensing*, vol. 36, no. 3, pp. 738-748, May 1998.
- [7] J. T. Johnson, L. Tsang, R. T. Shin, K. Pak, C. H. Chan, A. Ishimaru and Y. Kuga, "Backscattering enhancement of electromagnetic waves from two-dimensional perfectly conducting random rough surfaces: A comparison of Monte Carlo simulations with experimental data," *IEEE Trans. Anten. & Prop.*, vol. 44, no. 5, pp. 748-756, May 1996.
- [8] L. Tsang, C. H. Chan, K. Pak, H. Sangani, A. Ishimaru and P. Phu, "Monte Carlo simulations of large-scale composite random rough-surface scattering based on the banded-matrix iterative approach," *J. Optical Society of America*, Vol. 11, No. 2, pp.691-696, 1994.
- [9] G. Zhang, L. Tsang and K. Pak, "Angular correlation function and scattering coefficient of electromagnetic waves scattered by a buried object under a two-dimensional rough surface," *J. Opt. Soc. Am. A*, vol. 15, no. 12, pp. 2995-3002, December 1998.
- [10] M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo simulations of electromagnetic wave scattering from random rough surface with 3-D penetrable buried object: mine detection application using the SDFMM," *J. Optical Society of America A*, to appear in December 2001.
- [11] M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, "3-D subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces: use of the steepest descent fast multipole method (SDFMM)," *IEEE Trans. Geosci. & Rem. Sensing*, vol. 39, no. 6, pp. 1174-1182, June 2001
- [12] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [13] M. El-Shenawee and C. Rappaport, "Modeling Clutter from Bosnian and Puerto Rican Rough Ground Surfaces for GPR Subsurface Sensing Applications Using the SDFMM Technique," accepted for publication in the *J. Subsurface Sensing Technologies and Applications*, to appear.
- [14] M. El-Shenawee, "The Steepest Descent Fast Multipole Method for Scattering from Two Penetrable Shallow Objects Buried Under Two-Dimensional Randomly Rough Surface," submitted in June 2001 to the *IEEE Trans. Anten. & Propag.*, under review.
- [15] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Anten. & Prop.*, vol. AP-30, no. 3, pp.409-418, May 1982.
- [16] L. Medgyesi-Mitschang, J. Putnam, and M. Gedera, "Generalized method of moments for three-dimensional penetrable scatterers," *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1383-1398, April 1994.
- [17] C. A. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons Inc, Ch. 6, pp. 254-309, 1989.

- [18] C. Rappaport, S. Azevedo, T. Rosenbury, and J. Gough, and A. Dean, "Detecting antipersonnel mines with a handheld parabolic reflector transmitter/multistatic receiver impulse GPR," 2000 UXO/Countermining Forum, May 2000.
- [19] C. Rappaport, S. Azevedo, T. Rosenbury, and J. Gough, "Handheld forward-looking focused array mine detection with plane wave excitation," *SPIE Aerosense Conference*, Orlando, FL, pp. 1118-1126, April 2000.
- [20] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.
- [21] J. Curtis, "Dielectric properties of soils; various sites in Bosnia," *US Army Corp. of Eng., Waterways Experim.*, Station Data Rep., 1996.
- [22] Von Hippel, *Dielectric Materials and Applications*, Wiley (New York), p. 315, 1953.
- [23] F. T. Ulaby, *Fundamentals of Applied Electromagnetics*, Prentice Hall, Appendix B, 2001.
- [24] R. W. Freund, "A Transpose-free quasi-minimal residual algorithm for non-hermitian linear systems," *SIAM J. Sci. Comput.*, vol. 14, no. 2, pp. 470-482, March 1993.
- [25] A. L. Warrick, S.G Azevedo and J. E. MST, "Prediction of buried mine-like target radar signature using wideband electromagetic modeling," *SPIE Aerosense Conference*, vol. 3392, pp. 776-783, 1998.
- [26] A. van der Merwe, I. J. Gupta, "A novel signal processing technique for clutter reduction in GPR measurements of small, shallow land mines," *IEEE Trans. Geosci. & Rem. Sensing*, vol. 38, no. 6, pp. 2627-2637, November 2000.

List of Figures

Figure 1. Cross section of a target buried near a clutter object beneath the rough ground surface.

Figure 2. (a) Two scatterers, R_3 and R_4 , immersed in region R_2 immersed in region R_1 and (b)

Figure 2b. Triangular mesh for the rough ground (upper figures), for the oblate spheroid and the horizontal cylinder (lower figure).

Figure 3. (a) Top view of the target (oblate spheroid) and the tree root (horizontal cylinder), (b-e) The scattered electric field at multiple views for both the buried mine-like target and the tree root upon removing the scattering from ground when: (b) $\phi^i = 0^\circ$, (c) $\phi^i = 60^\circ$, (d) $\phi^i = 120^\circ$, and (e) $\phi^i = 240^\circ$.

All results are for $\sigma = 0.1\lambda_0$, $I_c = 0.5\lambda_0$, $\theta^i = 30^\circ$ and H-polarization.

Figure 4. The average scattered electric field at $z = 0.5\lambda_0$ from (a) the ground with no buried objects, (b) the ground with both the buried mine-like target and tree root, (c) the ground with only the buried tree root (no mine), and (d) the ground with only the buried mine-like target (no tree root). Data are from Fig. 3c (H-polarization).

Figure 5. The average scattered electric field upon removing the scattering from the rough ground for (a) both the mine-like target and tree root, (b) just the tree root, (c) just the mine-like target. Data are from Fig. 3c. (H-polarization).

Figure 6. The average scattered electric field at $z = 0.5\lambda_0$ from (a) the ground with no buried objects, (b) the ground with both buried the mine-like target and tree root, (c) the ground with only the buried tree root (no mine), and (d) the ground with only the buried mine-like target (no tree root). Data are from Fig. 3c for V-polarization.

Figure 7. The average scattered electric field upon removing the scattering from the rough ground for (a) both the mine-like target and tree root, (b) just the tree root, (c) just the mine-like target. Data are from Fig. 3c but for V-polarization

Figure 8. The standard deviation of the scattered electric fields for (a) just the tree root, (b) just the AP-mine-like target, for the H-polarization, (c) just the tree root, and (d) jut the AP-mine-like target, for the V-polarization. Data are from Fig. 3c.

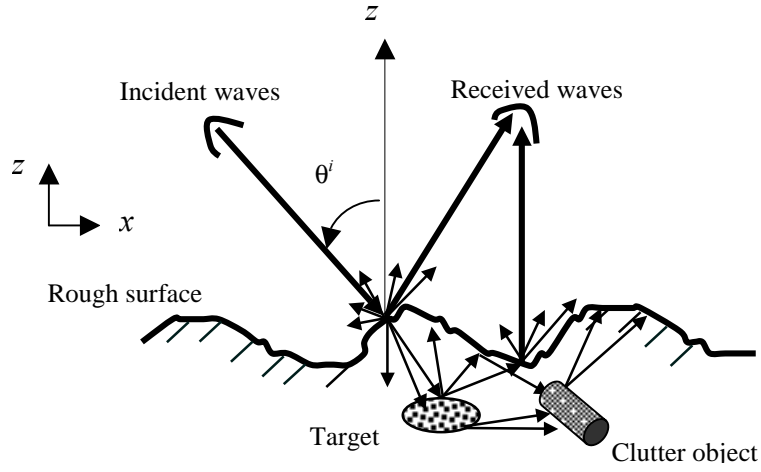


Figure 1. Cross section of a target buried near a clutter object beneath the rough ground surface.

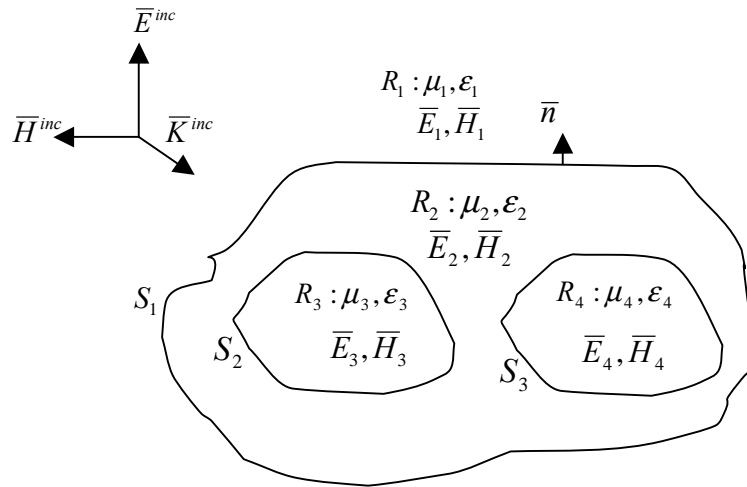


Figure 2a. Two scatterers, R_3 and R_4 , immersed in region R_2 immersed in region R_1 .

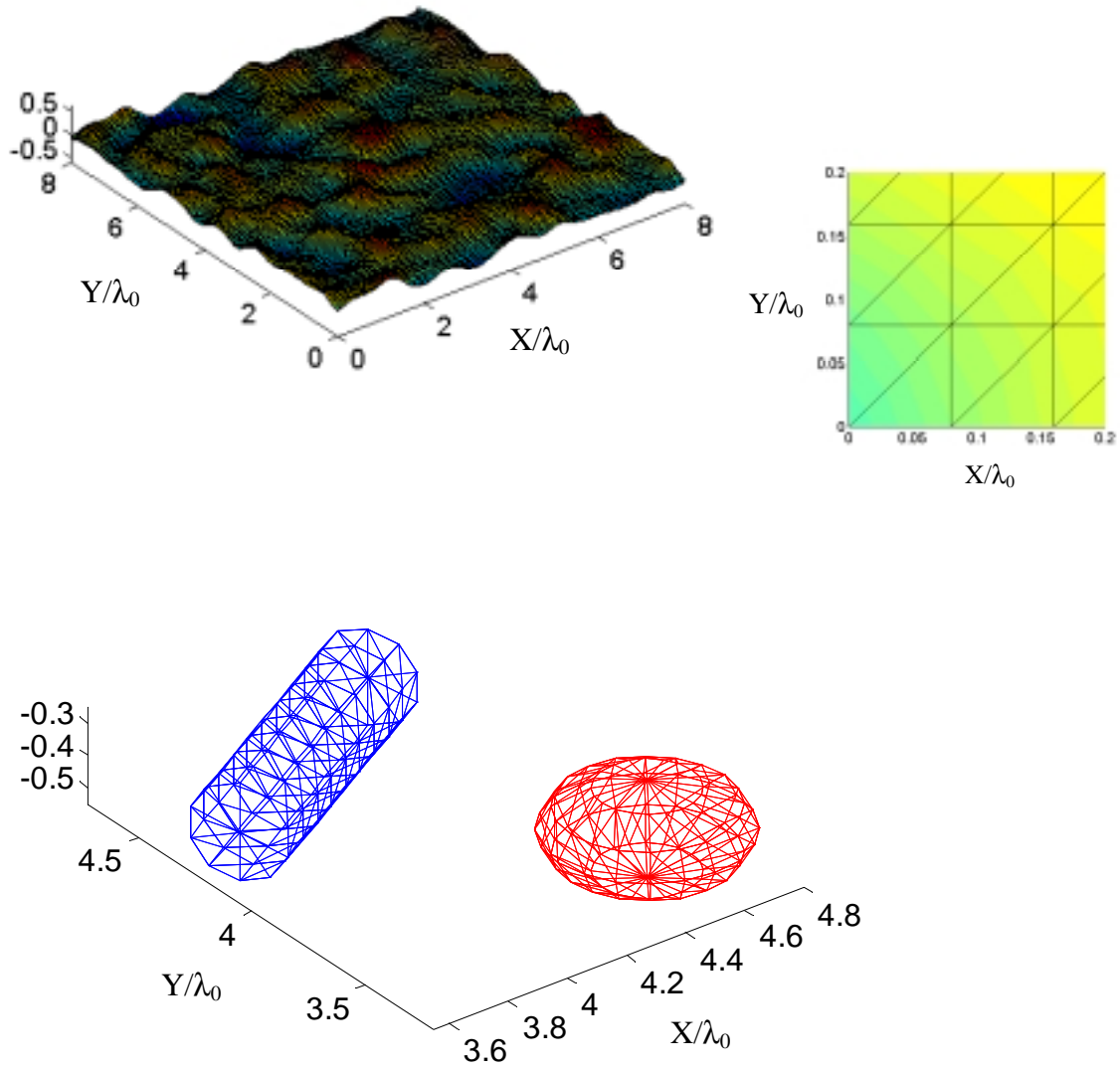


Figure 2b. Triangular mesh for the rough ground (upper figures), for the oblate spheroid and the horizontal cylinder (lower figure).

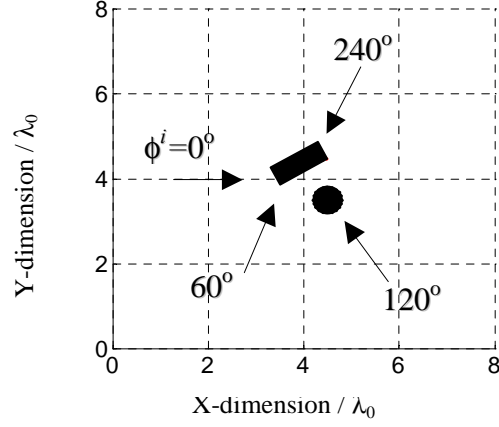


Figure 3a. Top view of the target (oblate spheroid) and the tree root (horizontal cylinder).

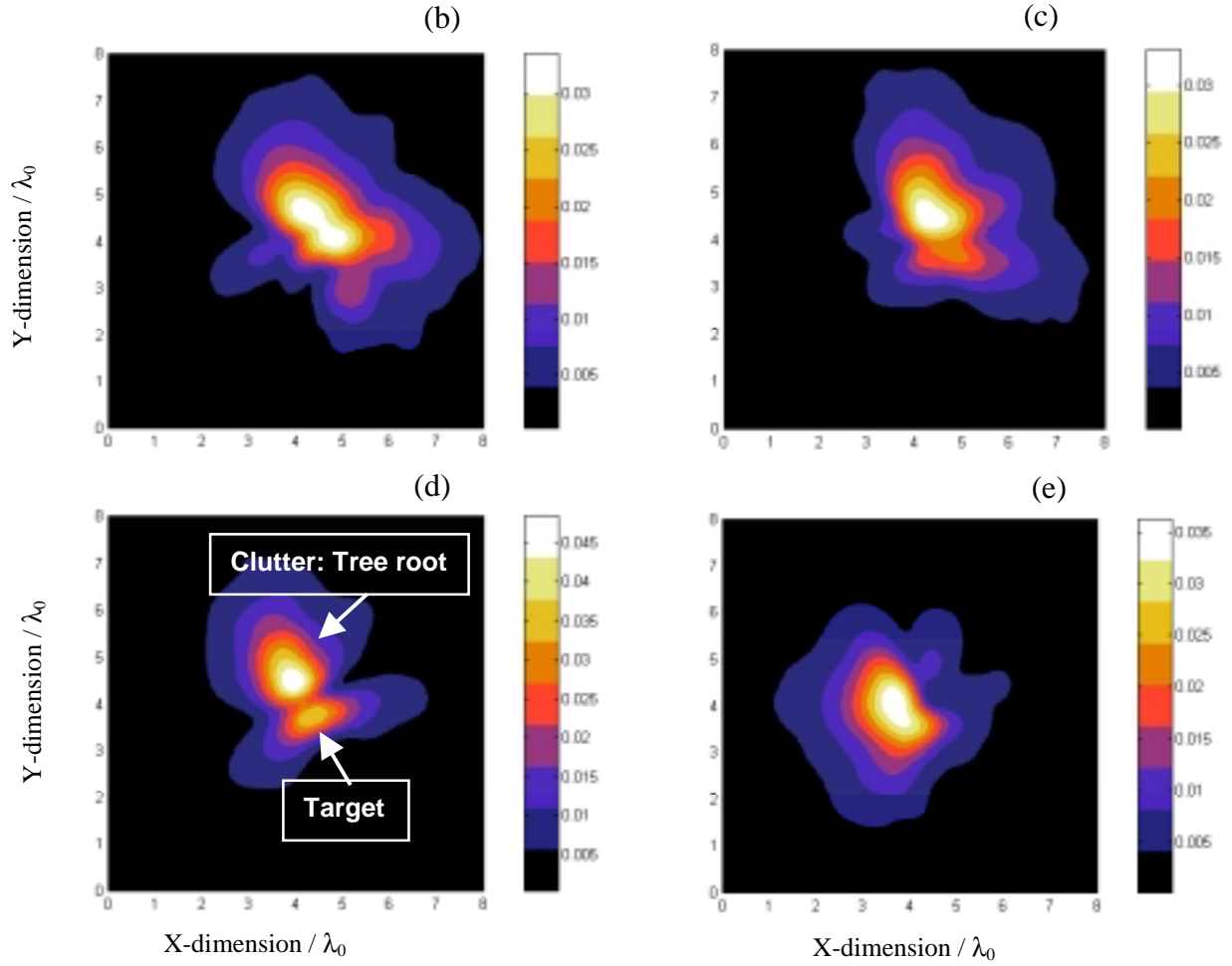


Figure 3. The scattered electric field at multiple views for both the buried mine-like target and the tree root upon removing the scattering from ground when: (b) $\phi^i = 0^\circ$, (c) $\phi^i = 60^\circ$, (d) $\phi^i = 120^\circ$, and (e) $\phi^i = 240^\circ$. All results are for $\sigma = 0.1\lambda_0$, $l_c = 0.5\lambda_0$, $\theta^i = 30^\circ$ and H-polarization.

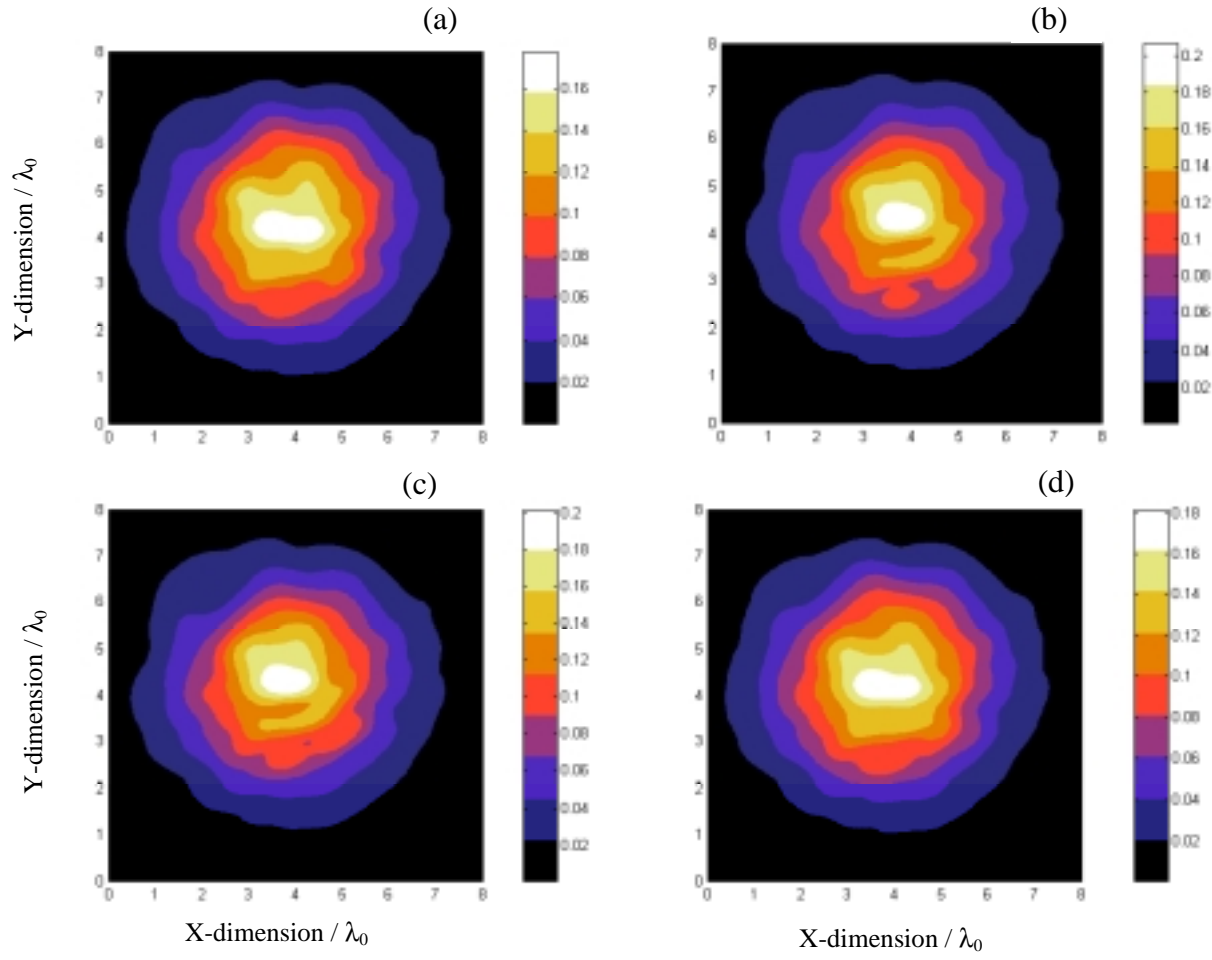


Figure 4. The average scattered electric field at $z = 0.5\lambda_0$ from (a) the ground with no buried objects, (b) the ground with both the buried mine-like target and tree root, (c) the ground with only the buried tree root (no mine), and (d) the ground with only the buried mine-like target (no tree root). Data are from Fig. 3c (H-polarization).

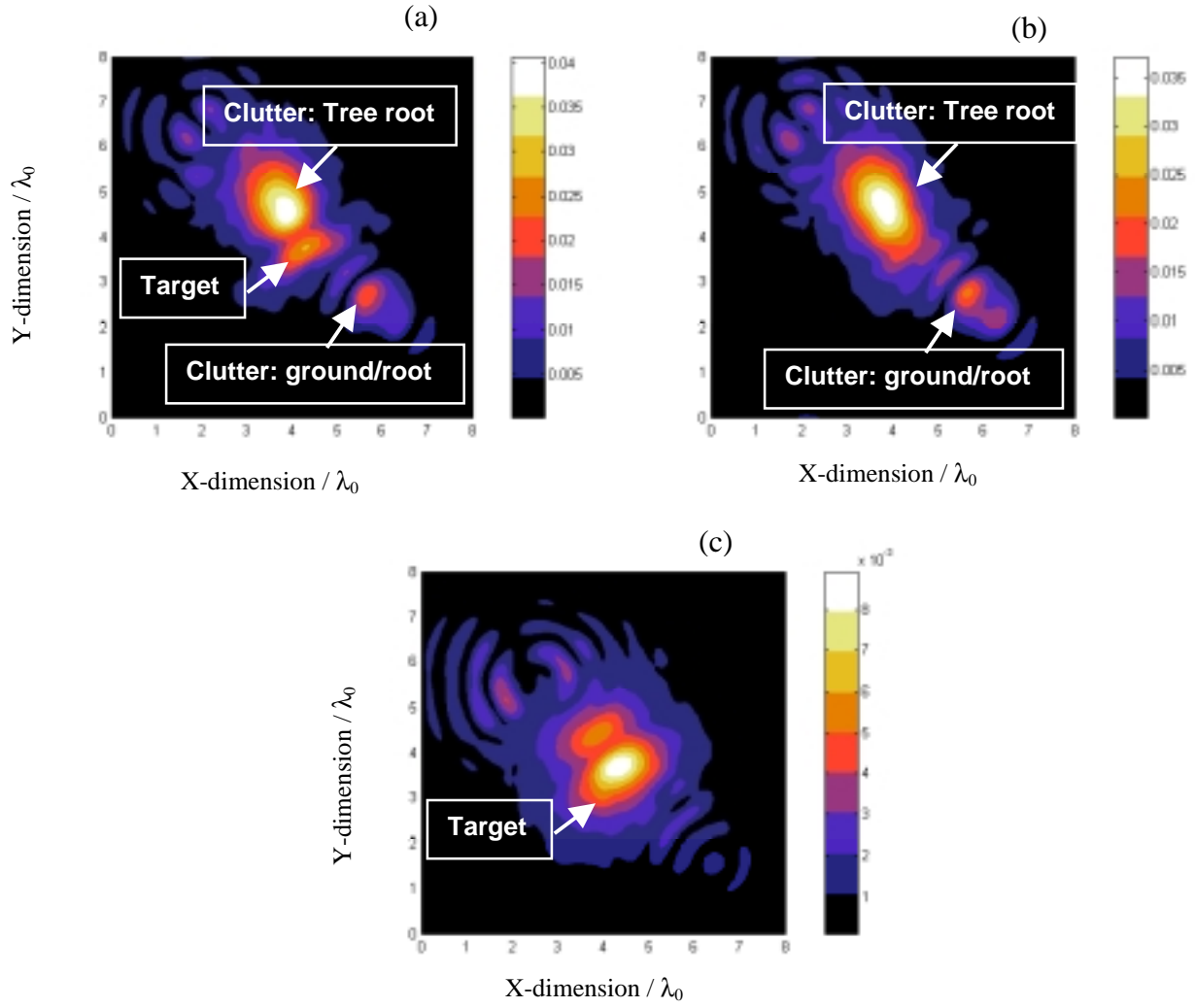


Figure 5. The average scattered electric field upon removing the scattering from the rough ground for (a) both the mine-like target and tree root, (b) just the tree root, (c) just the mine-like target. Data are from Fig. 3c. (H-polarization).

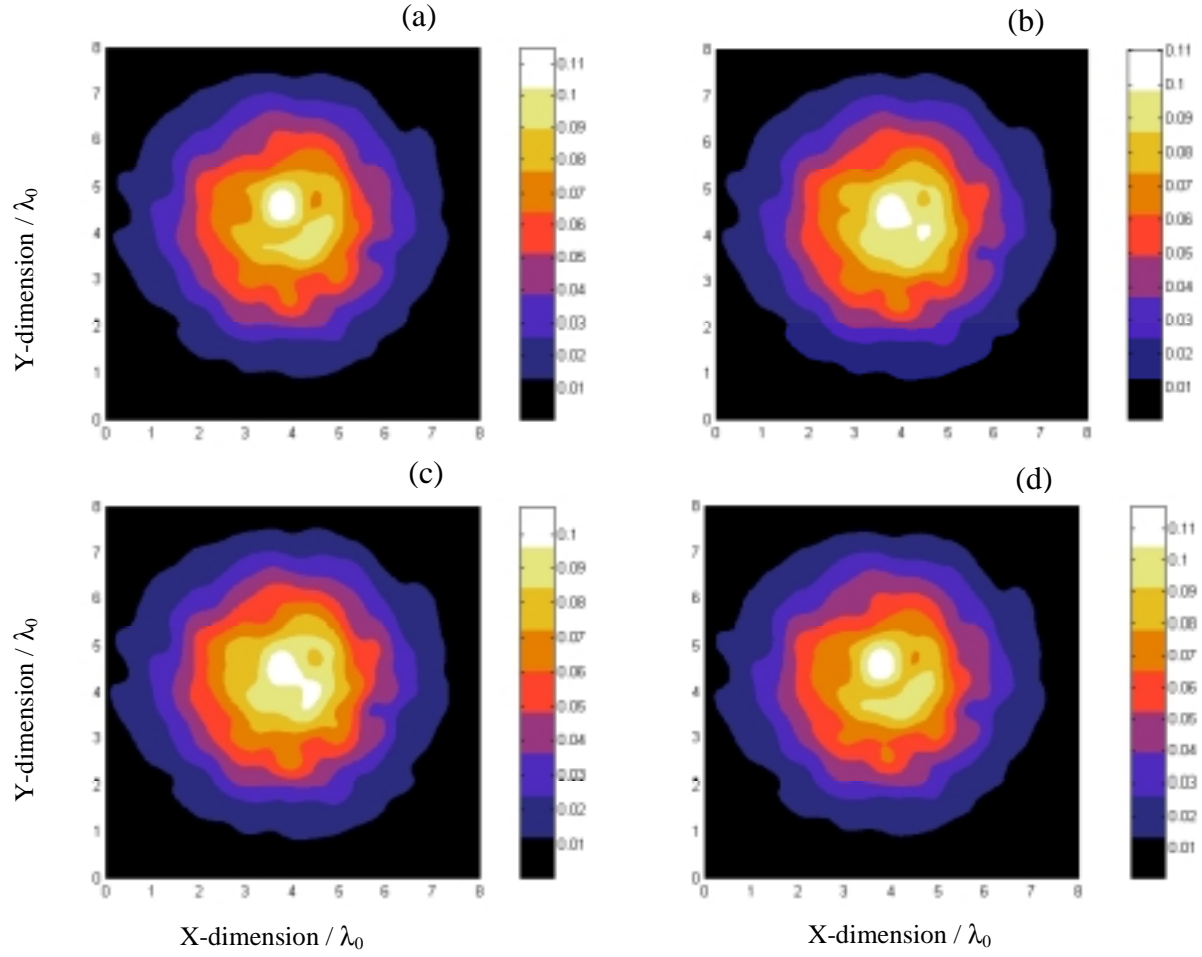


Figure 6. The average scattered electric field at $z = 0.5\lambda_0$ from (a) the ground with no buried objects, (b) the ground with both the buried mine-like target and tree root, (c) the ground with only the buried tree root (no mine), and (d) the ground with only the buried mine-like target (no tree root). Data are from Fig. 3c for V-polarization.

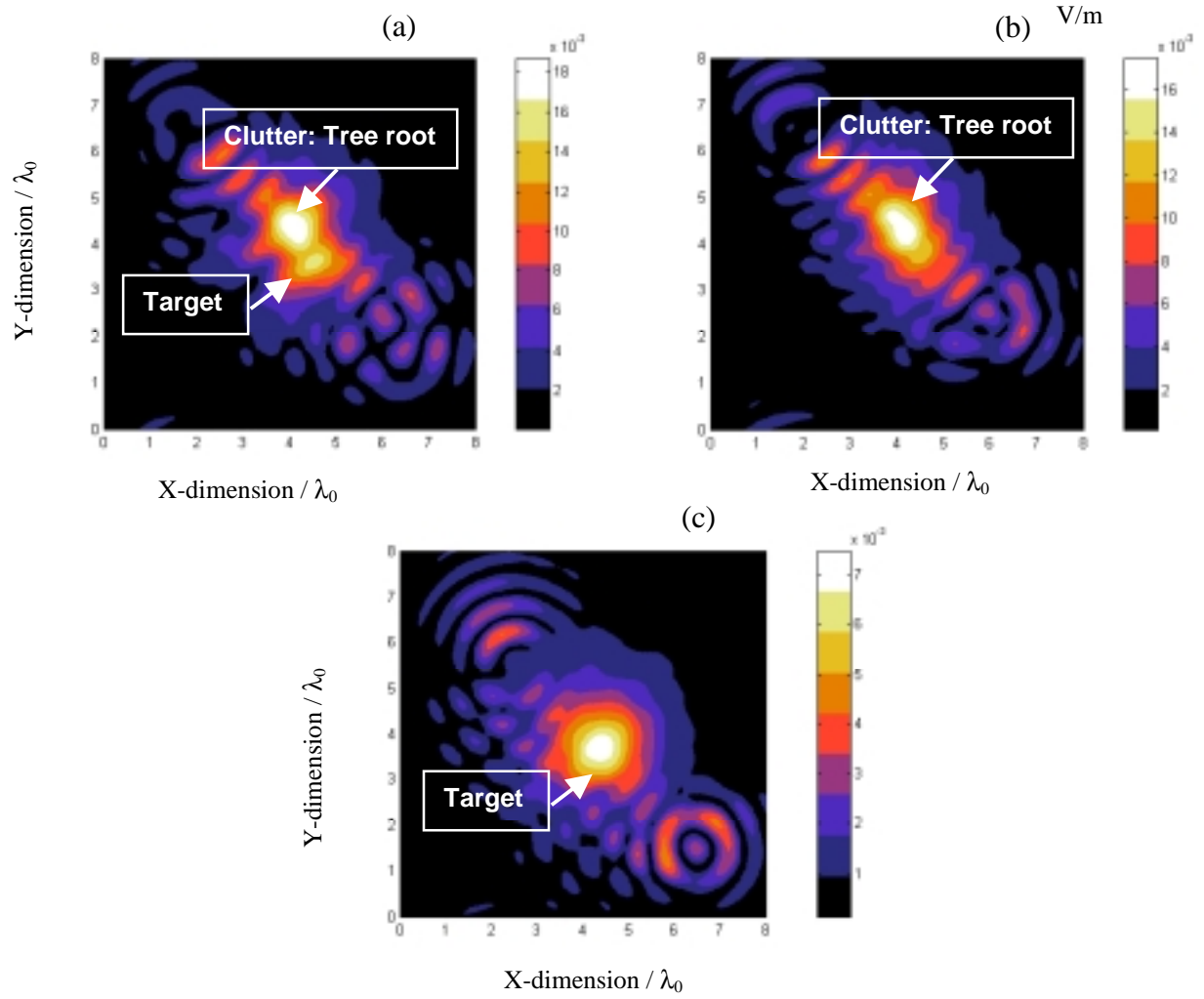


Figure 7. The average scattered electric field upon removing the scattering from the rough ground for (a) both the mine-like target and tree root, (b) just the tree root, (c) just the mine-like target. Data are from Fig. 3c but for V-polarization.

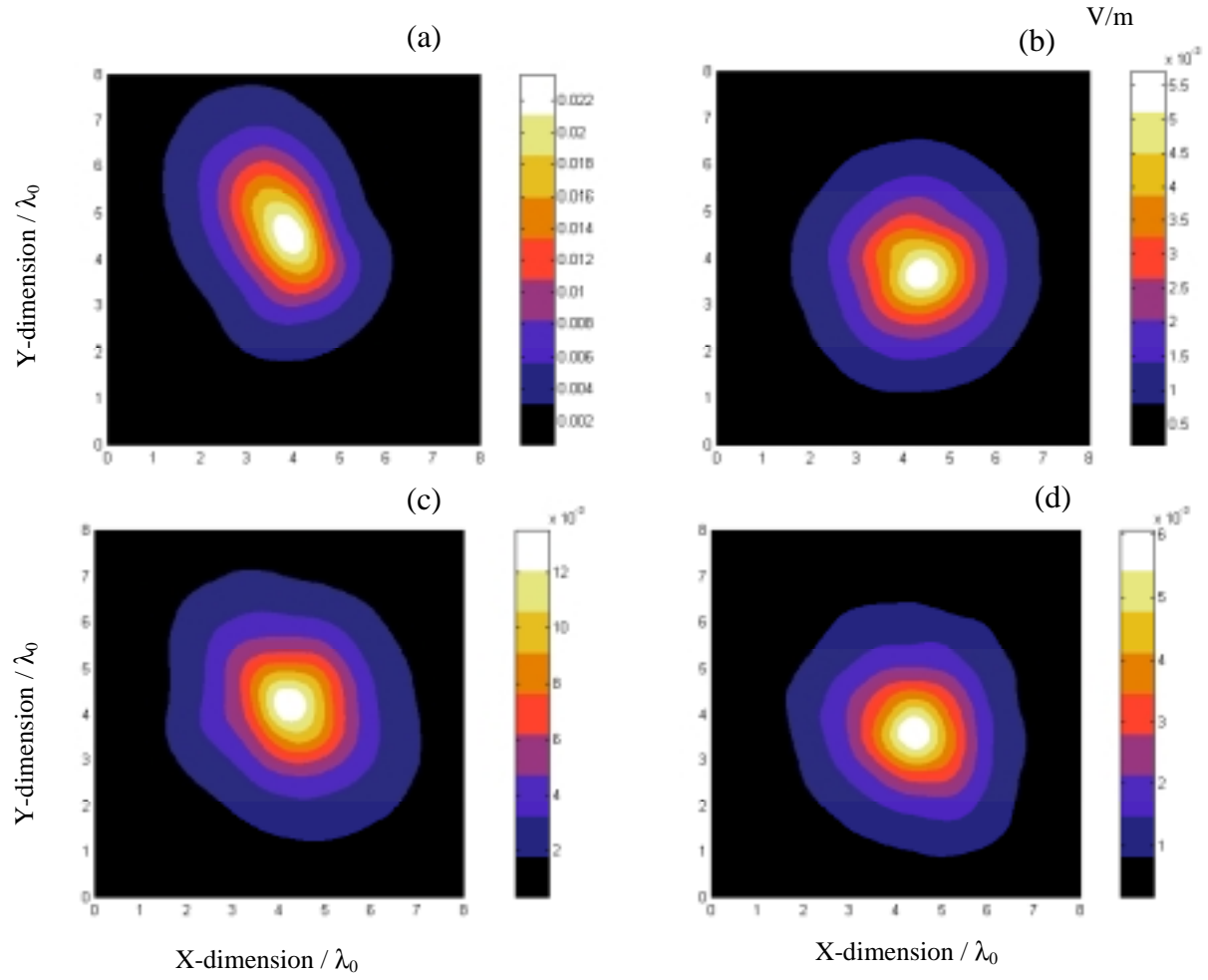


Figure 8. The standard deviation of the scattered electric fields for (a) just the tree root, (b) just the AP-mine-like target, for the H-polarization, (c) just the tree root, and (d) just the AP-mine-like target, for the V-polarization. Data are from Fig. 3c.

Using the PML ABC for Air/Soil Wave Interaction Modeling in the Time and Frequency Domains

Carey M. Rappaport, Scott Winton
Northeastern University, 235 Forsyth Building, Boston, MA 02115
email: RAPPAPORT@NEU.EDU

Abstract—The Perfectly Matched Layer (PML) absorbing boundary condition (ABC) is applied to two-dimensional FDTD and FDFD simulations of wave interactions with a two-layer air and soil geometry. Since the soil medium is lossy and dispersive, modeling of wave propagation and scattering is significantly more difficult than for free space and pure dielectrics. In addition, specifying an ABC which efficiently prevents reflections from both the free space termination and the adjacent soil is challenging. The theoretical basis for terminating dispersive media for both the time and frequency domains is presented, and simulation results for plane wave and point source excitations are demonstrated. For the former case, scattering is computed from a buried mine-like target. Although the absorbing characteristics of the PML for the air/soil interface are not as good as for free space, it is shown that maximum local reflections of as low as -15 dB for FDTD and -50 dB for FDFD are possible.

Keywords—FDTD, FDFD, PML, ABC, GPR, Ground Penetrating Radar, Computational Modeling.

I. INTRODUCTION

To understand the effects of the environment surrounding scattering targets, it is important to simulate numerically the propagation, interaction, and scattering of electromagnetic or acoustic waves by buried objects in the surrounding modelled clutter.

This work has been supported by The Army Research Office, Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013

The short pulse and stepped-frequency ground penetrating radar problems are open scattering geometries, with lossy, frequency dependent (temporally dispersive) media, requiring high performance absorbing boundary conditions [1-2]. Finite difference methods are very effective in modeling small-scale variations of environment, including: particular target shapes, rough surface boundaries, and volume inhomogeneities. The time domain version (FDTD) [3-5] is an efficient discretization of Maxwell's equations, which is robust and flexible and can conveniently deal with multiple inhomogeneities. The computation is particularly well-suited to supercomputer platforms, and it yields results that are suitable for animation and other high quality computer graphics capabilities. Memory requirements involve storage of the field values in each of millions of cells. For frequency-dependent media, such as soil, the convolutional relationship of permittivity and electric field requires storage of field values for several time cycles. The finite difference frequency domain method is similar to FDTD, but solves the self-consistent field distribution in open and closed spatial volumes for single frequency[6-8].

Berenger's Perfectly Matched Layer (PML) absorbing boundary condition (ABC) [9] has significantly increased the performance and popularity of the finite difference methods of electromagnetic field computation. While many analyses have considered applying the PML ABC to terminate layered media [10] and lossy dispersive media [11-13] computational grids, its application to layered dispersive media is less well-understood. In particular, the application to the common and important problem of a realistic soil half space, with sources in air above has seen little exposure in the literature. In the subsequent sections, this paper discusses layered, dispersive, lossy PML theory for the frequency and time domains, and then presents numerical examples to confirm this theory.

II. FREQUENCY DOMAIN PML THEORY FOR GENERAL LAYERED MEDIA

In the frequency domain, the equations for a two-dimensional transverse electric (relative to the longitudinal direction z into the paper) PML region terminating a lattice at $x = x_{\max}$ have the form [14]:

$$\frac{\partial H_z}{\partial y} = j\omega\epsilon_s E_x \quad (1a)$$

$$-\frac{\partial H_z}{\partial x} \left(\frac{1}{1 - j\sigma_e(x)/\omega\epsilon_0} \right) = j\omega\epsilon_s E_y \quad (1b)$$

$$\frac{\partial E_y}{\partial x} \left(\frac{1}{1 - j\sigma_m(x)/\omega\epsilon_0} \right) - \frac{\partial E_x}{\partial y} = -j\omega\mu_0 H_z \quad (1c)$$

where the medium (soil) permeability is assumed to be that of free space, and its complex permittivity is:

$$\epsilon_s = \epsilon_0(\epsilon'_s - j\sigma_s/\omega\epsilon_0) \quad (2)$$

In general, ϵ'_s and σ_s are functions of frequency. The electric and magnetic PML conductivities $\sigma_e(x)$ and $\sigma_m(x)$ govern the rate of wave attenuation in the absorbing layers. For the continuous case in which the PML medium is impedance matched to the scattering space it terminates, $\sigma_e(x) = \sigma_m(x)$ and each increases with distance $x - x_{\max}$ from the scattering space. However, for finite difference computation, in which the E-fields and H-fields are calculated on interlocking grids, these conductivities are alternately sampled values of the continuously increasing function. For a typical eight grid cell PML layer, the conductivity profile is specified as a power function of position with the form [15]:

$$\sigma_{i-\frac{1}{2}} = \left(\frac{2i-1}{16} \right)^p \frac{S_{\max}}{\Delta}, \quad i = 1, \dots, 8 \quad (3a)$$

$$\sigma_i = \left(\frac{2i}{16} \right)^p \frac{S_{\max}}{\Delta}, \quad i = 1, \dots, 8 \quad (3b)$$

for a PML beginning at $i = 0$. Note that magnetic conductivity $\sigma_m((i - \frac{1}{2})\Delta)$ is associated with $\sigma_{i-\frac{1}{2}}$ values if the PML begins on an electric field point. For a PML terminating a free-space scattering region, quasi-exhaustive search studies in both the time and frequency domains have shown that the best values of p and S_{\max} are 3.1 and 0.021, respectively [6, 15]. With this choice, the maximum reflected signal for typical time pulse signals for incident angles from normal through 60° is -100 dB.

Viewing the PML as an anisotropic mapping of the normal coordinate (x , in the case above), it is clear

that regardless of the material characteristics of the scattering medium ϵ_s , the profile of PML conductivities is unchanged from those of free-space, Eqns. (3). While the actual conductivities of the individual grid cells within the PML will incorporate both the background conductivity of the soil and the PML conductivities, as long as the frequency domain TE curl equations are written and evaluated as in Eqns. (1), there is no need consider these actual values.

If instead of separate first-order curl equations, the FDFD computation makes use of Helmholtz equation – by substituting Eqns. (1a) and (1b) into (2) – the same approach applies:

$$\frac{1}{k^{\text{rel}}(x)} \frac{\partial}{\partial x} \frac{1}{k^{\text{rel}}(x)} \frac{\partial H_z}{\partial x} + \frac{\partial^2 H_z}{\partial y^2} + k_s^2 H_z = 0 \quad (4)$$

where $k^{\text{rel}}(x) = 1 - j\sigma_m(x)/\omega\epsilon_0$, for the continuous PML case. The complete scattering medium material characteristics are included in the wavenumber $k_s = \omega\sqrt{\mu_0\epsilon_s}$.

To terminate a parallel layered region in the frequency domain with a PML ABC, all that is required is to specify the free-space PML conductivity profile as the termination for each layer. Again using the mapped coordinate interpretation, the common transverse wavenumber along the air/soil interface (that is, normal to the PML), k_x maps to $k_x k^{\text{rel}}(x)$ in the PML which terminates both layers. The wavenumber normal to the air/soil interface, $k_y(x)$, for any of the i^{th} PML sublayers is the same as in the scattering space. As a result, the reflection and transmission properties across the air/soil interface are the same within the PML as they are in the scattering region, with only a relative wave attenuation normal to the PML boundary (and along the air/soil interface) applied to waves in all layers.

III. TIME DOMAIN PML THEORY FOR GENERAL LAYERED MEDIA

The relative elegance of the layered PML in the previous section is lost in conversion to the time domain. This happens because the material parameters of each cell in the PML must be specified explicitly. The background σ_s and ϵ'_s must be combined with the σ_e and σ_m of the PML for each layer.

The time domain version of Eqns. (1) is readily available in the literature in several forms, but the

most common splits the H-field into a part associated with the first term of Eqn (1c), H_{zx} , and another with the second term of Eqn (1c), H_{zy} . Thus, in preparation for converting to the time domain, Eqn (1c) splits into two equations:

$$-\frac{\partial E_y}{\partial x} = \mu_0 \left[j\omega + \frac{\sigma_m}{\epsilon_0} \right] H_{zx} \quad (5a)$$

$$\frac{\partial E_x}{\partial y} = j\omega\mu_0 H_{zy} \quad (5b)$$

Using Eqn. (2), Eqn. (1b) becomes:

$$-\frac{\partial H_z}{\partial x} = \epsilon_0 \left[j\omega\epsilon'_s + \frac{\sigma_s}{\epsilon_0} + \frac{\epsilon'_s\sigma_e}{\epsilon_0} - j\frac{\sigma_s\sigma_e}{\omega\epsilon_0^2} \right] E_y \quad (6)$$

Transforming Eqns. (1a, 6, 5a, 5b) yields the set:

$$\frac{\partial H_z}{\partial y} = \epsilon_0\epsilon'_s \left[\frac{\partial E_x}{\partial t} + \frac{E_x}{\tau_s} \right] \quad (7a)$$

$$-\frac{\partial H_z}{\partial x} = \epsilon_0\epsilon'_s \left[\frac{\partial E_y}{\partial t} + \left(\frac{1}{\tau_s} + \frac{1}{\tau_e} \right) E_y + \frac{\int E_y dt}{\tau_s\tau_e} \right] \quad (7b)$$

$$-\frac{\partial E_y}{\partial x} = \mu_0 \left[\frac{\partial H_{zx}}{\partial t} + \frac{H_{zx}}{\tau_m} \right] \quad (7c)$$

$$\frac{\partial E_x}{\partial y} = \mu_0 \frac{\partial H_{zy}}{\partial t} \quad (7d)$$

with $\tau_{(e)} = \epsilon_0/\sigma_{(e)}$, $\tau_s = \epsilon'_s\epsilon_0/\sigma_s$, and $H_z = H_{zx} + H_{zy}$. A similar set exists for the TM polarization.

If the PML terminates a scattering medium with frequency-dependent lossy permittivity, a separate dispersive FDTD model must be used. Recursive convolution [16], differential equation [17-19], and Z-transform methods [20] can be used for this purpose. The conventional Debye and Lorentz dispersion models attempt to match both the real and imaginary parts of the complex material permittivity at once. Although these model parameters exist for many soils, once the conductivity of the PML layers are included — as in Eqn (5a) — new model parameters must be determined for each value in the PML. A less complicated and more automatic approach is to use a dispersion model which matches conductivity separately, keeping the real part of the complex permittivity constant with frequency. Excellent agreement with this type of model for soils is possible [21]. In this case, the first two terms on the right hand side of Eqn (6) are not changed in the presence of PML conductivity, the third term is

merely a constant scaling of the profiled PML conductivity, and the fourth term is small enough to be ignored for typical soils and most of the conductivities in the PML profile. Thus, for the dispersive conductivity/constant dielectric constant model, the integration term in Eqn. (7b) is neglected, and the dispersive model for the scattering space is modified by the addition of simple frequency-independent constants.

The choice of conductivity profile in the PML at the air/soil interface follows the same argument in the time domain as previously for the FDFD case. The σ_e and σ_m values have the form given by Eqns. (3). Many trials have suggested that a simple parabolic profile with $p = 2$ and $S_{\max} = 0.0063$, gives the best absorption performance for dispersive layered soil media, so the *effective* conductivities of the progressive PML sub-layers (starting on an E-field point) are given by:

$$\sigma_m = \sigma_{s,i-\frac{1}{2}} = \frac{0.0063\epsilon'_s}{\Delta} \left(\frac{2i-1}{16} \right)^2, \quad i = 1, \dots, 8 \quad (8a)$$

$$\sigma_e = \sigma_{s,i} = \sigma_s(\omega) + \frac{0.0063\epsilon'_s}{\Delta} \left(\frac{2i}{16} \right)^2, \quad i = 1, \dots, 8 \quad (8b)$$

That is, the conventional PML is modified when terminating the lossy dispersive soil region by replacing the real permittivity ϵ_0 with that of the soil $\epsilon'_s\epsilon_0$, and using a conductivity profile with electric and magnetic conductivities given by Eqns. (8).

IV. NUMERICAL RESULTS

Numerical experiments were conducted to validate the theory presented above. Two scattering cases, representative of typical ground penetrating radar applications, were considered: a point-source above a lossy, uniform half-space of soil; and a plane wave normally incident on a soil half-space with a landmine-like rectangular target buried 4.5 cm below the interface. In both cases the soil parameters used are those of Puerto Rican clay loam, measured at 10% moisture with density 1.4 g/cc [22]. At a typical GPR center frequency of 960 MHz, this soil has dielectric constant $\epsilon' = 6.2$ and electrical conductivity $\sigma = 0.035$ S/m. The target has the dielectric properties of TNT: $\epsilon' = 2.9$, loss tangent 0.001, corresponding to $\sigma = 0.00015$ S/m at 960 MHz.

The FDFD modeling was performed in two dimensions, with field sampled in $1/20$ soil wavelength increments ($\Delta = 6.3$ mm). First, a point source, located at 14 points above the ground surface and 14 points to the left of the center of the scattering space (to examine symmetry), was simulated in a 61 by 41 point grid surrounded by a free-space PML on exterior boundaries of the upper half-plane, and a lossy medium PML for the lower half plane. The total, real part of the electric field H_z is displayed in Figure 1. Clearly, the left/right symmetry, relative to the source point indicates that there are very few numerical artifacts from the left and right boundaries. To further verify the reflectionless nature of the combined PML, a scattering region with twice the horizontal and vertical dimensions was examined. The resulting calculation of real electric field, Figure 2, includes the previous simulation as a subset centered about its origin. Since the agreement between the two simulations is very close over their common region, it is reasonable to conclude that the absorbing boundary surrounding the smaller region is not contributing significantly to the calculated field, even when this ABC is only 6 grid points (0.12λ) from the source point. Figure 3 presents explicitly the magnitude of the difference between the electric fields calculated with the two different size grids. The differences are normalized to the calculated electric field at the upper right corner (27,20). It is clear that the worst error occurs in this corner, where the wave from the source encounters the upper boundary with 82° grazing incidence angle. Even so, this error represents a reflection artifact of 0.17%, or about -55 dB.

The second case modelled by FDFD, is representative of target scattering by a buried 7.6cm by 4.4cm rectangular non-metallic mine-like target. For this calculation, first the unperturbed normally incident plane wave H_{unp} is determined analytically, using standard reflection/ transmission theory, with transmission coefficient given by [23]:

$$T = \frac{2}{1 + \sqrt{\epsilon' - j\sigma/\omega\epsilon_0}} \quad (9)$$

for free space permeability. Thus, for normal incidence,

$$H_{\text{unp}} = TE_0 \sqrt{\frac{\epsilon_s}{\mu_0}} e^{-jk_s(y_0-y)} \quad (10)$$

with E-field magnitude E_0 at the ground surface at $y = y_0$. The total field, $H_t = H_s + H_{\text{unp}}$ satisfies the Helmholtz equation:

$$[\nabla^2 + k^2(x, y)] H_t = 0 \quad (11)$$

for space-dependent wavenumber $k(x, y)$. Since the unperturbed magnetic field in the soil satisfies the Helmholtz equation with the constant scatterer-free wavenumber k_s , Eqn (11) can be separated into an inhomogeneous Helmholtz equation as:

$$[\nabla^2 + k^2(x, y)] H_s = -[k^2(x, y) - k_s^2] H_{\text{unp}} \quad (12)$$

where the right hand side is non-zero only over the support of buried target.

Using the combined PML, applied just to the scattered field, completes the specification of the FDFD calculation. The field scattered from the rectangular target buried in soil, added to the unperturbed reflected and transmitted plane wave was again computed across a smaller and a larger grid. The real part of this total field for the 61 by 41 grid is shown in Figure 4. The target occupies horizontal grid points $-18\Delta < x < -6\Delta$, and vertical points $-6\Delta < y < -13\Delta$. Once again, the field is symmetric with respect to the symmetric rectangular scatterer. The computation over the larger grid is presented in Figure 5, and the magnitude of the difference between the two computations divided by the computed field value where the largest error occurs at (-12,-20) is shown in Figure 6. The largest error is even lower than for the point source illumination case above, 0.018% (-75 dB), occurring in the lower left bottom boundary at the point closest to the target.

For time domain calculations, the PML performance is considerably worse. For the point source case, the excitation was chosen to be a gaussian pulse, modulated at 960 MHz, at the same position as for the FDFD example: (-14,14). The spatial sampling is also the same $\Delta = 6.3\text{mm}$, and the time step is selected to be well within the FDTD Courant stability condition to be $\Delta t = 10$ ps. The soil model parameters are those of Rappaport, *et. al.* [21]:

$$D(t) = \epsilon_{\text{AV}} E(t) \quad (13)$$

$$\sigma(Z) = \frac{b_0 + b_1 Z^{-1} + b_2 Z^{-2}}{1 + a_1 Z^{-1}} \quad (14)$$

where $Z = \exp(j\omega\Delta t)$, corresponding to a unit time delay in FDTD calculations. Parameter value which

match the measured values of Puerto Rican clay loam to within 10% from 30 MHz to 3840 MHz are: $\epsilon_{Av} = 1.171$, $b_0 = 4.386$, $b_1 = -8.637$, $b_2 = 4.251$, and $a_1 = -0.975$. Using Eqn. (14), the electric current density is related to the electric field in the time domain by:

$$J^n + a_1 J^{n-1} = b_0 E^n + b_1 E^{n-1} + b_2 E^{n-2} \quad (15)$$

where the superscript denote time step. Eqn. (15) can be used along with Eqn. (13) in the discretized Ampere's law for dispersive FDTD calculations [20].

For the point source case, at the time just after the time pulse peak has entirely penetrated the ground surface, the H-field pattern is fairly symmetric with respect to the line of symmetry passing through the source point, as shown in Figure 7. The error is calculated by subtracting the field values from this 61 by 41 cell grid from those calculated on a much bigger grid (361 by 341 cells, centered at the same point as the 61 by 41 cell grid). This difference, divided by the field value at the point of maximum error (-30,-18) is displayed in Figure 8. The largest error is 15% (-16 dB). Although this error is quite large, most of the reflection comes from the wall terminating the soil rather than from the air soil interface.

The buried rectangular TNT mine case was also examined using FDTD. Figure 9 shows the total H-field for the same parameters given above. the left/right symmetry is better than for the point source FDTD case. Again error is calculated by comparison with the larger grid. Figure 10 shows the difference between the sample grid calculation and the much larger reference calculation normalized to the value of H-field at the point of largest error (28,-40). The largest magnitude errors are at the corners corresponding to deepest soil locations, with a value of 20% (-14 dB).

IV. CONCLUSIONS

The theory of PML grid terminations for both FDFD and FDTD for an air/dispersive soil half space has been developed and analyzed. The frequency domain results are considerably better than the FDTD results, with local reflection errors of -55 dB compared to -14 dB respectively for an 8 layer parabolically profiled PML. The need to match the PML attenuation profile in air with that of soil at the

air/soil boundary decreases the FDTD PML absorption performance compared to that of a uniform scattering volume. Further work to determine the best PML conductivity for the air/soil media problem is well indicated. For many practical FDTD simulations, reasonable results are possible by enlarging the grid enough to prevent side and back grid termination reflections from reinteracting with the target region.

V. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the code development and assistance of Misha Kilmer, and Ann Morgenthaler for many helpful discussions.

VI. REFERENCES

- [1.] B. Engquist and A. Majda, "Absorbing Boundary Conditions for the Numerical Simulation of Waves", *Mathematical Computation*, vol. 31, pp. 629–651, 1977.
- [2.] G. Mur, "Absorbing Boundary Conditions for the Finite-Difference Approximation of the Time-Domain Electromagnetic Field Equations", *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-23, pp. 377–382, November 1981.
- [3.] K. Yee, "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media", *IEEE Transactions on Antennas and Propagation*, vol. AP-14, pp. 302–307, 1966.
- [4.] A. Taflove and M. Brodwin, "Numerical solution of steady state electromagnetic scattering problems using the time-dependent maxwell's equations", "IEEE Transactions on Microwave Theory Technology", vol. 23, pp. 623–630, August 1975.
- [5.] A. Taflove and K. Umashankar, "The Finite-Difference Time-Domain (FD-TD) Method for Electromagnetic Scattering and Interaction Problems", *Journal of Electromagnetic Waves and Applications*, vol. 1, pp. 243–267, 1987.
- [6.] M. Kilmer, E. Miller, and C. Rappaport, "Preconditioners for structured matrices arising in subsurface object detection", *SIAM Journal on Scientific Computation*, in review for publication
- [7.] E. Marengo and C. Rappaport and E. Miller, "Optimum PML ABC Conductivity Profile in FDFD", *IEEE Transactions on Magnetics*, May 1999. vol. 35, pp. 1506–1509,
- [8.] C. Rappaport, M. Kilmer, and E. Miller, "Accuracy Considerations in Using the PML ABC with FDFD Helmholtz Equation Computation", in review for publication "International Journal of Numerical Modeling", 1999.
- [9.] J. Berenger, "A Perfectly Matched Layer for the Absorption of Electromagnetic Waves", *Journal of Computational Physics*, vol. 114, No. 1, pp. 185–200, October 1994.
- [10.] L. Zhao and A. Cangellaris, "GT-PML: Generalized Theory of Perfectly Matched Layers and Its Application to the Reflectiveness Truncation of Finite Difference Time-Domain Grids", *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, pp. 2555–2563, 1996.
- [11.] S. Gedney, "An Anisotropic PML Absorbing Media for the FDTD Simulation of Fields in Lossy and Dispersive Media", *Electromagnetics*, vol. 16, pp. 399–415, 1996.
- [12.] T. Uno, Y. He, and S. Adachi, "Perfectly Matched Layer Absorbing Boundary Condition for Dispersive Medium", *IEEE Transactions on Microwave Theory and Techniques*, vol. 7, pp. 264–266, September 1997.
- [13.] Y. Chen, W. Chew, and M. Oristaglio, "Application of Perfectly Matched Layers to the Transient Modeling of Subsurface EM Problems", *Geophysics*, vol. 62, pp. 1730–1736, December 1997.
- [14.] C. Rappaport, "Interpreting and improving the PML absorbing boundary condition using anisotropic lossy mapping of space", "IEEE Transactions on Magnetics", vol. 32, pp. 968–974, 1996.
- [15.] S. Winton and C. Rappaport, "Profiling the Perfectly Matched Layer to Improve Large Angle Performance", *IEEE Transactions on Antennas and Propagation*, in review
- [16.] R. Luebbers, F. Hunsberger, K. Kunz, R. Standler and M. Schneider, "A Frequency-Dependent Finite Difference Time Domain Formulation for Dispersive Materials", *IEEE Transactions on Electromagnetic Compatibility*, pp. 222–227, vol. 32, No. 3, March 1990.
- [17.] T. Kashiwa and I. Fukai, "A Treatment by the FD-TD Method for the Dispersive Characteristics Associated with Electronic Polarization", *Microwave and Guided Wave Letters*, vol. 16, pp. 203–205, June 1990.
- [18.] O. Gandhi, "A Frequency-Dependent Finite Difference Time Domain Formulation for General Dispersive Media", *IEEE Transactions on Microwave Theory and Techniques*, vol. 41, pp. 658–665, April 1993.
- [19.] F. Teixeira, W. Chew, M. Straka, M. Oristaglio, and T. Wang, "Finite Difference Time Domain Simulation of Ground Penetrating Radar on Dispersive Inhomogeneous, and Conductive Soils", "IEEE Transactions on Geoscience and Remote Sensing", vol. 36, November 1998. pp. 1928–1936,
- [20.] W. Weedon, and C. Rappaport, "A General Method for FDTD Modeling of Wave Propagation in Arbitrary Frequency-Dispersive Media", *IEEE Transactions on Antennas and Propagation*, pp. 401–410, March 1997.
- [21.] C. Rappaport, S. Wu, and S. Winton, "FDTD Wave Propagation Modeling in Dispersive Soil Using a Single Pole Conductivity Model", "IEEE Transactions on Magnetics", vol. 35, pp. 1542–1545, May 1999.
- [22.] J. Hipp, "Soil Electromagnetic Parameters as Functions of Frequency, Soil Density, and Soil Moisture", *Proceedings of the IEEE*, vol. 62, pp. 98–103, January 1974..
- [23.] D. Staelin, A. Morgenthaler, and J. Kong, *Electromagnetic Waves*, Prentice-Hall, Englewood Cliffs: pp. 126, 1994.

V. FIGURE CAPTIONS

Figure 1 Real part of the magnetic field computed using the Finite Difference Frequency Domain method for a 960 MHz point source 8.82cm above a clay loam soil half space. This 61 by 41 point grid is terminated by an 8-layer combined air/soil PML.

Figure 2 Real part of the magnetic field computed using the FDFD method for the same physical configuration of Figure 1, except with a larger computational grid: 121 by 81 points, terminated by the same 8-layer combined air/soil PML.

Figure 3 Magnitude of the difference over the common space between the magnetic fields calculated by FDFD on the small grid of Figure 1 and the big grid of Figure 2, normalized to the field value at the point (27,20).

Figure 4 Real part of the total magnetic field computed using the FDFD method for a 960 MHz plane wave normally incident on a clay loam soil half space with a 7.6 by 4.4cm rectangular TNT target buried 3.15cm below the ground surface. This 61 by 41 point grid is terminated by the same PML as in Figure 1.

Figure 5 Real part of the total magnetic field computed using the FDFD method for the same physical configuration of Figure 4, except with a larger computational grid: 121 by 81 points, terminated by the same 8-layer combined air/soil PML.

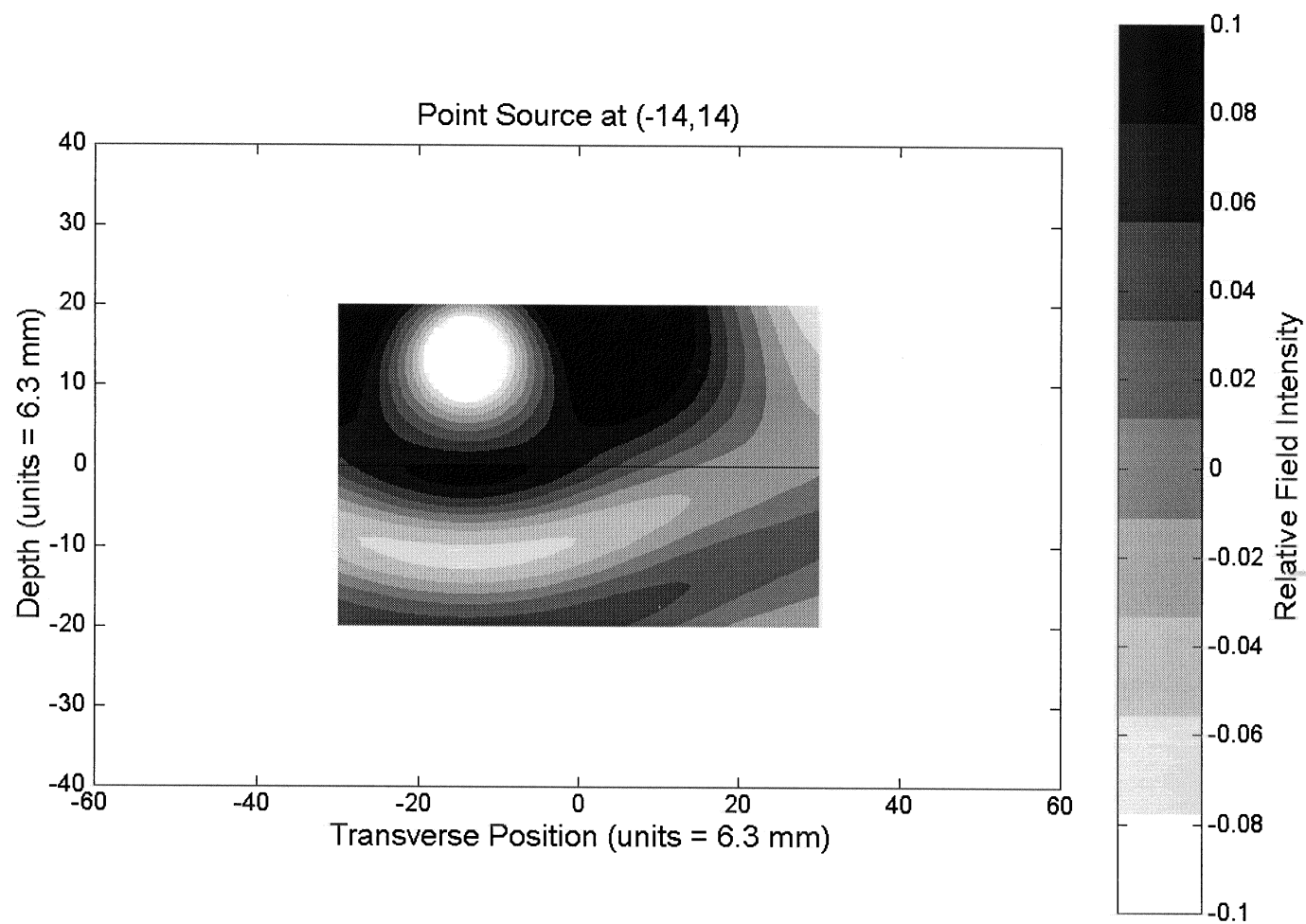
Figure 6 Magnitude of the difference over the common space between the magnetic fields calculated by FDFD on the small grid of Figure 4 and the big grid of Figure 5, normalized to the field value at the point (-12,-20).

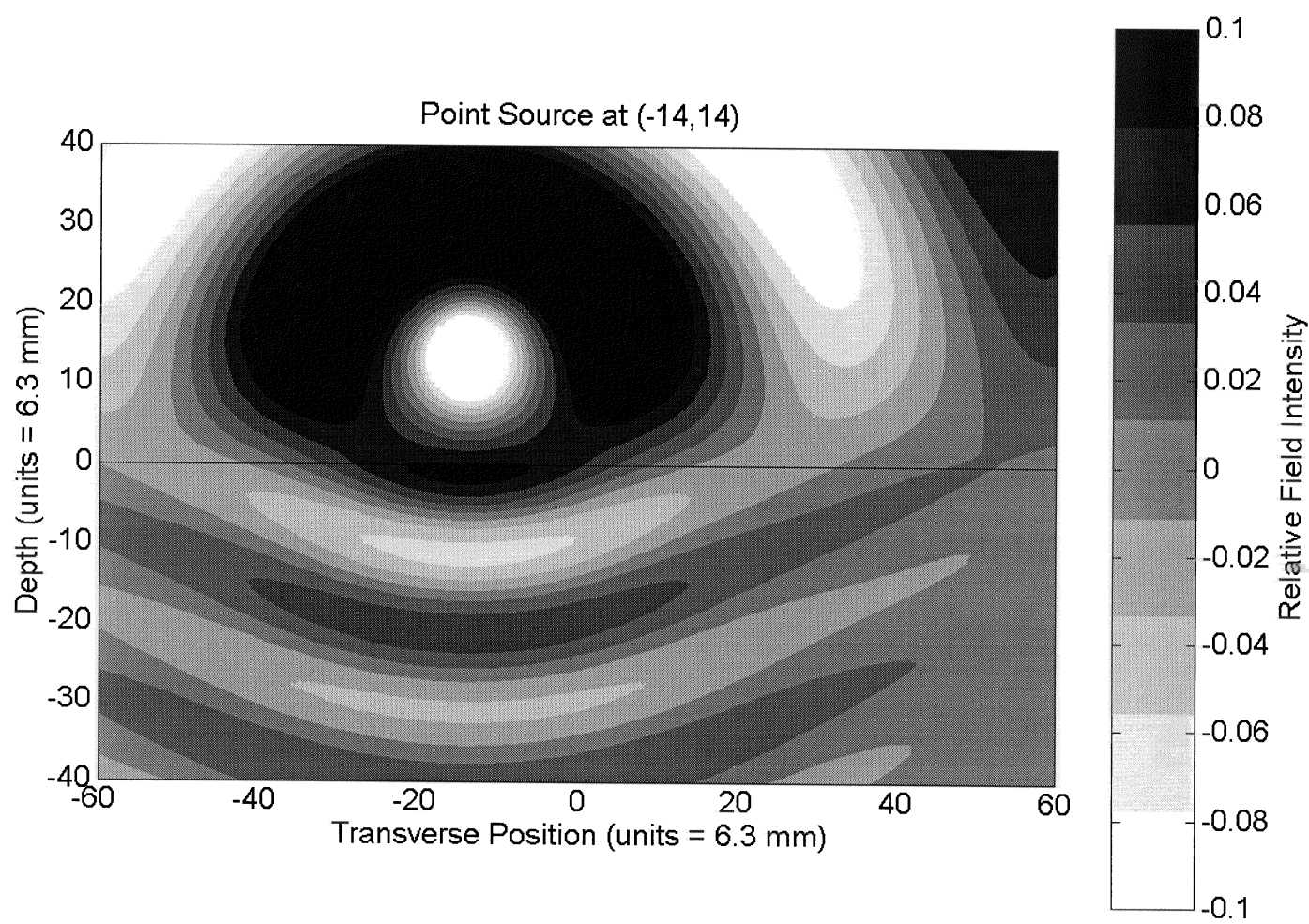
Figure 7 Real part of the magnetic field computed using the Finite Difference Time Domain method for a point source gaussian pulse modulated at 960 MHz, 8.82cm above a clay loam soil half space. This 61 by 41 point grid is terminated by an 8-layer combined air/soil PML with the same conductivity profile as for the FDFD simulations.

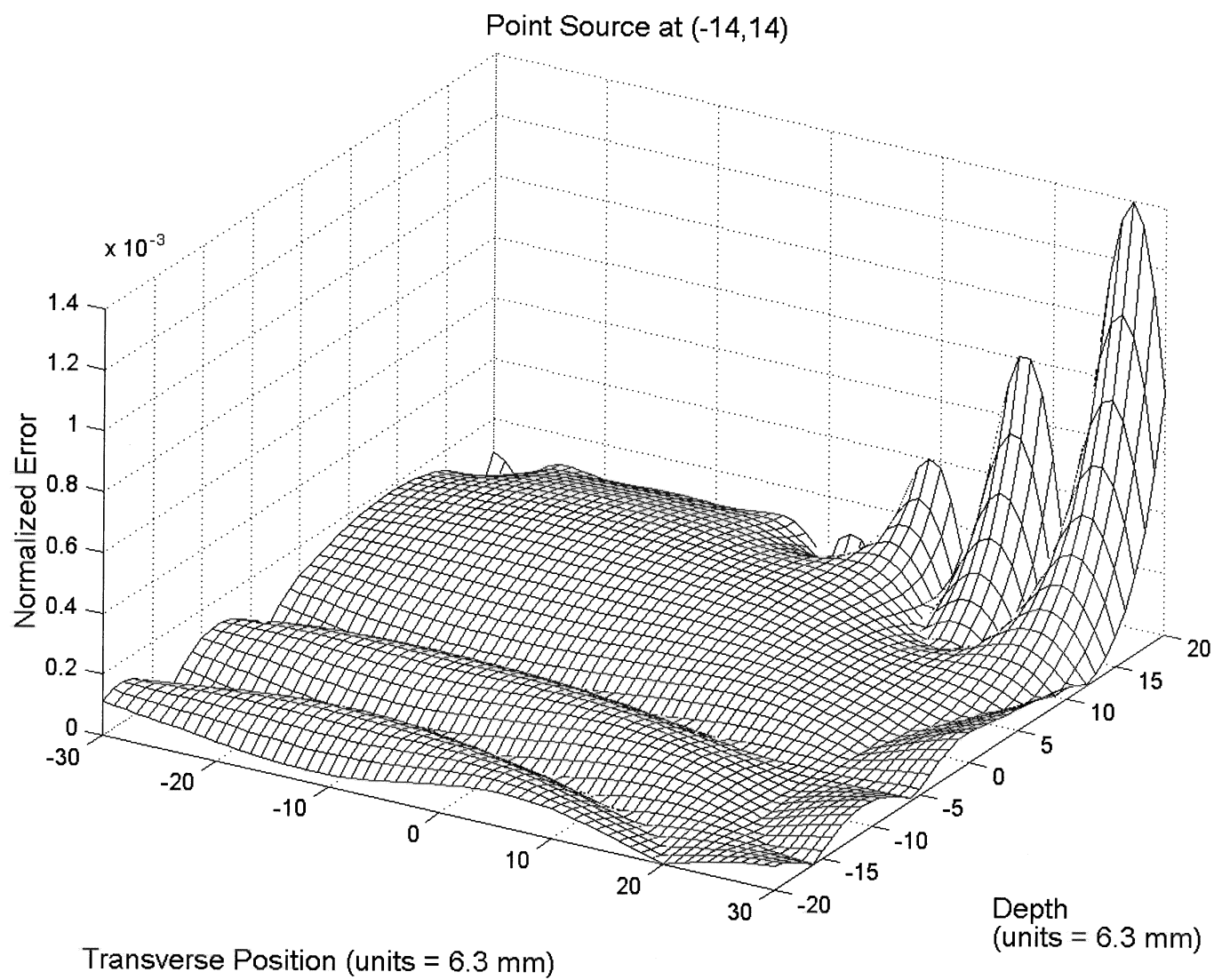
Figure 8 Magnitude of the difference between the magnetic fields calculated by FDTD on the grid of Figure 7 and a much bigger grid with extent 361 by 341, normalized to the field value at the point (-30,-18).

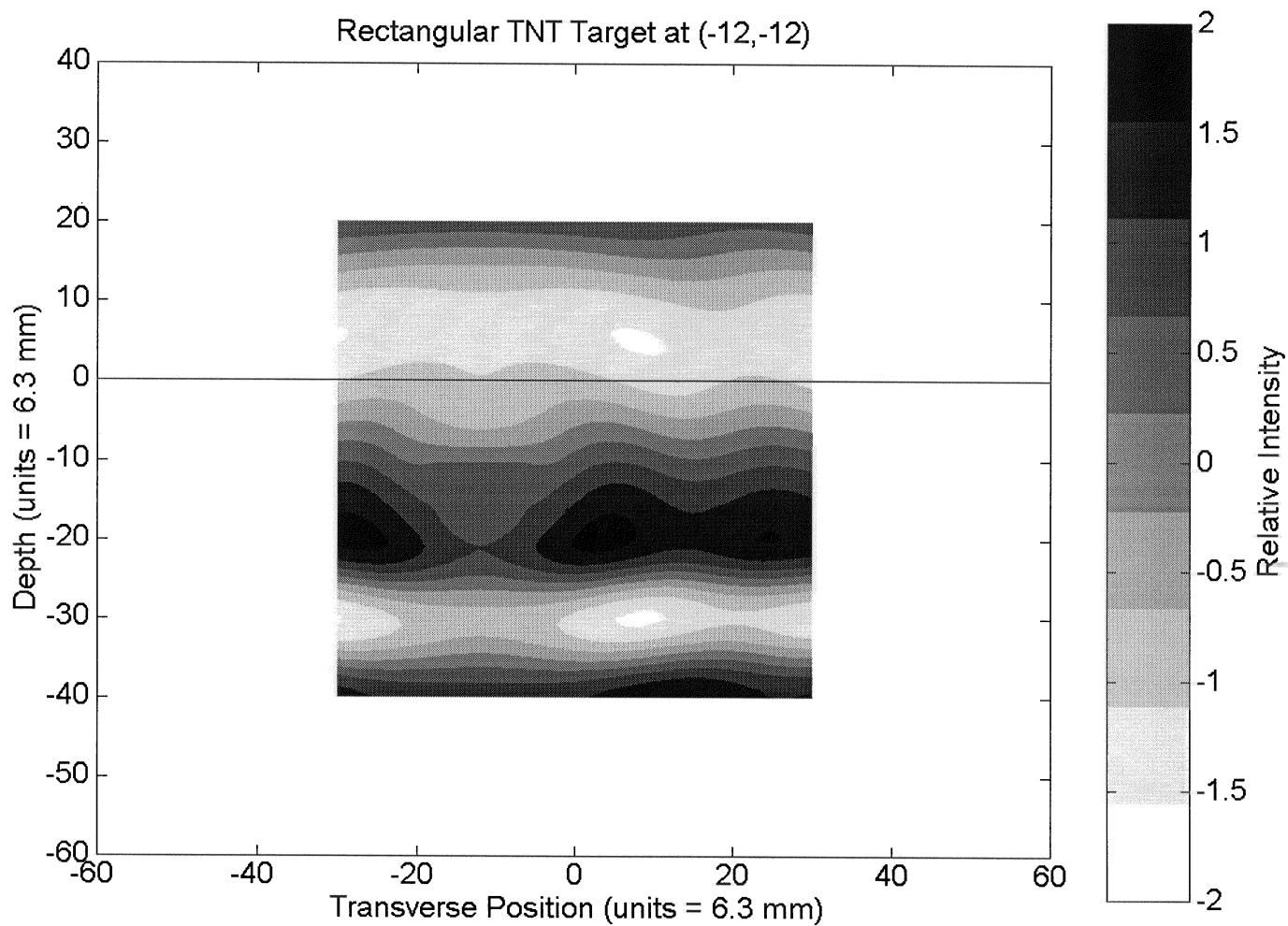
Figure 9 Real part of the magnetic field computed with FDTD for a plane wave gaussian pulse modulated at 960 MHz, normally incident on a clay loam soil half space, terminated by the same 8-layer combined air/soil PML.

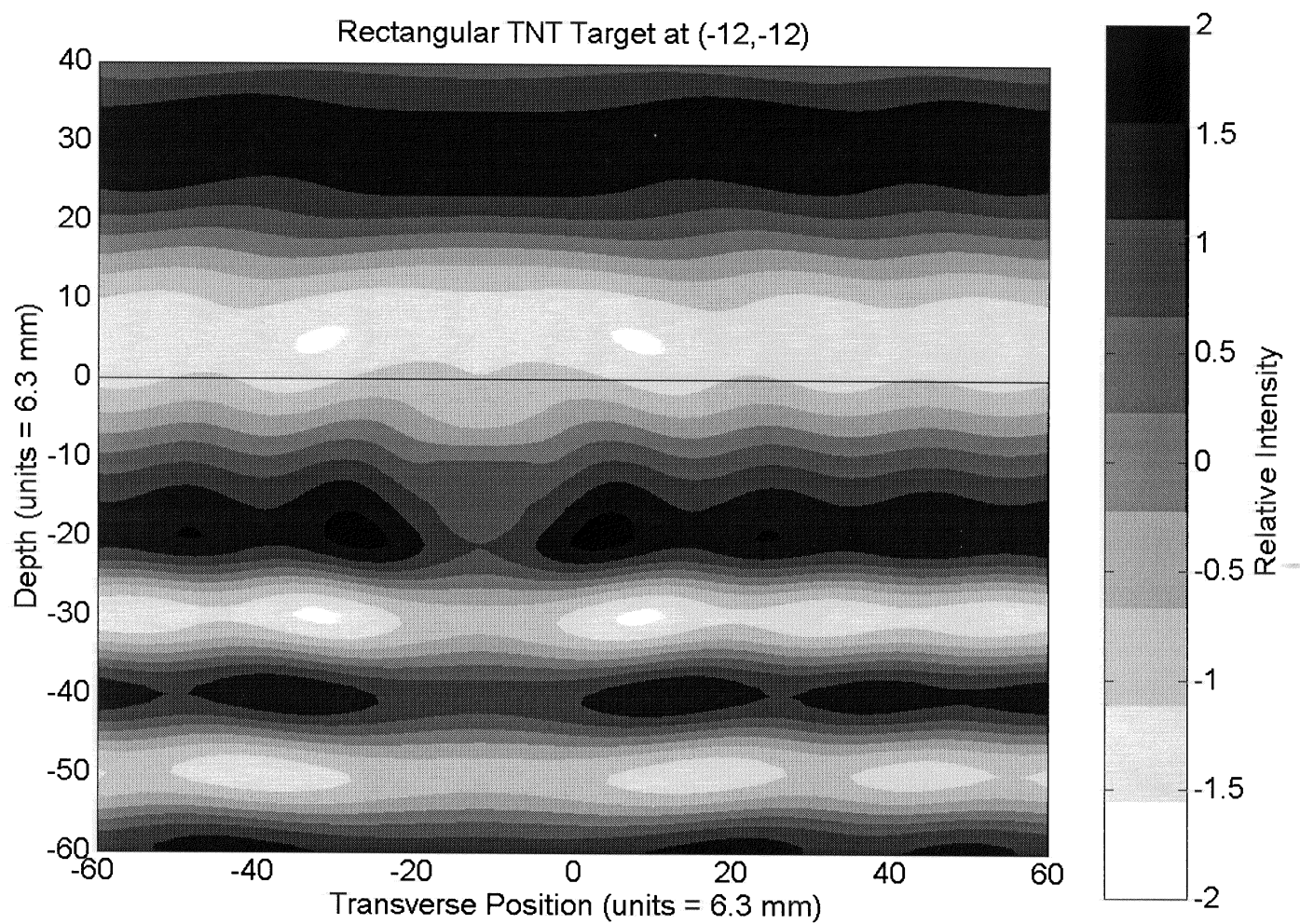
Figure 10 Magnitude of the difference between the magnetic fields calculated by FDTD on the grid of Figure 9 and a much bigger grid with extent 361 by 341, normalized to the field value at the point (28,-40).



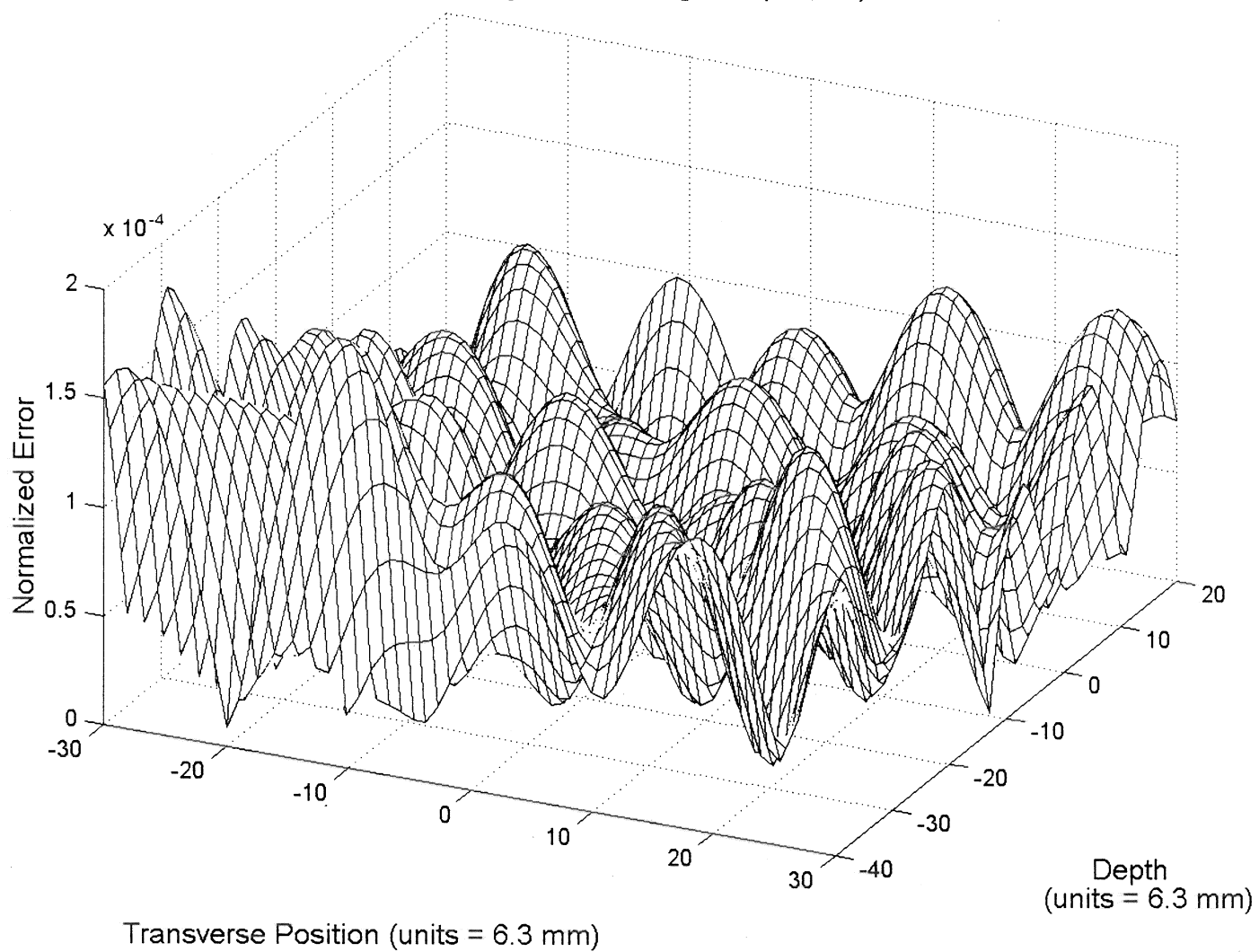


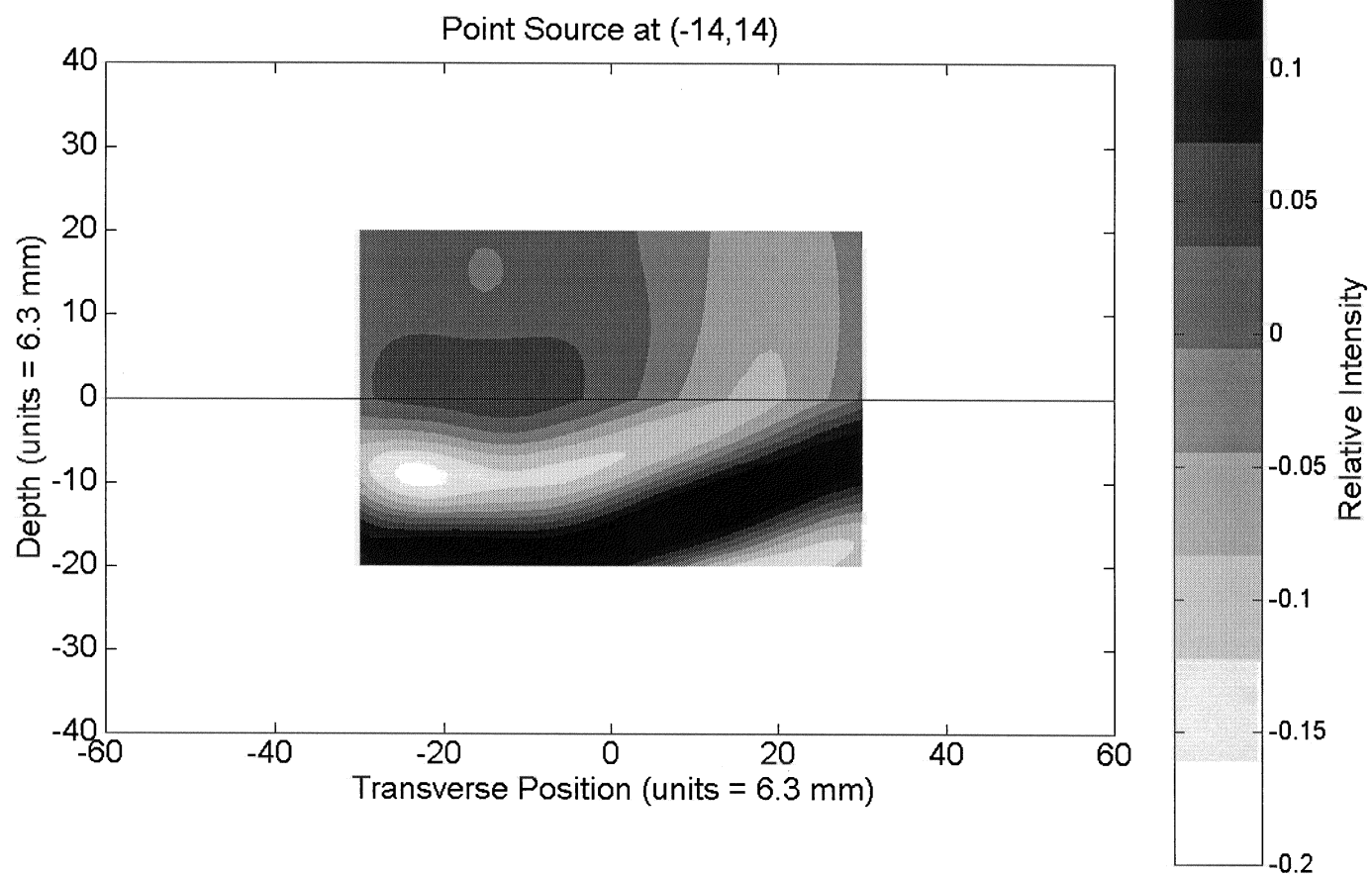


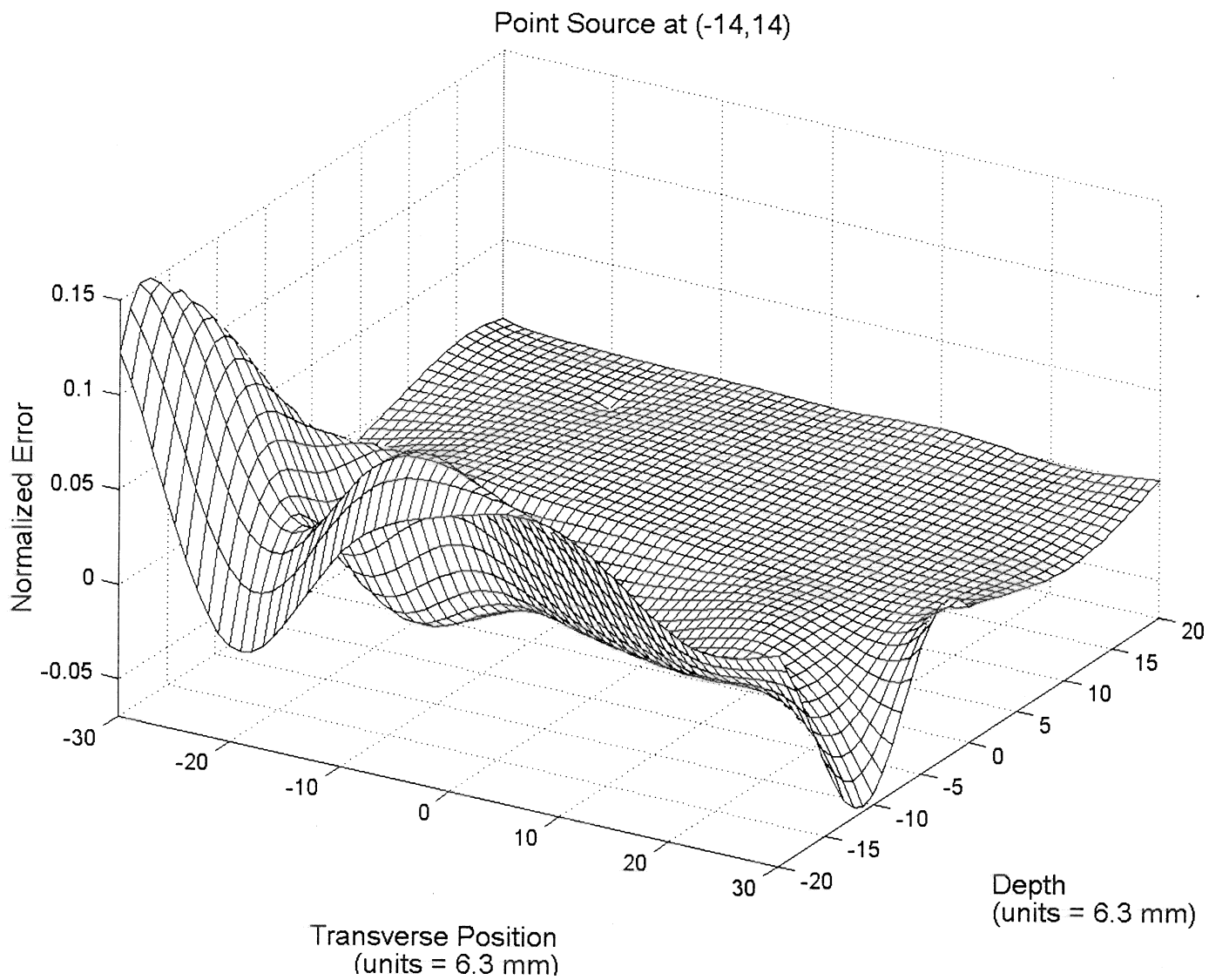


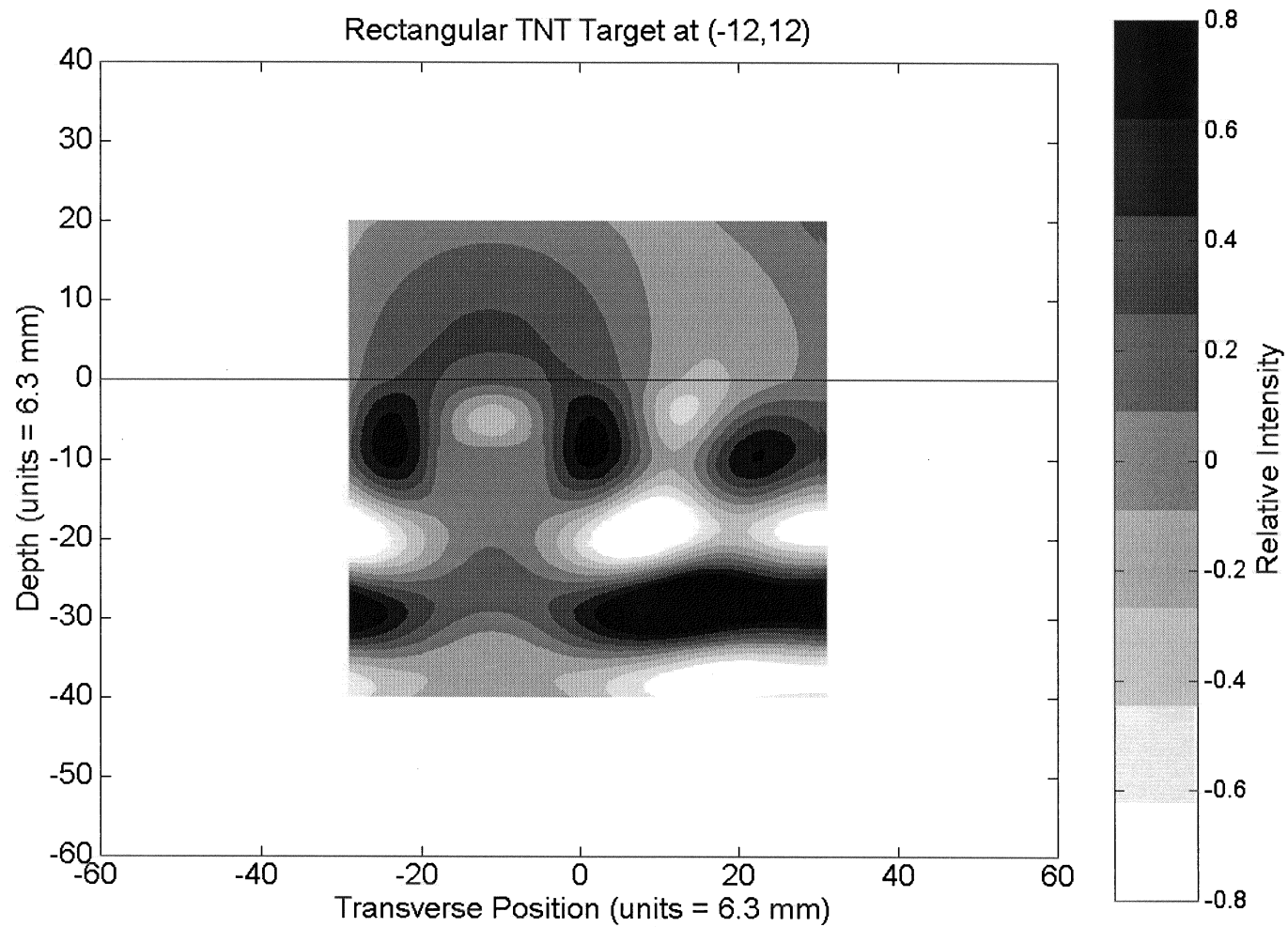


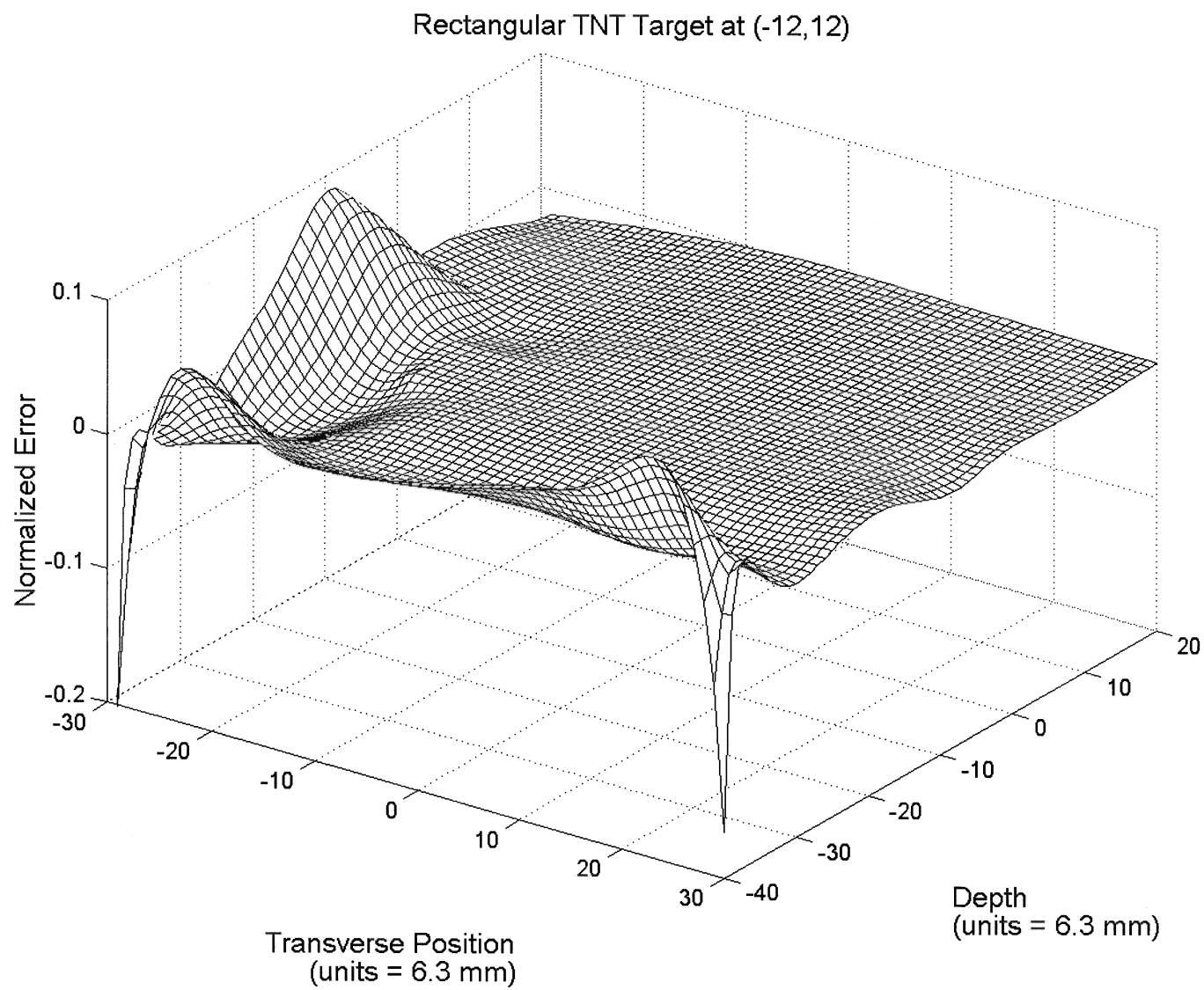
Rectangular TNT Target at (-12,-12)











Analysis and Characterization of Ultrawide-Band Scalar Volume Sources and the Fields They Radiate: Part II—Square Pulse Excitation

Edwin A. Marengo, *Member, IEEE*, Anthony J. Devaney, *Member, IEEE*, and Ehud Heyman, *Senior Member, IEEE*

Abstract—In a previous paper [1], we studied transient radiation from scalar collimated volume source distributions subjected to impulsive excitation. In this paper, we extend our analysis and results to the case of nonimpulsive excitation, paying special attention to the parameterization of the radiation pattern of three-dimensional (3-D) ellipsoid source distributions driven by square pulses of finite duration. We study the role both of the source's space distribution and of the square pulse duration on the generation of well-collimated short-pulse fields. In particular, we explore the source's angular and range resolutions as functions of the transverse and longitudinal dimensions of the source (for a fixed source volume) and of the pulse duration.

Index Terms—Transient propagation.

I. INTRODUCTION

IN a previous paper [1] we investigated the radiation properties of scalar wave fields $U(\mathbf{r}, t)$ radiated by three-dimensional (3-D) sources to the scalar wave equation

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) U(\mathbf{r}, t) = -4\pi Q(\mathbf{r}, t) \quad (1)$$

paying special attention to sources $Q(\mathbf{r}, t)$ having the space-time separable form of a "traveling wave"

$$Q(\mathbf{r}, t) = q_0(\mathbf{r})G(t - \hat{\mathbf{z}} \cdot \mathbf{r}/c) \quad (2)$$

where, without loss of generality, the main beam direction was chosen to be the positive z axis ($\hat{\mathbf{z}}$) and the space distribution $q_0(\mathbf{r})$ was normalized so that

$$\int d\mathbf{r}' q_0(\mathbf{r}') = 1. \quad (3)$$

In order to compare the radiation performance for different source parameters we have considered the source's volume V as invariant.

The class of sources defined in (2) represents a distribution of nondispersive isotropic point radiators, all of which radiate the same time signature $G(t)$ but with space-dependent

strength $q_0(\mathbf{r})$. The source defined in (2) could represent a discrete collection of point radiators (e.g., a 3-D antenna array of isotropic radiating elements) or a spatially continuous distribution of such radiators (note that in the discrete realization, the spectral shaping of the far field is also affected by the interelement spacings). A special case of such distribution is a uniformly distributed source wherein all point radiators are driven with the same strength: specific examples of uniform parallelepiped source distributions have been considered in [1].

In the present paper, we explore uniform ellipsoidal source distributions whose axis coincide with the z axis in (2) for which closed-form expressions can be obtained for all radiation directions and for all observation times. Furthermore, since the ellipsoidal shape may be continuously changed from a prolate to an oblate spheroidal shape, the closed-form results explain the different radiation characteristics of both elongated, quasi-linear traveling wave source distributions and quasi-planar broadside pulsed distributions, as well as those of 3-D sources with comparable longitudinal and transverse dimensions. Finally, while the analysis in [1] has concentrated on impulsive sources [i.e., $G(t) = \delta(t)$ in (2)], which generate near the beam axis singular field terms that cannot be parameterized, the emphasis here is placed on finite pulses (specifically on square pulses).

Our interest in the source distributions considered above arises from the fact that they yield highly collimated pulse beam fields with high degree of angular and range resolutions when they are driven by very short pulses $G(t)$ [1]. Motivation for the source structure in (2) is also provided by the possibility of building 3-D pulsed antenna arrays composed of identical electrically small photoconducting elements [2]–[4]. These antennas, made of III–V compound semiconductor wafers, radiate microwave pulses when triggered by short duration (e.g., picosecond) optical pulses. A 3-D array of such photoconducting elements can thus be realized using fiber optics and optical splitters, the latter to control the excitation strength associated with each element as dictated by $q_0(\mathbf{r})$ in (2) (the excitation strength can also be controlled by the voltage bias applied to each element [2], [3]). The progressive time delay in (2) can be controlled either by a delay network or simply by taking advantage of the propagation delay of the exciting optical wavefront along the z axis [5]. In addition, the antenna arrays described above are expected to have reduced microwave interelement coupling due to the short periods of photoconductivity associated with each element [5].

Manuscript received September 25, 1996; revised October 3, 1997. This work was supported in part by the U.S. Air Force Office of Scientific Research, Bolling AFB, Washington, DC, under Grant F49620-93-1-0093, by the Center for Electromagnetics Research, Northeastern University, Boston, MA, and by the Israel Science Foundation under Grant 574/95.

E. A. Marengo and A. J. Devaney are with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA.

E. Heyman is with the Department of Electrical Engineering and Physical Electronics, Tel Aviv University, Tel Aviv, 69978 Israel.

Publisher Item Identifier S 0018-926X(98)01493-8.

The source model defined in (2) can thus be viewed as an approximation to a 3-D distribution of electrically small (and thereby quasi-isotropic) photoconducting elements excited by a space-dependent amplitude $q_0(\mathbf{r})$ and with progressive time delay $\hat{\mathbf{z}} \cdot \mathbf{r}/c$ along the main beam direction $\hat{\mathbf{z}}$.

This paper generalizes our work in [1] by considering transient radiation from volume source distributions excited by finite duration square pulses. Our goal is to explore the role both of the excitation pulse duration W and of the length-to-width ratio $\ell \equiv L/a$ (where L and a are, respectively, the longitudinal and transverse dimensions of the source distribution) in establishing a highly collimated pulsed field with a high degree of angular and range resolutions. The angular resolution will be quantified from the directional properties of the *time-domain radiation pattern* $F(\hat{\mathbf{r}}, t)$ characterized by the peak amplitude and energy patterns while the range resolution will be determined from the pulse duration of $F(\hat{\mathbf{r}}, t)$. Special emphasis will be placed on the far-field properties near the main beam direction.

The work reported here encompasses that of Harmuth [6] where the peak amplitude, energy, and slope patterns of a linear (one-dimensional) array composed of elements that radiate far-field square pulses are derived using simple geometrical arguments. Also, the Radon transform radiation integral, which is the starting point of the present analysis, is a special case of the theory in [7] (formulated for electromagnetic sources) that applies to any time-dependent source distribution [not necessarily of the form in (2)]. We would also like to mention that even though we restrict our attention to ellipsoidal volume source distributions, the formulation presented here is rather general and can be applied to any 3-D source configuration of the form in (2). In what follows, we adopt the notation used in [1].

II. REVIEW OF THE GENERAL THEORY

The time-domain radiation pattern $F(\hat{\mathbf{r}}, t)$ is obtained by asymptotically evaluating the radiation field $U(\mathbf{r}, t)$ as $r \rightarrow \infty$, thus yielding [7], [8]

$$U(r\hat{\mathbf{r}}, t) \sim \frac{1}{r} F(\hat{\mathbf{r}}, t - r/c). \quad (4)$$

For the special case of the source in (2), $F(\hat{\mathbf{r}}, t)$ reduces to [1]

$$F(\hat{\mathbf{r}}, t) = G(t) * F_0(\hat{\mathbf{r}}, t) \quad (5)$$

where $*$ is used to denote temporal convolution and where the *time-domain pulsed-radiation pattern* $F_0(\hat{\mathbf{r}}, t)$ (e.g., the source's impulse response) is given by [1]

$$F_0(\hat{\mathbf{r}}, t) = \frac{c}{\xi} \bar{q}_0(\hat{\xi}, s) \Big|_{s=-ct/\xi} \quad (6)$$

where $\xi = \hat{\xi} \cdot \hat{\mathbf{r}} = \hat{\mathbf{r}} \cdot \hat{\mathbf{z}}$ (hence, $\xi = 2 \sin(\theta/2)$) and $\bar{q}_0(\hat{\xi}, s)$ is the Radon transform of $q_0(\mathbf{r})$ evaluated at the plane $\mathbf{r} \cdot \hat{\xi} = s$. Thus, for fixed $\hat{\mathbf{r}}$, the time-domain pulsed-radiation pattern $F_0(\hat{\mathbf{r}}, t)$ is defined by the projection of $q_0(\mathbf{r})$ onto the line directed along the unit vector $\hat{\xi}$.

A schematization of (6) is depicted in Fig. 1, which shows the planes $\mathbf{r} \cdot \hat{\xi} = s$ that lie perpendicular to the unit vector

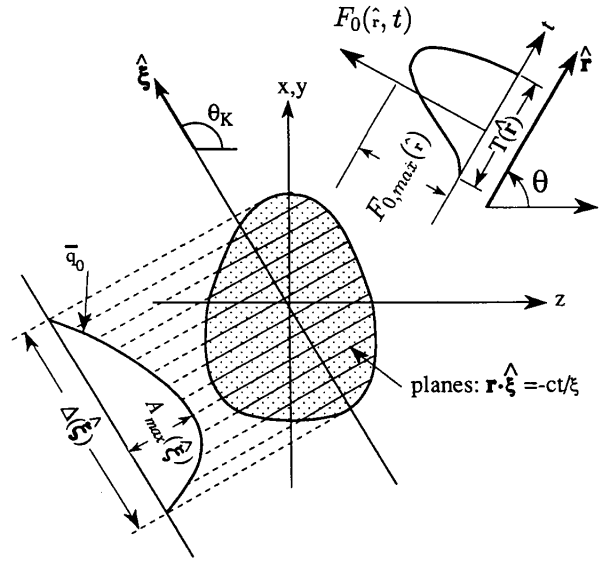


Fig. 1. Schematization of (6). $\bar{q}_0(\hat{\xi}, s)$ is the Radon transform along the $\hat{\xi}$ axis obtained by integrating $q_0(\mathbf{r})$ on surfaces orthogonal to $\hat{\xi}$. The time-domain pulsed-radiation pattern $F_0(\hat{\mathbf{r}}, t)$ is related to \bar{q}_0 via the scaling in (6).

$\hat{\xi}$ along which the projection is computed. $\bar{q}_0(\hat{\xi}, s)$ is shown in the left portion of the figure as the plot of the area of q_0 computed along the planes $\mathbf{r} \cdot \hat{\xi} = s$ with s being a parameter along the $\hat{\xi}$ axis. The time-domain pulsed-radiation pattern is shown in the right portion of the figure and is obtained from $\bar{q}_0(\hat{\xi}, s)$ by mapping $s \rightarrow -ct/\xi$ and scaling the amplitude by c/ξ . Referring to Fig. 1, we see that if the source dimension in the $\hat{\xi}$ direction is $\Delta(\hat{\xi})$, then the pulse duration of $F_0(\hat{\mathbf{r}}, t)$ along the observation direction $\hat{\mathbf{r}}$ is

$$T(\hat{\mathbf{r}}) = \xi \Delta(\hat{\xi})/c. \quad (7)$$

Similarly, the temporal peak amplitude—defined as the maximum amplitude (over all time t) of the time-domain pulsed-radiation pattern $F_0(\hat{\mathbf{r}}, t)$ in a given direction—is given by

$$F_{0,\max}(\hat{\mathbf{r}}) = c|A_{\max}(\hat{\xi})|/\xi \quad (8)$$

where $A_{\max}(\hat{\xi})$ is the maximum projection of the source's space distribution $q_0(\mathbf{r})$ onto the $\hat{\xi}$ axis.

III. TIME-DOMAIN PULSED-RADIATION PATTERN: ELLIPSOID DISTRIBUTIONS

We consider uniform volume-source distributions $q_0(\mathbf{r})$ with the shape of an ellipsoid that has rotational symmetry with respect to the z axis and with radial and axial dimensions a and L , respectively. To comply with the source normalization in (3), the source's magnitude is taken to be

$$q_0(\mathbf{r}) = \begin{cases} 1/V, & \frac{x^2+y^2}{a^2} + \frac{z^2}{L^2} \leq \frac{1}{4} \\ 0, & \text{else} \end{cases} \quad (9)$$

with $V = \frac{\pi}{6} a^2 L$ being the source's volume. Henceforth, we shall express the source dimensions a and L in terms of the volume V and the length-to-width ratio $\ell \equiv L/a$. Thus, $a = (6/\pi)^{1/3} V^{1/3} \ell^{-1/3}$ and $L = (6/\pi)^{1/3} V^{1/3} \ell^{2/3}$. By doing this we may now compare sources with the same volume

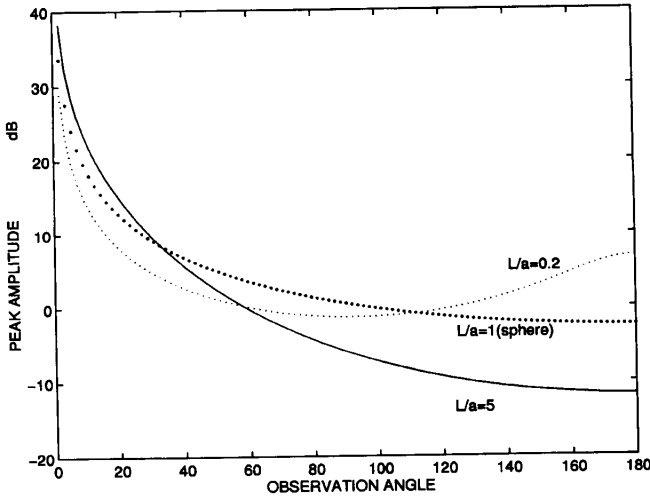


Fig. 2. Peak amplitude pattern of ellipsoidal source distributions under impulsive excitation $F_{0,\max}$ as function of θ (in degrees) for $\ell = 0.2, 1, 5$.

V (and thereby the same input energy¹) for different values of the parameter ℓ . Furthermore, when comparing elongated sources ($\ell > 1$) with flat sources ($\ell < 1$), we shall consider, in particular, sources with the same maximum dimension defined as the diameter of the smallest sphere completely enclosing the source. Therefore, we shall compare the results for an elongated source with, say, $\ell = \ell_e > 1$ with those of a flat source with $\ell = \ell_e^{-2}$.

Using the general Radon transform relation in (6), it is found in the Appendix that the time signature of $F_0(\hat{r}, t)$ for the ellipsoidally shaped distribution in (9) has, at a given observation direction \hat{r} with polar angle θ , the shape of a parabola

$$F_0(\theta, t) = \begin{cases} \frac{3}{2T(\theta)} - \frac{6}{[T(\theta)]^3} t^2 & |t| \leq \frac{1}{2}T(\theta) \\ 0 & \text{else} \end{cases} \quad (10)$$

where

$$T(\theta) = c^{-1}(6/\pi)^{1/3}\ell^{-1/3}V^{1/3}\sin\theta\sqrt{1+\ell^2\tan^2(\theta/2)} \quad (11)$$

is the time duration of $F_0(\theta, t)$. The peak amplitude $F_{0,\max}(\theta)$ is readily seen from (10) to be (see Appendix)

$$F_{0,\max}(\theta) = \max_t[|F_0(\theta, t)|] = \frac{3}{2T(\theta)}. \quad (12)$$

Following the discussion in [1], we define the energy radiation pattern

$$S(\hat{r}) = \pi \int_{-\infty}^{\infty} dt |F(\hat{r}, t)|^2 = \int_0^{\infty} d\omega |f(\hat{r}, \omega)|^2 \quad (13)$$

where $f(\hat{r}, \omega)$ is the frequency-domain radiation pattern, i.e., the temporal Fourier transform of $F(\hat{r}, t)$. Thus, the energy radiation pattern of ellipsoid distributions subjected to impulsive excitation is found to be

$$S(\theta) = \frac{6\pi}{5T(\theta)} \quad (14)$$

¹ Valid for sources of the form (2) that are uniformly excited.

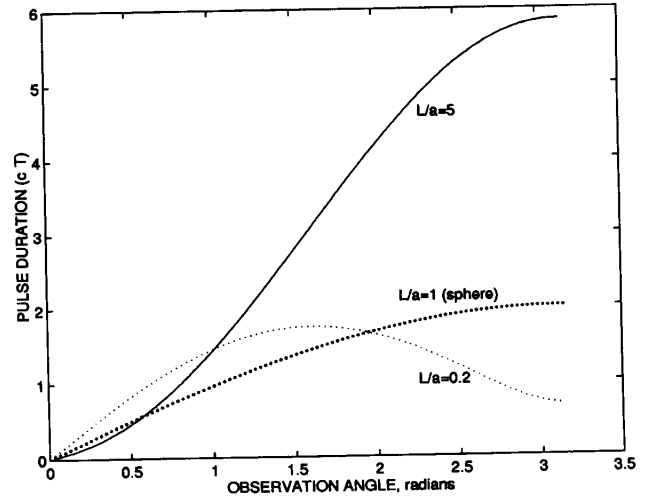


Fig. 3. Far-field pulse duration T of ellipsoidal source distributions under impulsive excitation versus θ (in radians) for $\ell = 0.2, 1, 5$.

hence, the peak amplitude radiation pattern $F_{0,\max}(\theta)$ and the energy radiation pattern $S(\theta)$ under impulsive excitation are identical within a multiplicative factor.

In the limit $\theta \rightarrow 0$ we obtain from (11)

$$T(\theta) \sim (6/\pi)^{1/3}V^{1/3}\ell^{-1/3}\theta/c \quad (15)$$

hence, $T(\theta)$ decreases monotonically with ℓ (for fixed $\theta \simeq 0$), while from (12), (14), (15), both $F_{0,\max}(\theta)$ and $S(\theta)$ increase monotonically with ℓ . Under impulsive excitation, elongated (large ℓ) sources thus produce far-field pulses of higher amplitude and shorter duration in the main-beam direction $\theta \simeq 0$.

Fig. 2 shows plots of the peak amplitude radiation pattern $F_{0,\max}(\theta)$ for ellipsoid source distributions with length-to-width ratios $\ell = 5, 1$, and 0.2 . These plots reveal that for a given $\theta \simeq 0$, higher radiated energy and peak amplitude (and thereby peak power) is available for larger ℓ (i.e., elongated distribution case). The presence of a secondary radiation lobe at $\theta = \pi$ for the small ℓ (planar source) case is also observed. Plots of the pulse duration $T(\theta)$ for the cases $\ell = 5, 1$, and 0.2 are shown in Fig. 3. We found that the θ dependence of the pulse duration is monotonic only for $\ell \leq 1$ and, more importantly, that the pulse duration for a fixed θ in the vicinity of the main-beam axis (i.e., $\theta \simeq 0$) is shorter for larger ℓ . For physical interpretation of the abovementioned effects, the interested reader is referred to [1].

IV. TIME-DOMAIN RADIATION PATTERN UNDER SQUARE-PULSE EXCITATION

In this section, we investigate the properties of the time-domain radiation pattern generated by scalar ellipsoid source distributions driven by unit amplitude square pulses having time duration W , i.e.,

$$G(t) = \begin{cases} 1, & 0 \leq t \leq W \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

We shall present the results in terms of the normalized pulse duration

$$\bar{W} = cW/V^{1/3}. \quad (17)$$

The time-domain radiation pattern $F(\hat{\mathbf{r}}, t)$ is now computed by convolving the square pulse $G(t)$ of (16) with the time-domain pulsed-radiation pattern $F_0(\theta, t)$ of (10) and (11). In performing the convolution, one identifies two distinct angular domains $\theta \leq \Theta_c(\ell, \bar{W})$ and $\theta \geq \Theta_c(\ell, \bar{W})$ where

$$\Theta_c(\ell, \bar{W}) = 2 \sin^{-1} \left[\frac{-1 + \sqrt{1 + (\pi/6)^{2/3} \ell^{2/3} (\ell^2 - 1) \bar{W}^2}}{2(\ell^2 - 1)} \right]^{1/2}. \quad (18)$$

For $(\pi/6)^{2/3} \ell^{2/3} (\ell^2 - 1) \bar{W}^2 \ll 1$ this expression yields

$$\Theta_c(\ell, \bar{W}) = 2 \sin^{-1} \left[\frac{1}{2} \left(\frac{\pi}{6} \right)^{1/3} \ell^{1/3} \bar{W} \right]. \quad (19)$$

This approximation also provides the limit of (18) at $\ell = 1$. Θ_c is defined by the condition that for $\theta < \Theta_c$, $T(\theta) < W$ while for $\theta > \Theta_c$, $T(\theta) > W$ where $T(\theta)$ is defined in (11). Accordingly, for $\theta < \Theta_c$ there is a time interval denoted as $T/2 < t < W - T/2$ where

$$G(t) * F_0(\theta, t) = \int dt F_0(\theta, t) = \int d\mathbf{r}' q_0(\mathbf{r}') = 1 \quad (20)$$

where we have also used the normalization in (3). By manipulating the convolution $G(t) * F_0(\theta, t)$ separately in these two angular regions we obtain (21), shown at the bottom of the page, for $\theta \leq \Theta_c$ and (22), shown at the bottom of the page, for $\theta \geq \Theta_c$. From (11), $T(\theta = 0) = 0$. Hence, from (21), $F(\theta = 0, t) = G(t)$ as expected.

A. Peak-Amplitude Pattern

The peak-amplitude radiation pattern $F_{\max}(\theta)$ is found from (21)–(22) to be

$$F_{\max}(\theta) = \max_t [|F(\theta, t)|] = \begin{cases} 1 & \theta \leq \Theta_c \\ \frac{3W}{2T(\theta)} - \frac{W^3}{2[T(\theta)]^3} & \theta \geq \Theta_c. \end{cases} \quad (23)$$

Note that in the small W limit, (23) yields $F_{\max}(\theta) \sim \frac{3W}{2T(\theta)}$, which agrees with the impulsive excitation result (12) (apart from the factor W arising from the fact that in the square pulse case $\int dt G(t) = W$).

Fig. 4(a)–(f) explores the combined role of the pulse width W and of the radiator shape on the peak amplitude radiation pattern $F_{\max}(\theta)$. The figure shows plots of $F_{\max}(\theta)$ for $\ell = 100, 10, 1$ (sphere), 0.1, and 0.01 and for various normalized

pulse durations \bar{W} [see (17)]. Note the angular region near the beam axis where $F_{\max} = 1$, which is more visible for the large W case. As expected, high directivity is obtained in the large or small ℓ limits. This result is related, of course, to the overall linear dimension of the source distribution (i.e., the diameter of the smallest sphere completely enclosing the source), which becomes increasingly larger for $\ell \gg 1$ and $\ell \ll 1$. The plots also provide insight into the need for excitation pulses of short duration (relative to the source's size) in order to achieve good angular resolution.

B. Peak-Amplitude Beamwidth: Θ_c

In order to explore the effect of ℓ and \bar{W} on the beamwidth, we shall refer to the angle Θ_c defined in (18). From (23) this angle is readily identified as the 0-dB beamwidth. We have also studied numerically the behavior of the 3-dB beamwidth of the peak-amplitude radiation pattern $F_{\max}(\theta)$, but the results (not shown) were similar to those obtained for Θ_c .

We consider first the limiting case $\bar{W} \ll 1$ (quasi-impulsive excitation). In this case, we obtain from (18)

$$\Theta_c(\ell, \bar{W})|_{\bar{W} \rightarrow 0} \sim (\pi/6)^{1/3} \ell^{1/3} \bar{W}. \quad (24)$$

For large planar sources ($\ell \ll 1$), this approximation is valid if $\ell \bar{W}^3 \ll 6/\pi$, while for large elongated sources ($\ell \gg 1$), it is valid if $\ell \bar{W}^{3/4} \ll (6/\pi)^{1/4}$. Thus, under these conditions the flat sources are seen to yield better peak amplitude directivity than the elongated ones. Note also that the impulsive excitation limit ($\bar{W} = 0$) yields $\Theta_c = 0$ independently of ℓ . This, of course, is in agreement with results reported in [1], where the normalized peak amplitude radiation patterns of elongated and planar sources were found to exhibit identical angular dependence in the vicinity of the main beam direction.

The condition in (24) applies for both elongated and flat sources as long as the excitation pulse is short enough as specified above. Another possible limit is for a large elongated source ($\ell \gg 1$) excited by a relatively long pulse such that $\ell \bar{W}^{3/4} \gg 1$. In this case (19) yields

$$\Theta_c(\ell, \bar{W}) \simeq \sqrt{2}(\pi/6)^{1/6} \ell^{-1/3} \bar{W}^{1/2} \quad (25)$$

which tends to zero as $\ell \rightarrow \infty$ subject to the condition $\ell \bar{W}^{3/4} \gg 1$.

Fig. 5 shows plots of Θ_c versus ℓ and parameterized by different values of \bar{W} . For a fixed \bar{W} , there is a value of ℓ (say ℓ_0) where Θ_c reaches a maximum and the angular resolution is poorest (for a source with a given volume and pulse length).

$$F(\theta, t) = \begin{cases} \frac{1}{2} + \frac{3}{2T}t - \frac{2}{T^3}t^3 & -T/2 \leq t \leq T/2 \\ \int dt F_0(\theta, t) = 1 & T/2 \leq t \leq W - T/2 \\ \frac{1}{2} + \frac{3}{2T}(W - t) - \frac{2}{T^3}(W - t)^3 & -T/2 + W \leq t \leq T/2 + W \end{cases} \quad (21)$$

$$F(\theta, t) = \begin{cases} \frac{1}{2} + \frac{3}{2T}t - \frac{2}{T^3}t^3 & -T/2 \leq t \leq W - T/2 \\ \frac{3W}{2T} - \frac{2}{T^3}(W^3 - 3W^2t + 3Wt^2) & W - T/2 \leq t \leq T/2 \\ \frac{1}{2} + \frac{3}{2T}(W - t) - \frac{2}{T^3}(W - t)^3 & T/2 \leq t \leq W + T/2 \end{cases} \quad (22)$$

Note also that for large \bar{W} 's, there is a range of ℓ where Θ_c reaches its maximal possible value π so that the source does not exhibit any peak-amplitude directivity.

We also explored analytically the variation of ℓ_0 with \bar{W} using (18), thus finding that ℓ_0 decreases monotonically with \bar{W} . Moreover, $\ell_0 \rightarrow \infty$ as $\bar{W} \rightarrow 0$ (quasi-impulsive case). Hence, optimal selection of the shape of the radiator is accomplished in the quasi-impulsive case ($\bar{W} \rightarrow 0$) by using the smallest possible value of ℓ (i.e., a planar source).

C. Energy and Correlation Patterns

Next, we shall explore the directional dependence both of the energy radiation pattern [defined in (13)] and of a radiation characterization we shall call the “correlation” or “matched filter” pattern $L(\theta)$, which we define as

$$L(\theta) = \frac{\max_t [|F(\theta, t) * F(\theta = 0, t)|]}{\max_t [|F(\theta = 0, t) * F(\theta = 0, t)|]} \quad (26)$$

where $*$ denotes temporal convolution. $L(\theta)$ provides a measure of the directional dependence of the (normalized) peak amplitude at the output of a matched filter tuned to the far-field pulse in the main-beam direction $F(\theta = 0, t) = G(t)$ (where a target or receiver may be expected to be located). This characterization thus simulates the angular resolution available in target detection or secure communication applications.

The normalized energy radiation pattern $\bar{S}(\theta) = S(\theta)/S(\theta = 0) = \frac{S(\theta)}{\pi \bar{W}}$ is obtained from (13), (21), and (22) and is given by

$$\bar{S}(\theta) = 1 - \frac{9}{35\bar{W}}T \quad (27)$$

for $\theta \leq \Theta_c$, where $T(\theta)$ is given in (11), and by

$$\bar{S}(\theta) = \frac{6}{5} \frac{W}{T} - \left(\frac{W}{T}\right)^3 + \frac{3}{5} \left(\frac{W}{T}\right)^4 - \frac{2}{35} \left(\frac{W}{T}\right)^6 \quad (28)$$

for $\theta \geq \Theta_c$. (We normalize S so as to be able to compare sources excited by pulses of different duration.)

The “correlation pattern” $L(\theta)$ is obtained from (26), (21), and (22), giving

$$L(\theta) = \begin{cases} 1 - \frac{3}{16\bar{W}}T(\theta), & \theta \leq \Theta_{cL} \\ \frac{3}{2T(\theta)}\bar{W} - \frac{1}{[T(\theta)]^3}W^3, & \theta \geq \Theta_{cL} \end{cases} \quad (29)$$

where Θ_{cL} is defined by $T(\Theta_{cL}) = 2W$.

1) *Special Case: Quasi-Impulsive Excitation:* Consider first the limit $\bar{W} \ll 1$. We examine this case using the small W limits of $\bar{S}(\theta)$ and $L(\theta)$ as computed using (27), (28), and (29). Here, $\Theta_c \ll 1$ as described in (24). We use the small θ approximations of $T(\theta)$ from (15). For $\theta \leq \Theta_c$ we therefore obtain from (27), with (15) and (24)

$$\bar{S}(\theta) \simeq 1 - \frac{9}{35} \left(\frac{6}{\pi}\right)^{1/3} \ell^{-1/3} \bar{W}^{-1} \theta = 1 - \frac{9}{35} \frac{\theta}{\Theta_c}. \quad (30)$$

This expression describes $\bar{S}(\theta)$ near the beam axis for $\theta \leq \Theta_c$ but, as follows from the second expression in (30), it does not describe the relevant range for $\theta > \Theta_c$ where $\bar{S}(\theta)$ may still be nonnegligible [note that $\bar{S}(\Theta_c)$ is about 0.75 of the value of \bar{S} at $\theta = 0$]. Hence, the effective beamwidth of the energy pattern

will be found from the analysis of \bar{S} in the region $\theta \geq \Theta_c$ where \bar{S} is described by (28). The leading term (in W) in this expression is given by $\bar{S}(\theta) \simeq \frac{6W}{5T} \simeq \frac{6}{5} \left(\frac{\pi}{6}\right)^{1/3} \ell^{1/3} \bar{W}^{-1} \theta$, where the second approximation follows from (15). The 3-dB beamwidth obtained by solving $\bar{S}(\theta) = \frac{1}{2}$ is therefore given by

$$\theta_{3\text{dB}} = \frac{12}{5} \left(\frac{\pi}{6}\right)^{1/3} \ell^{1/3} \bar{W} = \frac{12}{5} \Theta_c \quad (31)$$

and is smaller for flat sources (small ℓ). One therefore concludes that flat sources provide better energy focusing than elongated sources of comparable volume. Note in addition that since $S(\theta) = \pi W \bar{S}(\theta)$ we have $S(\theta)|_{W \rightarrow 0} \sim \frac{6\pi W^2}{5T}$, which is consistent with the impulsive source case in (14).

A similar analysis applies to the correlation pattern in (29). Thus, for small W and θ

$$L(\theta) \simeq \begin{cases} 1 - \frac{3}{16} \left(\frac{6}{\pi}\right)^{1/3} \ell^{-1/3} \bar{W}^{-1} \theta, & \theta < \Theta_{cL} \\ \frac{3}{2} \frac{W}{T(\theta)} \simeq \frac{3}{2} \left(\frac{\pi}{6}\right)^{1/3} \ell^{1/3} \bar{W}^{-1} \theta, & \theta > \Theta_{cL} \end{cases} \quad (32)$$

Thus, comparing (32) and (30) we note that the energy and correlation patterns exhibit a similar behavior in the vicinity of the main beam direction for sources with $\bar{W} \ll 1$.

In summary, the results above support the conclusion obtained in connection with the peak-amplitude pattern $F_{\max}(\theta)$, namely, that in the quasi-impulsive limit flat broadside sources provide narrower beamwidth than elongated sources.

2) *General Case: Simulation Results:* Plots of $\bar{S}(\theta)$ for the length-to-width ratios $\ell = 100, 10, 1, 0.1$, and 0.01 and for the normalized square-pulse durations $\bar{W} = 0.1$ and $\bar{W} = 1.0$ are shown in Fig. 6(a) and (b). These plots are also very similar to those of $L(\theta)$ (not shown here). Referring to Fig. 4, we also find that $\bar{S}(\theta)$ [and, thus, $L(\theta)$, too] depends on θ , \bar{W} , and ℓ essentially in the same manner as the peak-amplitude pattern considered in Fig. 4. Furthermore, the numerical study also revealed that the 3-dB beamwidths of both $\bar{S}(\theta)$ and $L(\theta)$ depend on ℓ and \bar{W} essentially in the same manner as the peak-amplitude beamwidth Θ_c considered in Fig. 5.

D. Pulse Duration and Range Resolution

The far-field pulse duration $\tau(\theta)$ is given by [see (21) and (22)]

$$\tau(\theta) = W + T(\theta) \quad (33)$$

where $T(\theta)$ of (11) is the pulse duration under impulsive excitation (i.e., $\tau|_{W=0} = T$). Plots of the normalized pulse duration $\bar{\tau}(\theta) \equiv \tau(\theta)c/V^{1/3}$ for various values of ℓ and \bar{W} are shown in Fig. 7. Note that near the main-beam direction $\bar{\tau}(\theta)$ is mainly controlled by \bar{W} . The shape of the radiator (i.e., ℓ) has, on the other hand, a significant effect away from the main direction (for example, one can easily show that away from the main direction, $\bar{\tau}$ varies monotonically with θ only for $\ell \geq 1$).

3) *Range Resolution for Highly Directive Sources:* For highly directive sources we may consider only the angular range near the beam axis. Specifically, we shall consider the angular range $\theta < \Theta_c$ where Θ_c is peak amplitude beamwidth in (18), which provides a reasonable figure of merit to the relevant angular domain.

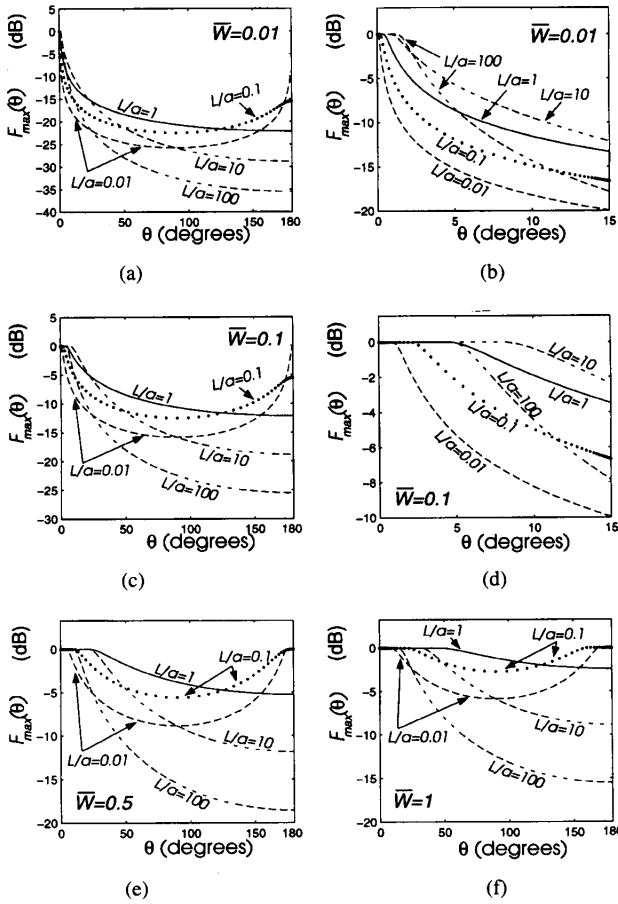


Fig. 4. Peak-amplitude pattern F_{\max} versus θ for $\ell = 0.01, 0.1, 1, 10, 100$, and for various values of \bar{W} . (a) $\bar{W} = 0.01$. (b) Small angles behavior of (a). (c) $\bar{W} = 0.1$. (d) Small angles behavior of (c). (e) $\bar{W} = 0.5$. (f) $\bar{W} = 1.0$.

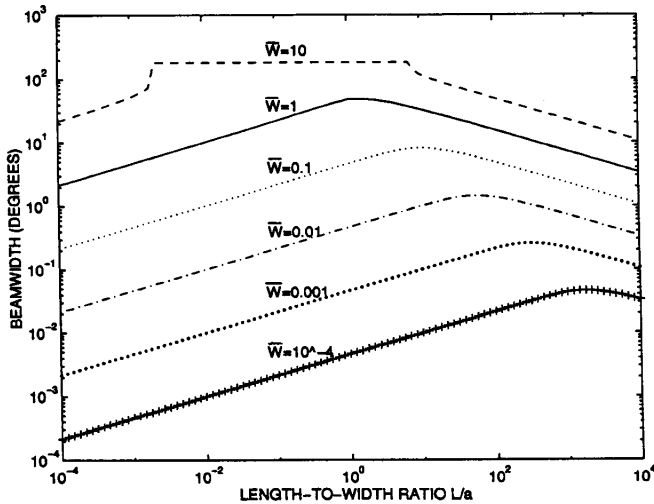


Fig. 5. Θ_c versus ℓ and parameterized by \bar{W} .

Using (33) together with (15) and (24) for small θ and Θ_c , the normalized pulse duration reduces to

$$\bar{\tau}(\theta) \simeq \bar{W} + (6/\pi)^{1/3} \ell^{-1/3} \theta = \bar{W}(1 + \theta/\Theta_c). \quad (34)$$

Next, we calculate the average pulse duration within the main-

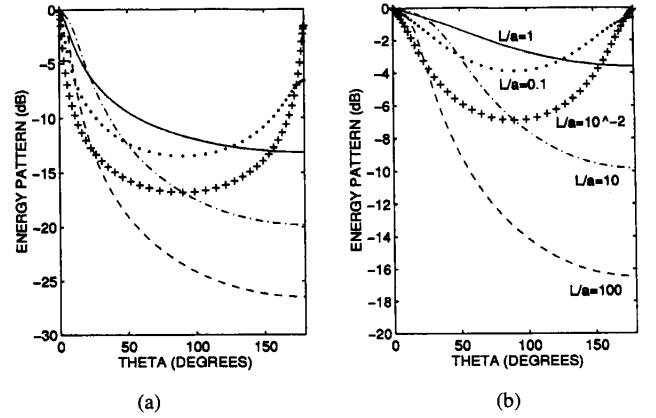


Fig. 6. Energy pattern \tilde{S} versus θ for $\ell = 0.01, 0.1, 1, 10, 100$, and for various values of \bar{W} . (a) $\bar{W} = 0.1$. (b) $\bar{W} = 1.0$.

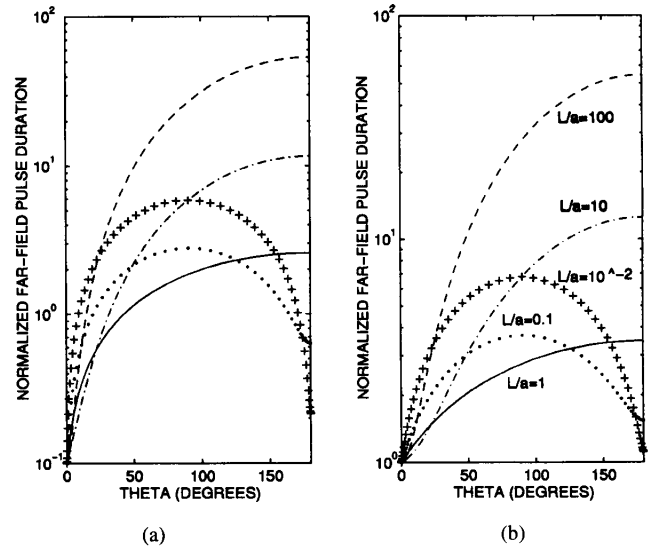


Fig. 7. Normalized far-field pulse duration $\bar{\tau}$ versus θ for $\ell = 0.01, 0.1, 1, 10, 100$, and for various values of \bar{W} . (a) $\bar{W} = 0.1$. (b) $\bar{W} = 1.0$.

beam zone. Denoting this parameter as $\langle \bar{\tau} \rangle$ we find from (34)

$$\langle \bar{\tau} \rangle \equiv \frac{1}{\Theta_c} \int_0^{\Theta_c} d\theta \bar{\tau}(\theta) \simeq \frac{3}{2} \bar{W}. \quad (35)$$

Thus, for highly directive source distributions, $\langle \bar{\tau} \rangle$ is—apart from an unessential factor—equal to the normalized pulse width \bar{W} , thereby being independent of the source shape parameter ℓ .

A drawback of the above definition is that Θ_c depends both on ℓ and \bar{W} . Hence, the average value of the far-field pulse duration is computed within different angular regions for different source configurations. An alternative approach would be to measure the average of $\bar{\tau}$ within a fixed angular region $\theta \leq \theta_0$ where θ_0 is some fixed small angle in the vicinity of the main direction. Analytical and computer-simulated results obtained for this case (not shown here) revealed that elongated sources provide better range resolution than planar sources of comparable size in the quasi-impulsive excitation case.

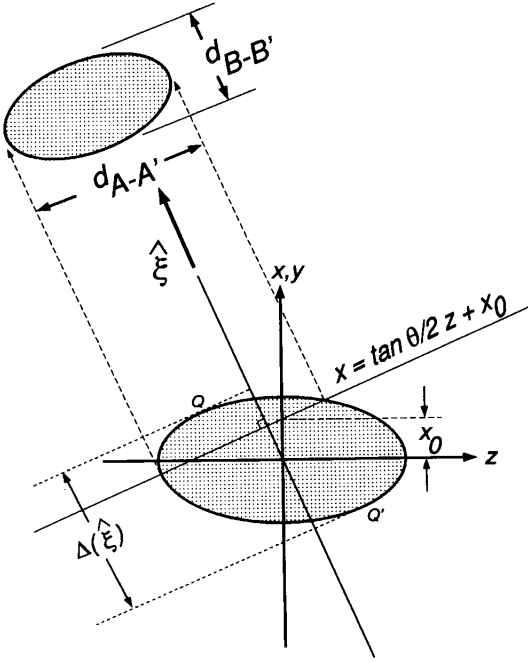


Fig. 8. Cross section of an ellipsoid volume source distribution used to compute the time-domain pulsed-radiation pattern $F_0(\hat{\mathbf{r}}, t)$ analytically.

V. CONCLUSION

This two-part sequence has dealt with the (far-field) radiation characteristics of collimated pulsed-source distributions having the “traveling-wave” form in (2). We have parameterized the radiation properties for sources of this class in both frequency and time domains. The time-domain route is based on radiation integrals in the form of planar projections of the time-dependent source distributions along the observation directions, thereby providing simple analysis and optimization tools with direct physical (geometrical) interpretation.

The general analysis tools were developed in the first paper. In that paper, we also considered explicitly the radiation characteristics of canonical parallelepiped source distributions under impulsive excitation. Such sources produce, however, singular field solutions along the main beam axis and, therefore, cannot be properly parameterized. This paper has considered finite duration signals. We studied, analytically, canonical source distributions with the shape of an ellipsoid whose axis coincides with the main-beam direction. Furthermore, by changing the ellipsoidal shape from a prolate to an oblate, we have clarified the different radiation characteristics obtained by an elongated quasi-linear traveling-wave source distribution or by a quasi-planar broadside source distribution.

We have been mainly concerned with the role both of the width of the excitation pulse and of the length-to-width ratio ℓ of the source for fixed-source volume. The source’s focusing properties associated with the peak amplitude, energy, and “correlation” patterns were found to be strongly affected by both the excitation pulse width and the shape of the source. High directivity is achieved using either large planar ($\ell \ll 1$) or elongated ($\ell \gg 1$) sources and short-duration excitation pulses (such that $\bar{W} \ll 1$).

We also found that the angular dependence of all patterns studied (peak amplitude, energy, “correlation pattern,” etc.) and, of the pulse duration, is monotonic only for $\ell \geq 1$ and that this result is independent of the normalized excitation pulse width \bar{W} . Planar broadside sources thus generate undesirable backward radiation lobes, which are not produced by elongated endfire sources.

We also explored the role of the length-to-width ratio on the directivity of the source when the normalized pulse duration is kept fixed. It was found that there is a breakdown region (or point $\ell = \ell_0$) where the beamwidth reaches its maximum (keeping \bar{W} fixed). Optimal selection of the shape of the radiator is accomplished at any of the extremes $\ell \gg \ell_0$ or $\ell \ll \ell_0$ (i.e., choosing ℓ away from the breakdown region). Moreover, we have also investigated the variation of the suboptimal length-to-width ratio ℓ_0 as a function of the width of the excitation pulse. Interestingly, ℓ_0 was found to decrease *monotonically* with the normalized pulse width \bar{W} . Moreover, $\ell_0 \rightarrow \infty$ as $\bar{W} \rightarrow 0$. Hence, minimum beamwidth is accomplished in the quasi-impulsive excitation case by using the smallest possible value of ℓ (i.e., a planar source).

The range resolution in the neighborhood of the main-beam axis was found to be controlled mainly by the width of the excitation pulse. For highly directive sources (such as quasi-impulsively excited sources) the average range resolution as computed within the peak amplitude beam region $\theta \leq \theta_c$ was found to be independent of the shape of the radiator.

APPENDIX

We refer to the geometry in Fig. 8, which schematically shows a cross section of the source in the direction $\hat{\xi} = \cos \theta/2 \hat{x} - \sin \theta/2 \hat{z}$ where, without loss of generality, the $\hat{\xi}$ axis is taken to be in the (x, z) plane. Denoting the point of intersection of the cross-sectional plane with the x axis as x_0 and the cross-sectional area by $A(\theta; x_0)$, we obtain from (6)

$$F_0(\theta, t) = \frac{c}{\xi} q_0 A(\theta; x_0) |_{x_0 \cos \theta/2 = -ct/\xi} \quad (36)$$

where, from (3), $q_0 = 1/V = \frac{6}{\pi a^2 L}$ and $\xi = 2 \sin(\theta/2)$.

The line that delimits the cross section shown is a second-order polynomial with the shape of an ellipse with axes $d_{A-A'}$ and $d_{B-B'}$ given by

$$d_{A-A'} = \frac{L \sqrt{\mu(\theta) - 4x_0^2/a^2}}{\mu(\theta) \cos(\theta/2)} \quad (37)$$

$$d_{B-B'} = a \sqrt{\frac{\mu(\theta) - 4x_0^2/a^2}{\mu(\theta)}} \quad (38)$$

where $\mu(\theta) = 1 + \ell^2 \tan^2(\theta/2)$. The area of this ellipse is, therefore, given by

$$A(\theta; x_0) = \frac{\pi}{4} \frac{La[\mu(\theta) - 4x_0^2/a^2]}{[\mu(\theta)]^{3/2} \cos(\theta/2)} \quad (39)$$

Substituting (39) in (36) yields (10). The peak amplitude is calculated using (8) with $A_{\max} = q_0 A(\theta; x_0 = 0) =$

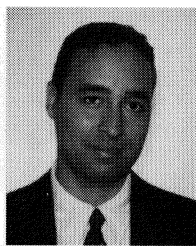
$\frac{\pi}{4} \frac{La}{\cos \theta/2 \sqrt{\mu(\theta)}}$, thus yielding

$$F_{0,\max}(\theta) = \frac{3c}{2\xi} \frac{1}{a \cos \theta/2 \sqrt{\mu(\theta)}}. \quad (40)$$

The pulse duration is calculated using (7). The quantity $\Delta(\hat{\xi})$ is obtained by finding the points Q and Q' in Fig. 8 where the ellipsoid is tangent to the plane $x = \tan \theta/2 z + x_0$ and later computing the projection of the distance $Q - Q'$ in the direction $\hat{\xi}$, giving $\Delta(\theta) = a \cos \theta/2 \sqrt{\mu(\theta)}$.

REFERENCES

- [1] E. A. Marengo, A. J. Devaney, and E. Heyman, "Analysis and characterization of ultrawideband scalar volume sources and the fields they radiate," *IEEE Trans. Antennas Propagat.*, vol. 45, pp. 1098–1106, July 1997.
- [2] D. W. Liu, P. H. Carr, and J. B. Thaxter, "Nonlinear photoconducting characteristics of antenna activated by 80 picosecond optical pulses," *IEEE Photon. Technol. Lett.*, vol. 8, pp. 815–817, June 1996.
- [3] D. W. Liu, J. B. Thaxter, and D. F. Bliss, "Gigahertz planar photoconducting antenna activated by picosecond optical pulses," *Opt. Lett.*, vol. 20, no. 14, pp. 1544–1546, July 1995.
- [4] D. H. Auston, K. P. Cheung, and P. R. Smith, "Picosecond photoconducting hertzian dipoles," *Appl. Phys. Lett.*, vol. 45, no. 3, pp. 284–286, Aug. 1984.
- [5] S. Mittlemann, P. Carr, and D. Liu, private communication, Rome Lab., Hanscom AFB, MA.
- [6] H. F. Harmuth, "Antennas for nonsinusoidal waves: Part: III-Arrays," *IEEE Trans. Electromagn. Compat.*, vol. 25, pp. 346–357, Aug. 1983.
- [7] A. Shlivinski, E. Heyman, and R. Kastner, "A unified antenna parameterization in the time and frequency domains," *IEEE Trans. Antennas Propagat.*, vol. 45, pp. 1140–1149, July 1997.
- [8] J. D. Jackson, *Classical Electrodynamics*. New York: Wiley, 1962.



Edwin A. Marengo (M'89) was born in Panama City, Panama, in 1967. He received the Licenciado degree in electromechanical engineering (*summa cum laude*) from the Technological University of Panama, Panama City, and the M.S.E.E. and Ph.D. degrees from Northeastern University, Boston, MA, in 1990, 1994, and 1997, respectively.

From 1989 to 1990, he was a Teaching Assistant at the Department of Electrical Engineering, Technological University of Panama. From 1991 to 1992 he held positions as Mechanical Engineer at Asesores Técnicos e Industriales, Panama City, as Systems Engineering Trainee at IBM, Panama City, and as Adjunct Instructor of Applied Dynamics and Thermal Machines at the Technological University of Panama and of Electronics Laboratory and Electric Machinery at Universidad Santa Maria La Antigua, Panama City. From 1992 to 1997 he was with the Department of Electrical and Computer Engineering, Northeastern University, first as a Fulbright Scholar (1992–1994)—sponsored by the U.S. Government—and later as a Stipended Graduate Research Assistant. He is currently a Post-Doctoral Research Associate at the Center for Electromagnetics Research, Department of Electrical and Computer Engineering, Northeastern University. His main research interests include transient wave radiation and scattering, computational electromagnetics, and ground-penetrating radar imaging.

Dr. Marengo is a member of the IEEE Antennas and Propagation Society, Phi Kappa Phi, and Eta Kappa Nu.

Anthony J. Devaney (M'87), for photograph and biography, see p. 1106 of the July 1997 issue of this TRANSACTIONS.

Ehud Heyman (S'80–M'82–SM'88), for photograph and biography, see p. 528 of the May 1995 issue of this TRANSACTIONS.

Estimation of Object Location from Wideband Scattering Data

George A. Tsihrintzis, Anthony J. Devaney, and Ehud Heyman

Abstract—We present a time domain algorithm for computation of the maximum likelihood estimate of the location of a known scattering object from wide-band scattering data acquired in a suite of scattering experiments. The algorithm consists of a three-step procedure: 1) data filtering, 2) time-domain backpropagation, and 3) coherent summation and is implemented via a number of forward and inverse Radon transforms integrated into a tomographic scheme. A computer simulation is included for illustration purposes.

Index Terms— Filtered backpropagation, inverse scattering, radon transform, tomography.

I. INTRODUCTION

Inverse scattering is the scientific discipline in which a scattering object is probed with incident waves in an attempt to estimate (reconstruct) its structure from scattered field measurements [1]. The scattering object is generally assumed to be embedded in a known, nonattenuating background medium and the objective is to quantitatively estimate the spatial distribution of its complex-valued index of refraction by inverting the mathematical mapping relating the probing wave, the refraction index, and the measurable total wave. This objective is nontrivial to achieve due to the inherent nonuniqueness and nonlinearity of the mapping from index of refraction to scattered field in any single scattering experiment [2]. The nonuniqueness issue can be partially addressed by employing a multiplicity of experiments, where the object is probed from several incident wave directions, and the full scattering data set is then available for the inversion. However, the issue of nonlinearity is significantly harder to address and, to date,

Manuscript received September 29, 1997; revised March 24, 1998. This work was supported in part by ARO Grant DAAG55-97-1-0013 and by AFOSR Grant F49620-96-1-0028. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. W. Clem Karl.

G. A. Tsihrintzis is with the Department of Informatics, University of Piraeus, Piraeus 185 34, Greece (e-mail: geoatsi@unipi.gr).

A. J. Devaney is with the Department of Electrical and Computer Engineering and the Center for Electromagnetics Research, Northeastern University, Boston, MA 02115 USA.

E. Heyman is with the Department of Electrical Engineering and Physical Electronics, Tel-Aviv University, Tel-Aviv 69978, Israel (e-mail: gat6v@virginia.edu).

Publisher Item Identifier S 1057-7149(99)05108-8.

research has only produced mathematical results or computationally intensive iterative algorithms as opposed to practically implementable inversion algorithms.

Over the past 20 years, an alternative approach to the inverse scattering problem has been employed based on certain linearizing approximations [3]–[6]. This approach has led to an expanded discipline within the regime of tomography, known as *diffraction tomography* (DT), which has reached today the stage of being implemented in prototype commercial tomographic scanners for ultrasonic [7], [8], seismic [9], [10], and optical [11] imaging systems. The success of the DT algorithms depends critically, however, on the two assumptions of linearity and availability of multiple experiments and in many cases, the linearity assumption fails, while different constraints (economic, safety, operating, geometric, or physical) limit the number of scattering experiments that can be performed and/or provide low signal-to-noise ratio (SNR) data. As a result, DT based scatterer reconstruction algorithms often suffer from high noise levels, poor resolution, and artifacts, such as streaking and geometric distortion.

To overcome the limitations discussed above a more modest inverse scattering problem was addressed by the first two authors, originally within the framework of linearized [12] and later exact [13] scattering theory. The goal of this more modest inverse problem, which was motivated, in part, by earlier work on a related problem in X-ray computed tomography [14], was to estimate the location of a *known* scatterer having unknown central location from noisy scattered field data. It was found that for monochromatic plane-wave probing the optimum (in the maximum likelihood sense) location estimate could be obtained via a filtered backpropagation algorithm [3], in which partial images formed by filtering and backpropagating scattered field data for different probing directions were coherently summed. The algorithm yields an *image* of the log likelihood function of the object's location and can be used for target detection and classification as well as target location estimation. The location estimation is optimum for any given number of scattering experiments and can yield a good estimate even from a single experiment as long as the wavelength of the probing radiation is comparable with the typical dimensions of the target [12].

The work in [12] and [13] is formulated and solved within the frequency domain and can be directly applied to wideband probing wavefields using standard frequency-domain synthesis. However, in many wideband applications a time-domain algorithm can be computationally more efficient as is evidenced by the time-domain algorithm recently developed for the standard reconstruction problem of DT [15]. In this correspondence, we develop such a time-domain algorithm for the object location problem within the framework of exact scattering theory. In particular, the basic framework for the location estimation problem of DT developed in [13] and the more recent time-domain formulation of standard DT presented in [15] are unified to result in a time-domain algorithm for optimal estimation of the location of a known, nondispersive scattering object from a set of wideband scattering experiments. The theory and estimation algorithm are developed within the frameworks of exact scattering theory and maximum likelihood theory and have applicability to the problems of target detection and classification as well as to target location estimation.

II. CONFIGURATION AND SCATTERING EQUATIONS

We consider the data collection configuration illustrated in Fig. 1, where a known wide-band, plane-wave pulse propagating in the direction characterized by the unit propagation vector \mathbf{s}_0 is incident on a nondispersive (frequency independent) scattering object and the scattered wave is measured over a plane located outside the

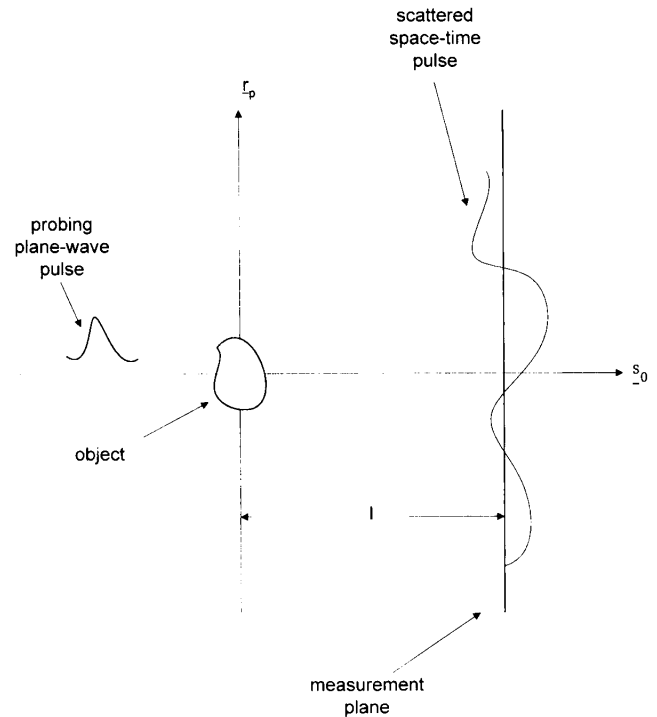


Fig. 1. Configuration of scattering experiments.

object and perpendicular to the unit propagation vector \mathbf{s}_0 . We will denote a point on the measurement plane with the coordinate vector $\mathbf{r}_0 = \mathbf{r}_p + l\mathbf{s}_0$ where \mathbf{r}_p is a two-dimensional (2-D) coordinate vector and $l > 0$ is the distance of the measurement plane from the origin. The object is assumed to be embedded in a dispersionless, nonattenuating homogeneous medium of wave velocity c_0 and characterized by its frequency independent distribution of complex-valued index of refraction $n(\mathbf{r}) = c_0/c(\mathbf{r})$, where $c(\mathbf{r})$ is the wave velocity distribution inside the scatterer.

The interaction of the incident wave pulse with the object results in the formation of a time-dependent wavefield $\Psi(\mathbf{r}, t)$ whose Fourier amplitude $\psi(\mathbf{r}, \omega)$ satisfies the time-independent inhomogeneous Helmholtz equation

$$(\nabla^2 + k^2)\psi(\mathbf{r}, \omega) = -k^2 O(\mathbf{r})\psi(\mathbf{r}, \omega) \quad (2.1)$$

where $k = \frac{\omega}{c_0}$ is the wavenumber in the background medium and $O(\mathbf{r}) = n^2(\mathbf{r}) - 1$ is known as the *object function* and is the quantity whose determination is the ultimate goal of inverse scattering theory and DT. The solution of the Helmholtz equation (2.1) can be decomposed into the sum of the incident field and a scattered field in the form $\psi(\mathbf{r}, \omega) = \psi_0(\mathbf{r}, \omega) + \psi^s(\mathbf{r}, \omega)$ where the scattered field component is related to the object function and total field via the integral equation

$$\psi^s(\mathbf{r}, \omega) = k^2 \int d^3 r' O(\mathbf{r}') \psi(\mathbf{r}', \omega) G(\mathbf{r}, \mathbf{r}', \omega) \quad (2.2)$$

where $G(\mathbf{r}, \mathbf{r}', \omega) = \exp(ik|\mathbf{r} - \mathbf{r}'|)/[4\pi|\mathbf{r} - \mathbf{r}'|]$ is the outgoing wave Green function to the Helmholtz equation [16].

The scattered field satisfies a certain source/field translation property that is a key ingredient of the theory and results presented in [13]. In particular, consider an object function that results from some other object function $O_0(\mathbf{r})$ that is shifted to some location \mathbf{R}_c ; i.e., $O(\mathbf{r}; \mathbf{R}_c) = O_0(\mathbf{r} - \mathbf{R}_c)$. The source/field translation property proved in [13] relates the field $\psi^s(\mathbf{r}, \omega; \mathbf{R}_c)$ scattered by the shifted object function to the field $\psi^s(\mathbf{r}, \omega; 0)$ scattered by the centered object

function

$$\psi^s(\mathbf{r}, \omega; \mathbf{R}_c) = e^{ik\mathbf{s}_0 \cdot \mathbf{R}_c} \psi^s(\mathbf{r} - \mathbf{R}_c, \omega; 0). \quad (2.3)$$

Equation (2.3) states that the field scattered by a shifted object function is equal to the shifted field scattered by the centered object function multiplied by the phase shift factor $\exp(ik\mathbf{s}_0 \cdot \mathbf{R}_c)$. This equation, which holds within the framework of exact scattering theory, is the key result that allows the inverse problem of estimation of the location of a known object from scattered field data to be solved within exact scattering theory.

A. Time-Domain Plane-Wave Spectra of the Scattered Wave

In the frequency domain solution of the object location estimation problem presented in [13] the *plane wave expansion* of the scattered field plays a key role. In the work presented here we will make use of the time-domain equivalent of this plane wave expansion which was employed in [15] in obtaining the time-domain inversion algorithm for conventional diffraction tomography. We first review the standard frequency domain plane wave expansion which expresses the scattered field $\psi^s(\mathbf{r}, \omega)$ as a superposition of homogeneous and evanescent plane waves that propagate from the scatterer into any half-space lying outside the scatterer support. We will employ the notation used above and assume that the scatterer is located in the half-space $z = \mathbf{s}_0 \cdot \mathbf{r} < l$. Then the scattered field throughout the half-space $\mathbf{s}_0 \cdot \mathbf{r} > l$ is expressible in the plane wave expansion [3], [4]

$$\psi^s(\mathbf{r}, \omega) = \left(\frac{k}{2\pi}\right)^2 \int d^2\xi A(\xi, \omega) e^{ik(\xi \cdot \mathbf{r}_p + \zeta z)} \quad (2.4)$$

where $\mathbf{r} = \mathbf{r}_p + z\mathbf{s}_0$ with \mathbf{r}_p being a 2-D coordinate vector in the plane passing through the origin and perpendicular to \mathbf{s}_0 . The quantity ζ is given by

$$\zeta = \begin{cases} \sqrt{1 - |\xi|^2}, & \text{if } |\xi| \leq 1 \\ i\sqrt{|\xi|^2 - 1}, & \text{else.} \end{cases} \quad (2.5)$$

The plane wave expansion equation (2.4) decomposes the scattered wavefield $\psi^s(\mathbf{r}, \omega)$ into a superposition of propagating (corresponding to $|\xi| \leq 1$) and evanescent (corresponding to $|\xi| > 1$) plane waves. Because the evanescent waves decay exponentially with distance from the scatterer their contribution to the scattered field becomes negligible at distances greater than a few wavelengths from the scatterer location. Thus, the integration region in equation (2.4) is effectively limited to the homogeneous region $|\xi| \leq 1$ at all field points located more than a few wavelengths from the scatterer.

The plane wave amplitude $A(\xi, \omega)$ can be expressed directly via a spatial Fourier transform of the boundary value field $\psi(\mathbf{r}, \omega)$ over any plane $z \geq l$. In particular, we find using (2.4) that

$$A(\xi, \omega) = e^{-ik\zeta z} \tilde{\psi}_z^s(k\xi, \omega) \quad (2.6)$$

where

$$\tilde{\psi}_z^s(\mathbf{K}_p, \omega) = \int d^2r_p e^{-i\mathbf{K}_p \cdot \mathbf{r}_p} \psi^s(\mathbf{r}, \omega)|_z$$

is the 2-D spatial Fourier transform of the scattered field on the z plane. It follows immediately from (2.6) that the spatial Fourier transform of the scattered field over any two planes, z_1, z_2 satisfies the translation property

$$e^{-ik\zeta z_1} \tilde{\psi}_{z_1}^s(k\xi, \omega) = e^{-ik\zeta z_2} \tilde{\psi}_{z_2}^s(k\xi, \omega).$$

By making use of the source/field translation property (2.3) and the above result we can relate the spatial Fourier transform of the field scattered by a shifted object function $O_0(\mathbf{r} - \mathbf{R}_c)$ to that of an unshifted object function $O_0(\mathbf{r})$. In particular, the spatial Fourier

transforms of these two fields over the measurement plane $z = l$ are related through the equation

$$\tilde{\psi}_l^s(k\xi, \omega; \mathbf{R}_c) = e^{ik[(1-\zeta)\mathbf{s}_0 - \xi] \cdot \mathbf{R}_c} \tilde{\psi}_l^s(k\xi, \omega; 0) \quad (2.7)$$

where $\tilde{\psi}_l^s(k\xi, \omega; \mathbf{R}_c)$ is the transform of the scattered field generated by the shifted object function and $\tilde{\psi}_l^s(k\xi, \omega; 0)$ the transform of the field scattered by the unshifted object function.

The plane wave expansion equation (2.4) and the associated equations (2.6) and (2.7) have time-domain equivalents. In particular, the time-domain plane-wave amplitude of the wavefield is defined as the temporal Fourier transform of the frequency domain plane wave amplitude, viz,

$$A(\xi, \tau) = \frac{1}{2\pi} \int d\omega A(\xi, \omega) e^{-i\omega\tau}. \quad (2.8)$$

On making use of (2.6), we can express the time-domain plane wave amplitude $A(\xi, \tau)$ in terms of the time-domain scattered field over the plane $z = l$ in the form

$$\begin{aligned} A(\xi, \tau) &= \frac{1}{2\pi} \int d\omega \int d^2r_p \psi^s(\mathbf{r}, \omega)|_{z=l} e^{-i\omega(\tau + \frac{\xi \cdot \mathbf{r}_p}{c_0} + \frac{\zeta}{c_0}l)} \\ &= \int d^2r_p \psi^s\left(\mathbf{r}, \tau + \frac{\xi \cdot \mathbf{r}_p}{c_0} + \frac{\zeta}{c_0}l\right)\Big|_{z=l}. \end{aligned} \quad (2.9)$$

Equation (2.9) is recognized as a Radon transform [17] of the boundary value of the wavefield $\psi^s(\mathbf{r}, t)$ in the three-dimensional (3-D) space (\mathbf{r}_p, t) and has, thus, been termed a slant-stack transform [15].

III. ESTIMATION OF OBJECT LOCATION

We consider a scatterer described by the shifted object function $O(\mathbf{r}; \mathbf{R}_c) = O_0(\mathbf{r} - \mathbf{R}_c)$ where O_0 is a known function and \mathbf{R}_c is an unknown scatterer location. The object is probed in a set of scattering experiments where, in each experiment, the incident wave is a wideband plane-wave pulse propagating in the direction \mathbf{s}_0 and the scattered field is measured over a measurement plane as illustrated in Fig. 1. We assume that in each experiment the scattered field measurements are corrupted by additive white Gaussian noise (AWGN) according to the signal model:

$$Y(\mathbf{r}_p, t, \mathbf{s}_0) = X(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{R}_c) + N(\mathbf{r}_p, t, \mathbf{s}_0) \quad (3.1)$$

where

$$X(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{R}_c) = F(\mathbf{r}_p, t, \mathbf{s}_0) \odot \Psi_l^s(\mathbf{r}_p + l\mathbf{s}_0, t, \mathbf{s}_0; \mathbf{R}_c) \quad (3.2)$$

is the time-dependent scattered field Ψ_l^s on the measurement plane $z = l$ filtered by the space-time convolutional (measurement) filter $F(\mathbf{r}_p, t, \mathbf{s}_0)$. In (3.1), $N(\mathbf{r}_p, t, \mathbf{s}_0)$ is a zero-mean white Gaussian noise process¹ in the variables \mathbf{r}_p, t , and \mathbf{s}_0 , i.e.,

$$\mathcal{E}\{N(\mathbf{r}_p, t, \mathbf{s}_0)N(\mathbf{r}'_p, t', \mathbf{s}'_0)\} = \sigma_N^2 \delta(\mathbf{r}_p - \mathbf{r}'_p) \delta(t - t') \delta_{\mathbf{s}_0, \mathbf{s}'_0}. \quad (3.3)$$

The inverse problem that we address is that of estimating the unknown parameter vector (object location) \mathbf{R}_c from the measurements filtered and noisy measurements of the scattered field $Y(\mathbf{r}_p, t, \mathbf{s}_0)$.

¹Gaussian noise of arbitrary color can be handled by expanding the algorithm of this section to include a proper whitening filter.

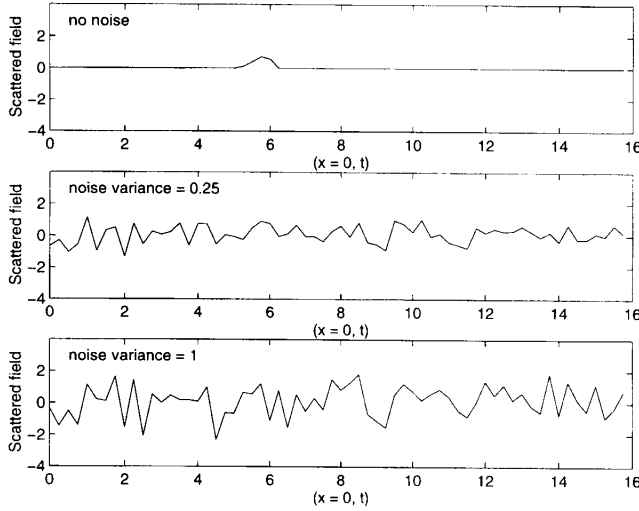


Fig. 2. Measured scattered field at $(x = 0, z = 5)$ for time $0 < t < 16$ and different noise levels.

A. Log Likelihood Function

The maximum likelihood estimate (MLE) of the object location vector \mathbf{R}_c is obtained by maximizing the *log likelihood function* [18]

$$L(\mathbf{r}_c) = \sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} dt \int d^2 r_p Y(\mathbf{r}_p, t, \mathbf{s}_0) \tilde{X}(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c) - \frac{1}{2} \sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} dt \int d^2 r_p |\tilde{X}(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c)|^2 \quad (3.4)$$

with respect to test location \mathbf{r}_c . We have the following theorem.

Theorem 1: In the absence of evanescent plane waves the second term in (3.4) is independent of object location \mathbf{r}_c and can be dropped from the log likelihood function.

We first make use of Parseval's theorem to write

$$\sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} dt \int d^2 r_p |\tilde{X}(\mathbf{r}_p, t, \mathbf{s}_0; \mathbf{r}_c)|^2 = \left(\frac{1}{2\pi} \right)^3 \sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} d\omega k^2 \int_{|\xi| \leq 1} d^2 \xi |\tilde{X}(k\xi, \omega, \mathbf{s}_0; \mathbf{r}_c)|^2 \quad (3.5)$$

where $k = \omega/c_0$ is the wavenumber and $\tilde{X}(k\xi, \omega, \mathbf{s}_0; \mathbf{r}_c)$ denotes the space-time Fourier transform of the filtered scattered field data. Making use of (2.7) we have that

$$\tilde{X}(k\xi, \omega, \mathbf{s}_0; \mathbf{r}_c) = e^{ik[(1-\zeta)\mathbf{s}_0 - \xi] \cdot \mathbf{r}_c} \tilde{X}(k\xi, \omega, \mathbf{s}_0; 0) \quad (3.6)$$

which, when substituted into (3.5) establishes the theorem since the quantity $(1-\zeta)\mathbf{s}_0 - \xi$ is purely real over the homogeneous region $|\xi| \leq 1$.

Theorem 2: In the absence of evanescent plane waves the log likelihood function can be expressed in the form

$$L(\mathbf{r}_c) = -\frac{1}{(2\pi c_0)^2} \sum_{\mathbf{s}_0} \int_{|\xi| \leq 1} d^2 \xi \frac{\partial^2}{\partial \tau^2} \times [\hat{Y}(\xi, \tau, \mathbf{s}_0) \hat{X}(\xi, \tau, \mathbf{s}_0; 0)] \Big|_{\tau = -\frac{[(1-\zeta)\mathbf{s}_0 - \xi] \cdot \mathbf{r}_c}{c_0}} \quad (3.7)$$

where

$$\hat{X}(\xi, \tau; \mathbf{s}_0; 0) = \int d^2 r_p \tilde{X}\left(\mathbf{r}_p, \tau + \frac{\xi \cdot \mathbf{r}_p}{c_0}; \mathbf{s}_0; 0\right) \\ \hat{Y}(\xi, \tau, \mathbf{s}_0) = \int d^2 r_p Y\left(\mathbf{r}_p, \tau + \frac{\xi \cdot \mathbf{r}_p}{c_0}, \mathbf{s}_0\right)$$

are the slant stacks of $\tilde{X}(\mathbf{r}_p, \tau, \mathbf{s}_0; 0)$ and $Y(\mathbf{r}_p, \tau, \mathbf{s}_0)$ [cf., (2.9)].

We again start with Parseval's theorem to find that the log likelihood function, in the absence of the second term, can be expressed in the form

$$L(\mathbf{r}_c) = \left(\frac{1}{2\pi} \right)^3 \sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} d\omega k^2 \int_{|\xi| \leq 1} d^2 \xi \tilde{Y}(k\xi, \omega, \mathbf{s}_0) \\ \times \tilde{X}(k\xi, \omega, \mathbf{s}_0; \mathbf{r}_c) \\ = \frac{1}{(2\pi)^3 c_0^2} \sum_{\mathbf{s}_0} \int_{-\infty}^{\infty} d\omega \int_{|\xi| \leq 1} d^2 \xi \omega^2 \tilde{Y}(k\xi, \omega, \mathbf{s}_0) \\ \times \tilde{X}(k\xi, \omega, \mathbf{s}_0; 0) e^{-i\omega \frac{[(1-\zeta)\mathbf{s}_0 - \xi] \cdot \mathbf{r}_c}{c_0}} \quad (3.8)$$

where we have put $k = \omega/c_0$ and made use of (3.6). The quantity

$$\tilde{Y}(\xi, \tau; \mathbf{s}_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \tilde{Y}(k\xi, \omega, \mathbf{s}_0) e^{-i\omega \tau} \\ = \int d^2 r_p \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega y(\mathbf{r}_p, \omega; \mathbf{s}_0) e^{-i\omega(\tau + \frac{\xi \cdot \mathbf{r}_p}{c_0})} \\ = \int d^2 r_p Y\left(\mathbf{r}_p, \tau + \frac{\xi \cdot \mathbf{r}_p}{c_0}, \mathbf{s}_0\right)$$

is recognized as being the slant stack of the measured field data on the measurement plane where we have denoted the time Fourier transform of the measured field data by $y(\mathbf{r}_p, \omega; \mathbf{s}_0)$. Similarly,

$$\hat{X}(\xi, \tau; \mathbf{s}_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \tilde{X}(k\xi, \omega, \mathbf{s}_0; 0) e^{-i\omega \tau}$$

is the slant stack of $\tilde{X}(\mathbf{r}_p, t; \mathbf{s}_0; 0)$. It then follows, again using Parseval's theorem, that

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \omega^2 \tilde{Y}(k\xi, \omega, \mathbf{s}_0) \tilde{X}(k\xi, \omega, \mathbf{s}_0; 0) e^{-i\omega \frac{[(1-\zeta)\mathbf{s}_0 - \xi] \cdot \mathbf{r}_c}{c_0}} \\ = \int_{-\infty}^{\infty} d\tau \frac{\partial^2}{\partial \tau^2} [\hat{Y}(\xi, \tau; \mathbf{s}_0) \hat{X}(\xi, \tau; \mathbf{s}_0; 0)] \\ \times \delta\left[\tau + \frac{[(1-\zeta)\mathbf{s}_0 - \xi] \cdot \mathbf{r}_c}{c_0}\right] \\ = \frac{\partial^2}{\partial \tau^2} [\hat{Y}(\xi, \tau; \mathbf{s}_0) \hat{X}(\xi, \tau; \mathbf{s}_0; 0)] \Big|_{\tau = -\frac{[(1-\zeta)\mathbf{s}_0 - \xi] \cdot \mathbf{r}_c}{c_0}}$$

Finally, on substituting the above expression into (3.8) we obtain the desired result.

Equation (3.7) can be interpreted as follows: For each incident wave pulse, the scattered field data are Radon-transformed with respect to their space-time coordinates, filtered in Radon space, and inverse Radon-transformed (one inverse Radon transform per value of $\mathbf{s}_0 \cdot \mathbf{r}_c$) into object space to form partial images of the log likelihood function. The Radon-space filter consists of the complex conjugate of the time-domain plane-wave spectra of the field scattered by the centered object $O_0(\mathbf{r})$. Finally, the partial images are coherently superimposed.

IV. COMPUTER ILLUSTRATION

We consider a single scattering experiment in a 2-D geometry in which a scatterer lies in the (x, z) -plane and is infinitely long and uniform along the y -axis. The probing pulse is incident along the direction of the positive z -axis and data are measured along the line $(x, z = l)$. The field scattered by the centered scatterer (corresponding to the object located at the origin) is the pulse

$$w_0^s(\mathbf{r}, t) = \frac{\sigma^2}{c_0^2} \frac{q\left(t - \frac{|\mathbf{r}|}{c_0}\right)}{\sqrt{|\mathbf{r}|}} \quad (4.1)$$

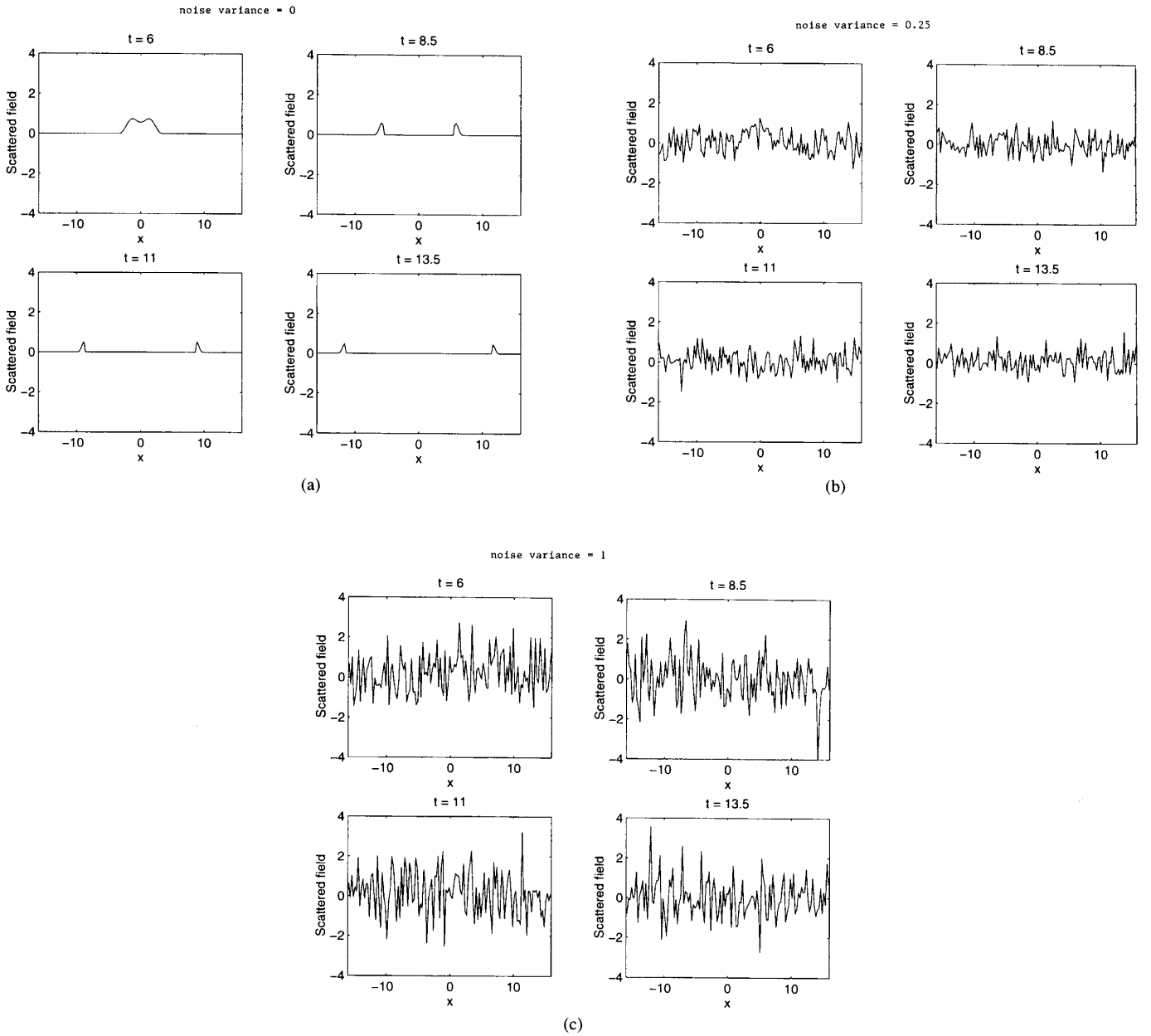


Fig. 3. Snapshots of scattered field along the measurement line $(x, z = 5)$ for times $t = 0, t = 8.5, t = 11,$ and $t = 13.5$. (a) Noise variance = 0. (b) Noise variance = 0.25. (c) Noise variance = 1.

where $\mathbf{r} = (x, z)$ denotes a point in the (x, z) plane and σ^2 is the scatterer cross section and

$$q(t) = \begin{cases} \frac{2\beta}{T} [\cos(\beta t^2) [1 + 2\beta t^2] - 1], & \text{if } 0 < t < T \\ 0, & \text{else.} \end{cases} \quad (4.2)$$

Equation (4.1) is an approximation of the field scattered by a line scatterer of cross section σ^2 when probed with a plane-wave pulse. The scattered field generated by a shifted line scatterer is obtained using (2.3).

We assume the scatterer to be located at the origin of the (x, z) -plane and chose the parameters $\sigma^2 = 1, c_0 = 1, \beta = 1, T = 1,$ and $l = 5$. For the additive Gaussian noise, we examined three different cases with corresponding variances 0, 0.25, and 1. In Fig. 2, we show the scattered pulse at $(x = 0, z = 5)$ for time $0 < t < 16$ for the three noise levels. Fig. 3 shows four snapshots of the field along

the measurement line $(x, z = 5)$ taken at times $t = 6, t = 8.5, t = 11,$ and $t = 13.5$ for the three noise levels. Finally, the likelihood functions computed via (3.7) are shown in Fig. 4 for the three noise levels.

V. SUMMARY, CONCLUSIONS, AND FUTURE WORK

In this correspondence, we established that the log likelihood function for estimation of the location of a known scattering object from noisy wideband scattering data can be computed via a time-domain, filtered backpropagation algorithm consisting of a sequence of direct and inverse Radon transforms of the space-time measurements. Target classification/identification can also be performed via a similar algorithm, in which a bank is employed of filters matched to various target signatures. A computer simulation of a single scattering experiment was performed to illustrate the procedure, which revealed high algorithm performance even

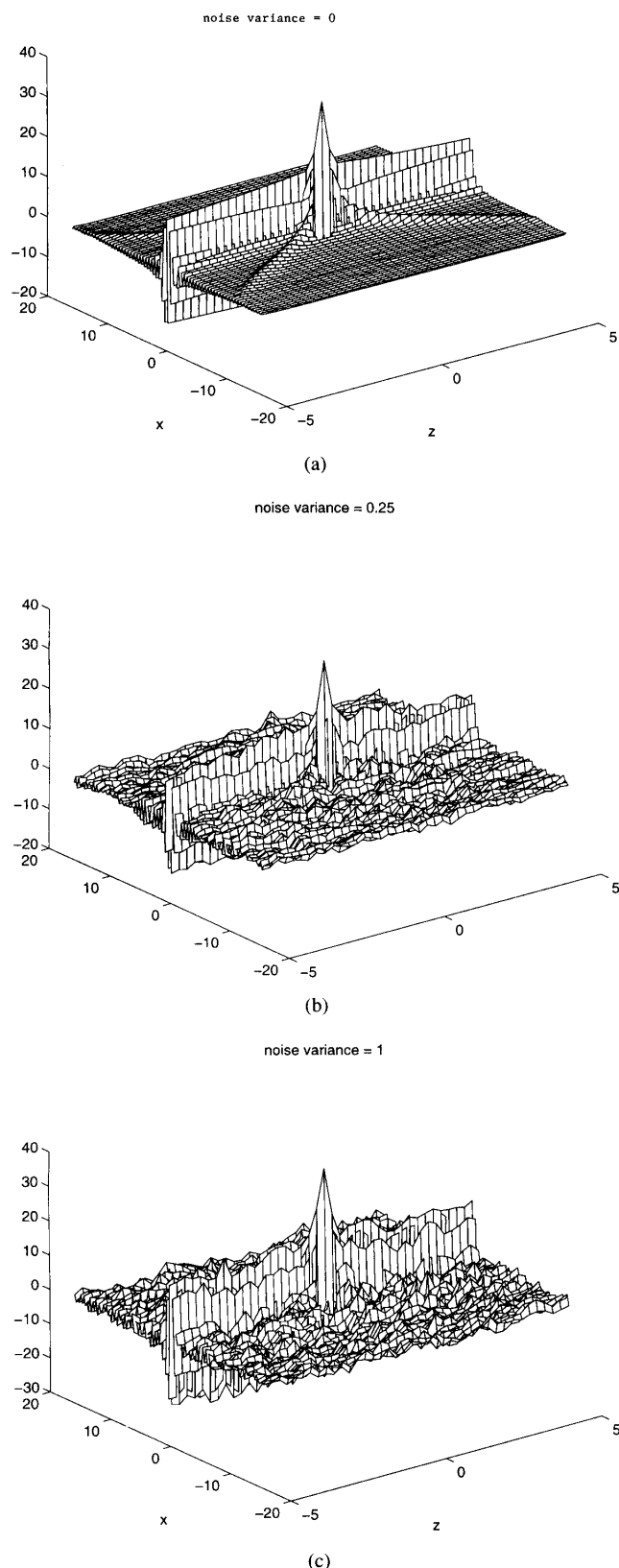


Fig. 4. Log likelihood functions for estimation of object location. (a) Noise variance = 0. (b) Noise variance = 0.25. (c) Noise variance = 1.

in the case of low SNR. Further performance improvements can be achieved if a multiplicity of scattering experiments are utilized.

Related research issues to be addressed in the future include the derivation of proper location estimation algorithms for the cases of measurement planes that remain fixed from scattering experiment to experiment. This is the case in geophysical surveys in which the sensor array is fixed in space and several scattering experiments are performed, each with a different probing plane-wave pulse [19]. Another avenue of future research is the derivation of nonparametric algorithms for detection, location estimation, and classification of stochastic scattering objects.

REFERENCES

- [1] R. G. Newton, *Scattering Theory of Waves and Particles*. Berlin, Germany: Springer-Verlag, 1982.
- [2] A. J. Devaney, "Nonuniqueness in the inverse scattering problem," *J. Math. Phys.*, vol. 19, p. 1526, 1978.
- [3] —, "A filtered backpropagation algorithm for diffraction tomography," *Ultrason. Imag.*, vol. 4, p. 336, 1982.
- [4] —, "Reconstructive tomography with diffracting wavefields," *Inv. Probl.*, vol. 2, p. 161, 1986.
- [5] E. Wolf, "Three dimensional structure determination of semi transparent objects from homographic data," *Opt. Commun.*, vol. 1, p. 153, 1969.
- [6] —, "Principles and development of diffraction tomography," in *Trends in Optics*, A. Consortini, Ed. New York: Academic, 1996.
- [7] N. Sponheim, L.-J. Gelius, I. Johansen, and J. J. Stamnes, "Quantitative results in ultrasonic tomography of large objects using line sources and curved detector arrays," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 38, p. 370, 1991.
- [8] G. A. Tsihrintzis and A. J. Devaney, "Application of a maximum likelihood estimator in an experimental study of ultrasonic diffraction tomography," *IEEE Trans. Med. Imag.*, vol. 12, pp. 545–554, 1993.
- [9] R. B. Pratt and M. H. Worthington, "The application of diffraction tomography to cross-hole seismic data," *Geophysics*, vol. 53, p. 1284, 1988.
- [10] A. Witten and W. C. King, "Acoustical imaging of subsurface features," *J. Environ. Eng.*, vol. 116, p. 166, 1990.
- [11] M. H. Maleki, A. J. Devaney, and A. Schatzberg, "Tomographic reconstruction from optical scattered intensities," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1356–1363, 1992.
- [12] A. J. Devaney and G. A. Tsihrintzis, "Maximum likelihood estimation of object location in diffraction tomography," *IEEE Trans. Signal Processing*, vol. 39, p. 672, 1991.
- [13] G. A. Tsihrintzis and A. J. Devaney, "Maximum likelihood estimation of object location in diffraction tomography, part II: Strongly scattering objects," *IEEE Trans. Signal Processing*, vol. 39, p. 1466, 1991.
- [14] D. J. Rossi and A. S. Willsky, "Reconstruction from projections based on detection and estimation of objects, parts I and II: Performance analysis and robustness analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, p. 886, 1984.
- [15] T. Melamed, Y. Ehrlich, and E. Heyman, "Short-pulse inversion of inhomogeneous media: A time-domain diffraction tomography," *Inv. Probl.*, vol. 12, pp. 977–993, 1996.
- [16] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics, Parts I, II*. New York: McGraw-Hill, 1953.
- [17] S. Deans, *The Radon Transform and Some of Its Applications*. New York: Wiley, 1983.
- [18] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley, 1968.
- [19] A. J. Devaney, "Geophysical diffraction tomography," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-22, p. 3, Jan. 1984.

Statistical approach to multi-channel spatial modeling for the detection of mine-like targets

Robert A. Weisenseel^a

W. Clem Karl^a

David A. Castañón^a

Charles A. Dimarzio^b

^aBoston University, Electrical and Computer Engineering Department

^bNortheastern University, Electrical and Computer Engineering Department

ABSTRACT

We present a statistically-based method for the enhancement and detection of mines and mine-like targets, in multi-channel imagery. Standard approaches to such multi-channel image processing take advantage of the correlation across channels within a pixel, but typically do not exploit the spatial dependency between pixels. This work aims to construct appropriate spatial statistical models for multi-channel mine imagery and apply these models to allow both image enhancement as well as direct and improved detection of anomalies (i.e., targets) in such data.

We base the method on a Markov Random Field (MRF) model that incorporates a priori information about both the target's and the background's spatial characteristics. In particular, we find a Maximum A Posteriori (MAP) detector of mine targets in background under the prior assumption target pixels are locally spatially dependent. We implement our algorithm on polarimetric and thermal data obtained from the Remote Minefield Detection System (REMIDS), with favorable results compared to a Maximum Likelihood (ML) detector that performs detections on a pixel-by-pixel basis, i.e. without spatial correlation.

Keywords: mine, MRF, multispectral, polarimetric, thermal, statistical, spatial, REMIDS, UXO, detection

1. INTRODUCTION

The need for mine and unexploded ordnance (UXO) detection and removal is growing in both military and humanitarian applications. In places like Cambodia, the threat of mines to the general populace is overwhelming. Since the end of the Cold War, there has been a growth in smaller, regional conflicts where threats are often not from high-tech weaponry, but from inexpensive ordnance. Bosnia is just one example. Mines are one of the least expensive weapons available, and their threat often far outlasts the conflict for which they are emplaced. Mine detection is therefore necessary in two roles: detecting minefield obstacles for military intelligence and detecting individual mines for eventual removal.

One specific subject of mine detection involves wide-area surveillance. In one case, military forces need to chart possible impediments to ground movement accurately over broad swaths of territory. In humanitarian applications, surveyors examining terrain for mine clearance can limit the area searched with wide-area surveillance. For both situations, it is highly desirable to conduct mine searches from the air to minimize the danger to personnel. Unfortunately, ground penetrating radar and quasistatic electromagnetic approaches are somewhat limited in range.

Optical techniques can meet many of the requirements of wide-area minefield detection. One area of much interest is polarimetric sensing. In the near infrared domain, mines have a highly polarizing characteristic. While this feature alone is inadequate to detect mines, it can be a powerful tool when combined with multispectral information, such as thermal imaging or imaging spectrometers. For all-weather capabilities, an active infrared sensor in combination with a thermal sensor is nearly ideal.

However, while polarization and multispectral capabilities, when fused in a rational, statistical manner, can go a long way toward perfecting mine detection from an airborne platform, they can fail under some conditions. For example, during diurnal crossover periods or other situations of low thermal contrast, polarization methods alone cannot make up for the lost information from the thermal channel. Under these conditions, we may require other sources of information to obtain good detection rates while restricting false alarms. In fact, mine detection rates are high using only polarization information, but reducing the false alarm rate is a harder problem, requiring that we apply additional information. Statistical spatial models of prior knowledge can help reduce false alarm rates in a reasonable, quantifiable way. By encoding prior spatial information, such as shape, smoothness, or pattern, in a

statistical form, we can fuse that information with the available data to improve the mine detection process under a wide range of operational conditions.

2. BACKGROUND

2.1. Sensor

Our data consist of three infrared imagery channels generated by the U.S. Army Engineer Waterways Experiment Station's Remote Minefield Detection System (REMIDS).^{1,2} The first two channels are imaged using an active infrared system at the $1.05\ \mu\text{m}$ wavelength. One channel is percent polarization, $(P - S)/(P + S)$, and the other channel is total reflectance, $(P + S)$, where P is reflectance in parallel polarization and S is reflectance in cross polarization. The third channel is a passive thermal infrared channel operating over the $8\text{--}12\ \mu\text{m}$ range. The sensor is mounted on an airborne platform, represented in Figure 1. The data that we use for this study has a resolution

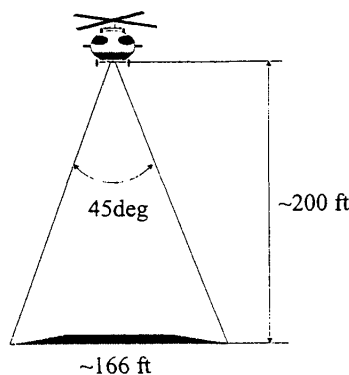


Figure 1. REMIDS Thematic Representation

of 2-3 inches per pixel, although later system upgrades have increased this resolution.

Several researchers have shown that polarization characteristics are advantageous in the mine detection process.^{1,3,4} Man-made objects such as mines and other unexploded ordnance (UXO) are significant near infrared polarizers compared to natural backgrounds, which tend to be random polarizers. REMIDS only uses linear polarization features, but the same algorithmic techniques can be extended to circular polarization as well. Additionally, thermal data alone is sensitive to weather and time of day, especially diurnal thermal crossover.

The active near-IR sensor used in REMIDS is capable of all-weather imaging, but this all-weather capability comes at the price of a laser source, its power supply, and other support equipment. A passive polarization sensor could be implemented cheaply, but would have weather-dependent detection and false alarm rates. An alternative system is a passive hyperspectral polarimetric imager,³ but this again increases the system cost.

2.2. Algorithms

Because detecting individual mines from the air is often prone to false alarms, many researchers have concentrated on detecting minefields, which is to say a collection of mines that can be grouped in some way. Some have used collinearity,⁵ and some have used Poisson point models⁶⁻⁸ as a basis for detecting minefields. However, collinearity is useful only when studying patterned minefields, and Poisson point models are limited by their preprocessor's ability to screen out false detections of individual mines. To obtain better minefield detection, a clear solution is to either improve the detection of individual mines or reduce the number of individual mine false alarms, or both. A good mine detection model can then be combined with a minefield detection model with the end result being improved detection of both mines and minefields.

One way to improve the individual mine detection process is to model the system statistically. Statistical model-based approaches can provide many advantages. Bayesian approaches provide a rational framework in which to model and fuse both information and its uncertainty. Statistical models frequently provide confidence measures of results.

They make fusion of information from disparate sources, especially prior information, a reasonable and quantifiable process.

One appropriate system for detecting mines in background is a Maximum A Posteriori (MAP) detector. With x being the image where each pixel can take on two states, mine or background, and y being the multichannel data image, the formula for determining the values of x is:

$$\hat{x} = \operatorname{argmax}_x p(x|y) = \operatorname{argmax}_x p(y|x)p(x) \quad (1)$$

where $p(x|y)$ is the posterior probability distribution of the mine location image conditioned on the multichannel data image, $p(y|x)$ is the likelihood of the data image conditioned on the mine locations, and $p(x)$ is the prior probability distribution of the mine location image. Thus, we just need three things: an observation probability model, a prior probability model, and an algorithm for maximizing over all possible mine location images x .

3. METHOD

For our observation model, we assume that each pixel of data, y_i , is conditionally independent of all other data pixels, conditioned on knowledge of mine presence at that pixel, x_i . Formally,

$$p(y|x) = p(y_1, \dots, y_N | x_1, \dots, x_N) = \prod_{i=1}^N p(y_i | x_i) \quad (2)$$

where N is the number of pixels in the image. We also assume that each pixel is an identically distributed Gaussian random 3-vector, with one element for each of the three channels: percent polarization, total reflectance, and thermal,

$$p(y_i | x_i) = \frac{1}{|2\pi\Sigma_{x_i}|} \exp \left\{ -\frac{1}{2} (y_i - \mu_{x_i})^T \Sigma_{x_i}^{-1} (y_i - \mu_{x_i}) \right\} \quad (3)$$

where μ_{x_i} and Σ_{x_i} are the mean and the covariance, respectively, of the observation variable, given that the observation is of type x_i . This type can be either mine or background, which we will denote as 1 and 0, respectively. For the purposes of this paper, we have assumed that the mean and covariance are known for both background and mines, and we have estimated the mean and a full covariance matrix for both mines and background directly from the data. In practice, we would need to estimate these parameters from the data, possibly by using a tool such as the Expectation-Maximization algorithm.

For our prior model, we have chosen to introduce spatial correlation of mine or background using a Markov Random Field (MRF) model.⁹ Some previous work has been done on using MRFs in mine detection and classification problems.^{10,11} In one case, the focus was on detecting boundaries using an MRF boundary structure model. In another case, the author based his approach on detecting deviations from an autoregressive (MRF) model that he computed from the data. Neither of these studies dealt with the problem of wide-area surveillance. We have chosen to use a type of discrete “smoothness,” or region-based, model, under the assumption that the mines we are detecting extend over multiple pixels and thus any pixel with a mine will be near several other pixels containing mines:

$$p(x) = \frac{1}{\mathcal{Z}} \prod_{k=1}^N \exp \left\{ -\alpha \sum_{j \in \mathcal{N}_k} (x_k - x_j)^2 \right\} \quad (4)$$

The \mathcal{Z} term is a normalizing constant given by,

$$\mathcal{Z} = \sum_x \prod_{k=1}^N \exp \left\{ -\alpha \sum_{j \in \mathcal{N}_k} (x_k - x_j)^2 \right\} \quad (5)$$

and \mathcal{N}_k is the set of four nearest neighbors around pixel k , as shown in Figure 2. Those familiar with statistical mechanics may recognize this as a variant form of the Ising magnet model. In simpler terms, Equation 4 states that each pixel has a higher probability of being the same as its neighbors than of looking different. We can then combine

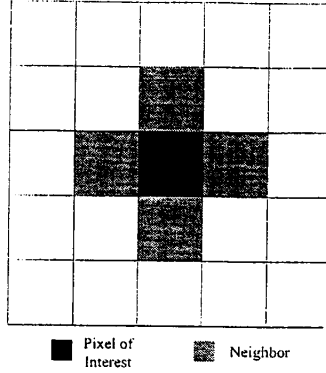


Figure 2. Neighborhood Structure of the Spatial Model

these two models into a function proportional to the posterior distribution with α as a weighting constant between the prior and likelihood distributions:

$$p(x|y) \propto e^{-H(x|y)} \quad (6)$$

where

$$H(x|y) = \sum_{k=1}^N \left[\alpha \sum_{j \in \mathcal{N}_k} (x_k - x_j)^2 + \frac{1}{2} (y_k - \mu_{x_k})^T \Sigma_{x_k}^{-1} (y_k - \mu_{x_k}) \right] \quad (7)$$

The last thing we need is a means to maximize this joint distribution, and hence the posterior distribution, for all x . We have chosen to implement this maximization using a simulated annealing algorithm because the distribution of x is discrete.¹² Thus, we introduce a “temperature” parameter to the distribution:

$$p(x|y) \propto e^{-H(x|y)/T} \quad (8)$$

To obtain an optimal solution, T should start at a sufficiently high temperature and be reduced slowly according to a logarithmic cooling schedule. However, to reduce the number of computations necessary, we chose to initialize the system at a relatively low temperature with a pixel-by-pixel Maximum Likelihood (ML) solution using the likelihoods given in Equation 3. We then reduced the temperature with an exponential cooling schedule:

$$\frac{1}{T} = \gamma e^{\beta k} \quad k = 1, 2, \dots \quad (9)$$

While this approach is not guaranteed to achieve a globally optimal solution, the initialization makes this algorithm unlikely to do worse than the ML solution and the exponential cooling schedule gives fast convergence. Because only local interactions are involved in this processing and the number of local operations required are small, this technique is well-suited to massive parallelization, either on multiple processors or on application specific ICs.

4. RESULTS

For our test we had access to only two data sets, both from 1991 test flights of REMIDS over Fort Drum, New York. The first flight was conducted at 9 am on an overcast day, resulting in a complete lack of thermal contrast between mines and background. The second flight was conducted at 3 pm on a sunny day. Representative segments of all channels of both scenes are shown in Figures 3 and 4. The mines in these images are surface patterned anti-tank mines, but note that our technique does not use any pattern based information, so it is equally applicable to scatterable mines as long as sufficient spatial resolution is available.

The good thermal and percent polarization contrast in the sunny imagery made near perfect detection possible with simple techniques, such as thresholding the three images separately and combining them with boolean operations, so this data will not be considered here. However, the cloudy imagery removed the advantages conveyed by the thermal channel. A likelihood ratio test based on a Gaussian model for each hypothesis, mine and background, produced a very high false alarm rate (Figure 5) Using the MAP estimator with its a priori “smoothness” model

significantly reduced the false alarms, as demonstrated in Figure 6. An empirically determined ROC curve appears in Figure 7. We computed this ROC by taking the total number of false alarms in a 710×9000 pixel ($166\text{ft} \times 2100\text{ft}$) image as the x-axis and the ratio of detected mines to actual mines as the probability of correct detection on the y-axis. A detection (both correct detection and false alarms) was any continuously 4-connected region of mine pixels.

We found that we could also achieve good minefield detection simply by counting the number of detections in a processed block. In each case where a minefield was present in a block of the cloudy day data, the number of individual mine detections significantly exceeded some threshold between 4 and 10, as shown in Table 1. This result

	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Block 9
ML	299	260	522	44	433	233	174	183	341
MAP	2	0	1	0	11	12	12	0	3
True	0	0	0	0	11	11	11	0	0

Table 1. Number of Mine Detections in Each Data Block

shows that we have found a simple tool for minefield detection that does not depend on patterned minefield layouts. In addition, our method can still be combined with other statistical models of mine layouts, to further improve both minefield and individual mine detection.

Note that the only ground truth available to us at the time of this writing comes from human evaluation of the data. Note also that our algorithms, both the ML and the MAP detectors, detected every individual mine. The only errors in these results arose from false alarms. Because of the small data set, however, it would not be prudent to attempt to estimate true error rates from these data.

Processing a $710 \times 1000 \times 3$ data block through twenty iterations in MATLAB using non-optimized code required approximately 2.5 minutes on a Sun Ultra 1. Large speedups are possible by going to native compiled code rather than interpreted code and by implementing this algorithm in a massively parallel environment.

5. CONCLUSIONS

We have demonstrated the incorporation of a spatial correlation structure in a mine detector using a statistical approach to fuse a priori spatial knowledge with multichannel imagery. We implemented this approach on three channel data from the REMIDS sensor, comprised of percent polarization and total reflectance from an active sensor in the near infrared domain and a single passive thermal infrared channel. This MAP formulation resulted in a noticeable improvement over an ML hypothesis test under reasonably common conditions where the thermal channel provided little or no discriminating information. Limited data availability precluded a more thorough evaluation of performance measures.

This approach required no information about the pattern of the minefield or the shape of the mines. We only required the information that mines form multi-pixel regions. Hence, there is no reason why this approach would not work equally well on scatterable mines. In addition, because we formulated this algorithm as a statistical model, we can fuse the results of this detection process with other sources of information in a rational way. For example, we could have included a global statistical model of the mine locations in a minefield to better detect minefields. In future work, we aim to examine the buried minefield problem.

6. ACKNOWLEDGMENTS

We would like to thank Ricky Goodson and Ernesto Cespedes of the US Army Corps of Engineers Waterways Experimental Station's Environmental Laboratory for graciously sharing their REMIDS data. This work is partially supported by the Army Research Office under Grant ARO DAAG55-97-1-0013, the Air Force Office of Scientific Research under Grant F49620-96-1-0028, and the National Institutes of Health under Grant NINDS 1 R01 NS34189.

REFERENCES

1. B.H. Miles, E.R. Cespedes, R.A. Goodson, "Polarization-Based Active/Passive Scanning System for Minefield Detection", *Polarization and Remote Sensing*, Walter G. Egan, editor, Proc. SPIE V 1747, pp239-252, SPIE, San Diego, (1992)
2. G. Maksymonko, K. Breiter, "ASTAMIDS Minefield Detection Performance at Aberdeen Proving Ground Test Site", *Detection and Remediation Technologies for Mines and Mine-Like Targets II*, Abinash C. Dubey, Robert L. Barnard, editors, Proc. SPIE V 3079, pp726-737, SPIE, Orlando, (1997)
3. M.A. LeCompte, F.J. Iannarilli, D.B. Nichols, R.R. Keever, "Multispectral IR Signature Polarimetry for Detection of Mines and Unexploded Ordnance (UXO)", *Detection Technologies for Mines and Mine-Like Targets*, Abinash C. Dubey, Ivan Cindrich, James M. Ralston, Kelly Rigano, editors, Proc. SPIE V 2496, pp180-192, SPIE, Orlando, (1995)
4. B.A. Barbour, S. Kordella, M.J. Dorset, B.L. Kerstiens, "Mine Detection Using a Polarimetric IR Sensor", *The Detection of Abandoned Landmines: a Humanitarian Imperative Seeking a Technical Solution*, EUREL International Conference, V 431, pp78-82, IEE, Edinburgh, (1996)
5. D.E. Lake, B. Sadler, S. Casey, "Detecting Regularity in Minefields Using Collinearity and a Modified Euclidean Algorithm", *Detection and Remediation Technologies for Mines and Mine-Like Targets II*, Abinash C. Dubey, Robert L. Barnard, editors, Proc SPIE V 3079, pp500-507, SPIE, Orlando (1997)
6. C.E. Priebe, T.E. Olson, D.M. Healy, "Exploiting Stochastic Partitions for Minefield Detection", *Detection and Remediation Technologies for Mines and Mine-Like Targets II*, Abinash C. Dubey, Robert L. Barnard, editors, Proc SPIE V 3079, pp508-518, SPIE, Orlando (1997)
7. N. Cressie, A.B. Lawson, "Models and Inference for Clustering of Locations of Mines and Mine-Like Objects", *Detection and Remediation Technologies for Mines and Mine-Like Targets II*, Abinash C. Dubey, Robert L. Barnard, editors, Proc SPIE V 3079, pp519-530, SPIE, Orlando (1997)
8. M.M. Hayat, J.A. Gubner, W. Chang, "Detection of Signals Modeled by Interaction Point Processes Using a Poisson Approximation", *Detection and Remediation Technologies for Mines and Mine-Like Targets II*, Abinash C. Dubey, Robert L. Barnard, editors, Proc SPIE V 3079, pp531-536, SPIE, Orlando (1997)
9. G. Winkler, *Image Analysis, Random Fields, and Dynamic Monte Carlo Methods: a Mathematical Introduction*, Chapters 3 and 7, pp531-536, Springer-Verlag, New York, (1995)
10. M.G. Bellow, "A Markov Random Field Based Anomaly Screening Algorithm", *Detection Technologies for Mines and Mine-Like Targets*, Abinash C. Dubey, Ivan Cindrich, James M. Ralston, Kelly Rigano, editors, Proc. SPIE V 2496, pp466-474, SPIE, Orlando, (1995)
11. X. Hua, J. Davidson, N. Cressie, "Mine Boundary Detection Using Markov Random Fields", *Detection Technologies for Mines and Mine-Like Targets*, Abinash C. Dubey, Ivan Cindrich, James M. Ralston, Kelly Rigano, editors, Proc. SPIE V 2496, pp626-636, SPIE, Orlando, (1995)
12. S. Geman, D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v PAMI-6 no 6, pp721-741, 1984

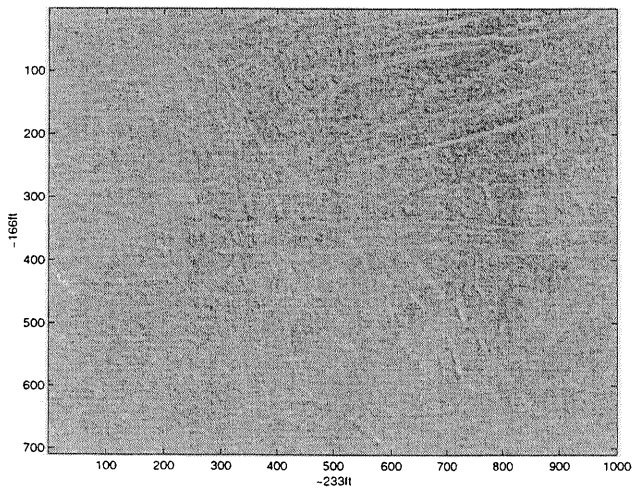
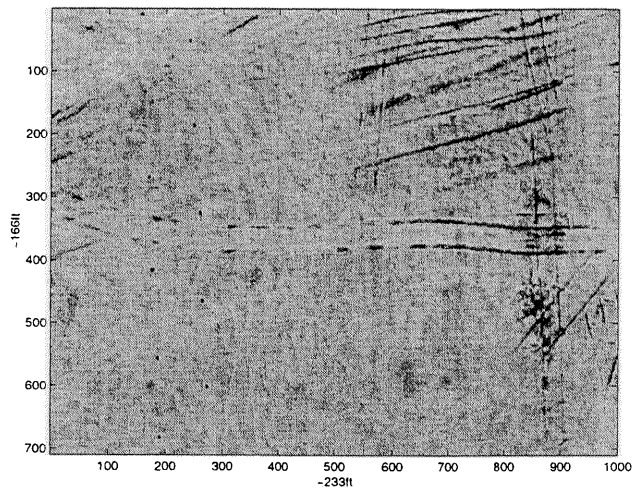
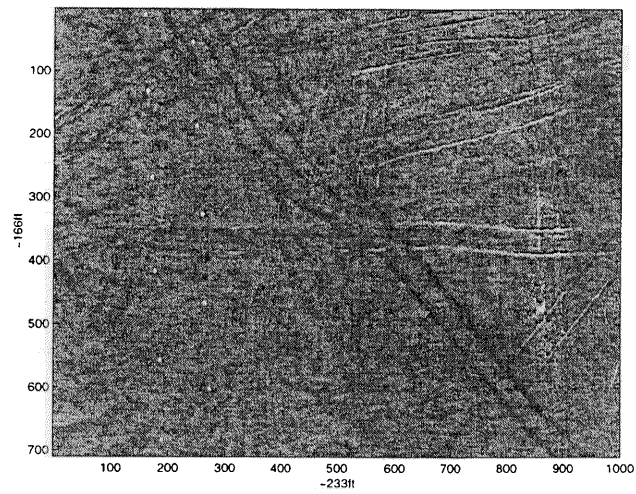


Figure 3. 9 am, Overcast, Percent Polarization, Reflectance, and Thermal

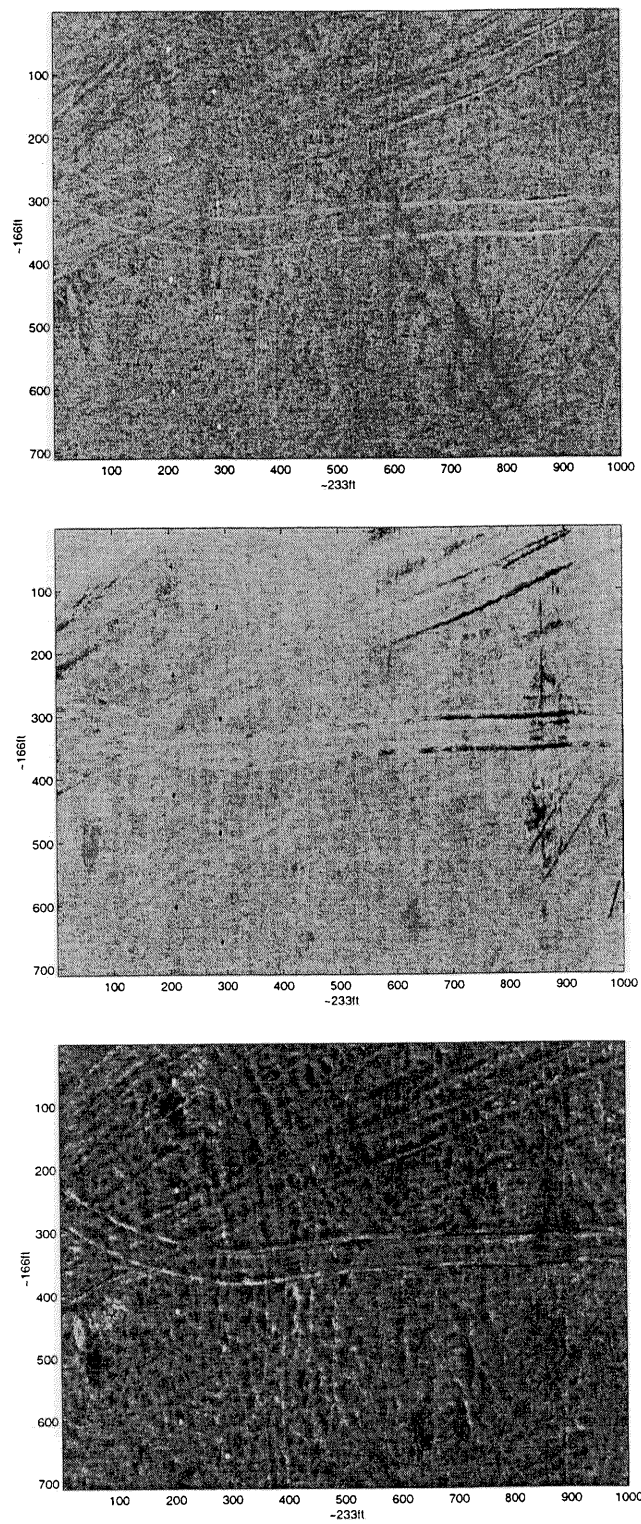


Figure 4. 3 pm, Sunny, Percent Polarization, Reflectance, and Thermal

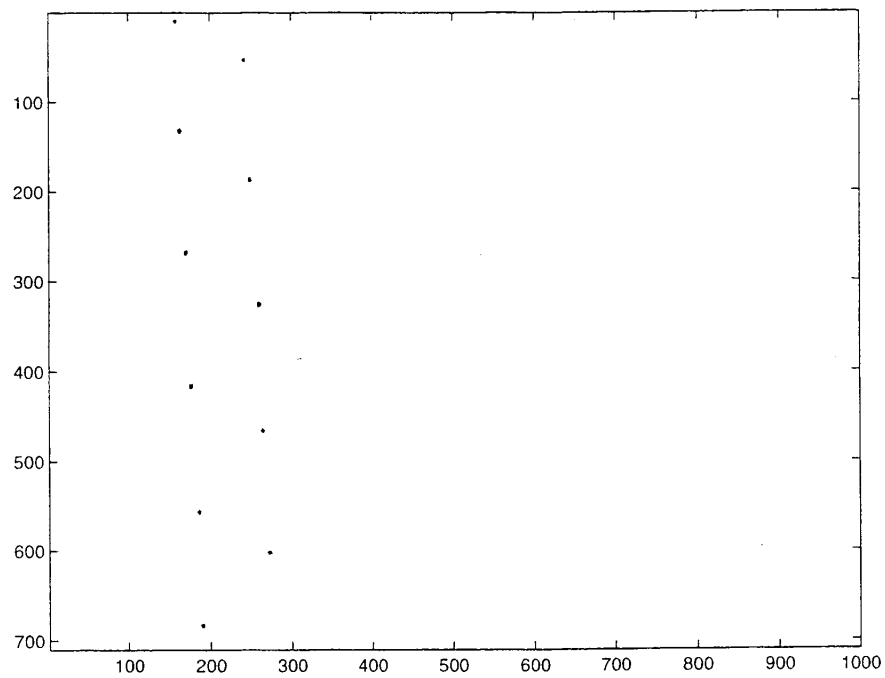


Figure 5. Typical Maximum Likelihood Result Based on Gaussian Model

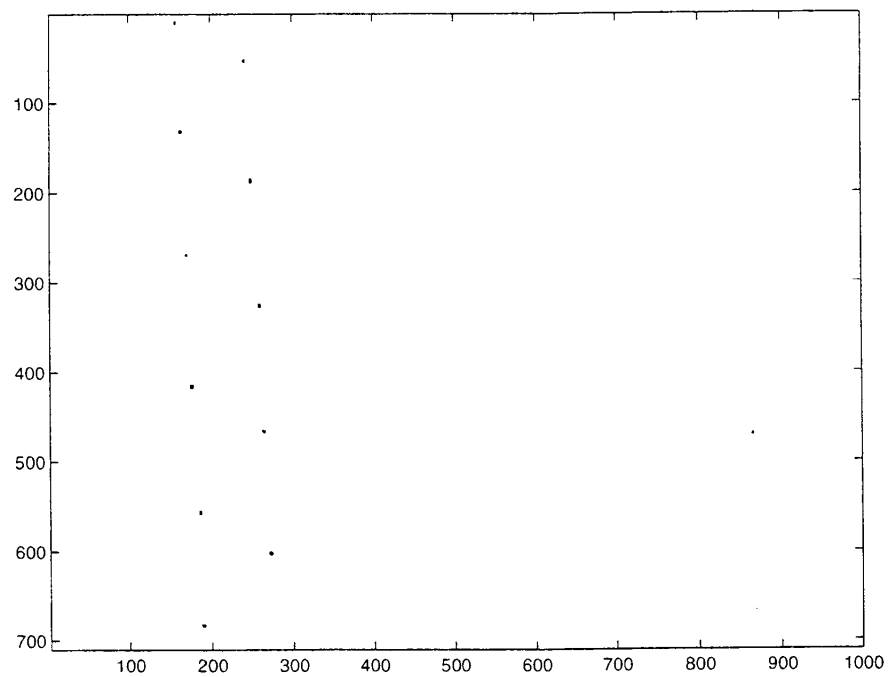


Figure 6. Typical MAP Detection Result with MRF Prior Model

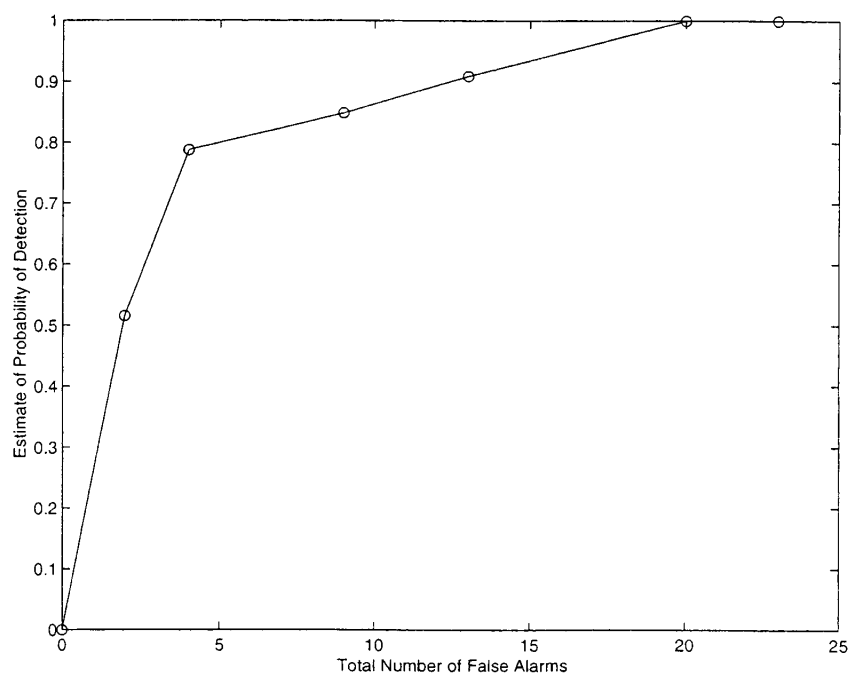


Figure 7. Experimental Receiver Operating Characteristic (ROC) for MAP Detector

Time-dependent plane wave and multipole expansions of the electromagnetic field

Edwin A. Marengo^{a)} and Anthony J. Devaney

*Center for Electromagnetics Research, Northeastern University,
Boston, Massachusetts 02115*

(Received 24 November 1997; accepted for publication 8 April 1998)

A new and conceptually simple derivation is presented of the time-dependent multipole expansion of the electromagnetic field radiated by a time-varying, localized, volume charge-current distribution. The analysis is based on a new time-dependent plane wave representation of the electromagnetic field (i.e., an “angular spectrum expansion” in the time domain), also derived in the paper. Expressions are given for the time-dependent plane wave spectra in terms of a fourfold Radon transform representation of the transverse current distribution. Two alternative expressions for the time-dependent multipole moments are derived; the first gives them in terms of the spectral amplitude vectors of the corresponding time-dependent plane wave representation while the second gives them in terms of a weighted radial-temporal average of the current distribution. Thus the analysis also sheds light on the relationship between the time-dependent plane wave and multipole expansions in their common domains of validity. © 1998 American Institute of Physics.
[S0022-2488(98)01007-X]

I. INTRODUCTION

In describing transient wave phenomena, descriptions directly in space-time domain (where the sources and fields are well-localized) are usually more efficient and insightful than their conventional, frequency domain counterparts. Examples include the pulsed beams (PBs) and the focus wave modes (FWMs), among others.^{1,2} This investigation is concerned with the time domain analogs of more traditional (frequency-domain-based) representations, namely, the plane wave and multipole expansions. Thus, the electromagnetic field radiated by a time-varying (e.g., pulsed), localized, volume charge-current distribution is expressed as a sum of time-dependent spherical (multipole) wave functions. We also obtain, via analytic signal representation, a new time-dependent plane wave representation of the electromagnetic field (i.e., an “angular spectrum expansion” in the time domain). The relationship between the expansion coefficients (moments) of the time-dependent plane wave and multipole expansions in their common domains of validity is also examined. The work reported here is an extension to the electromagnetic field case of the time domain multipole theory for scalar fields developed in Ref. 3. It is also an extension to the time domain of the frequency domain plane wave representations and multipole expansions of the electromagnetic field developed in Ref. 4.

Time-dependent multipolelike (spherical wave) expansions of electromagnetic radiation fields are not new. Early work in this area is associated with the names of Granzow,⁵ Davidon⁶ and Campbell *et al.*⁷ More recent contributions include those of Heyman and Devaney,³ where a scalar time-dependent multipole theory for volume source distributions is developed and Hansen,⁸ where time-dependent multipole fields are applied to address theoretical aspects of spherical near field scanning. However, even though work in this field is not new, both the procedure we use to derive our results as well as many of the results are new. For instance, the time domain multipole theories in Refs. 6 and 7 originate from the Debye-potential formalism⁴ while ours originates from the time-dependent plane wave representation, also derived in the paper. Granzow's⁵ and Hansen's⁸ focus is on the diffraction problem for prescribed field data over a boundary while ours is on the radiation problem for volume current distributions. Furthermore, our formulation is conceptually

^{a)}Electronic mail: emarengo@cdsp.neu.edu

simple and insightful in describing the relationship between the time-dependent expansion coefficients (moments) of the time-dependent plane wave and multipole expansions in their common domains of validity. Aspects of our time-dependent plane wave representations of the electromagnetic field have also appeared before in connection with time domain planar near field scanning.⁹ Our formulation, however, is given in terms of charge-current distributions (i.e., radiation problem), unlike those in Ref. 9, which are given in terms of fields prescribed over a boundary (i.e., diffraction problem). More importantly, our analysis not only corroborates those of previous contributions but also yields new, alternative expressions for both types of expansions.

Frequency domain plane wave and multipole expansions of scalar and electromagnetic fields have played a key role in classical and quantum-mechanical theories of wave radiation, propagation and diffraction and in defining fundamental limits in antenna theory.^{10,11} Thus, in analogy to their frequency domain counterparts, the time-dependent expansion schemes considered here could be useful in treating canonical radiation and scattering problems directly in the time domain. They could also be of interest in diffraction problems formulated in terms of near or far field data (e.g., time domain near field scanning⁸ and inverse diffraction from far field data⁴). Furthermore, they could also help in elucidating transient radiation phenomena (e.g., fundamental limitations of antennas driven by short pulses) and—by doing so—also in defining relevant figures of merit for transient radiation.

The remainder of the paper is organized as follows. Section II reviews known results on plane wave and multipole expansions for scalar fields. We examine, in analytic signal representation, the Weyl expansion (or angular spectrum expansion) of the scalar field generated by a localized scalar source distribution that is turned on at some initial time (say $t=0$). We then derive, again in analytic signal representation, a time domain multipole expansion of the scalar field based on the corresponding time-dependent plane wave representation. Methodologically, we Fourier invert to the time domain the corresponding frequency domain multipole expansion. The final form of the time-dependent multipole expansion is given explicitly in terms of real signals using the real spherical harmonics $S_{\ell,m}^{(1)}(\theta, \phi)$ and $S_{\ell,m}^{(2)}(\theta, \phi)$ (also known as tesseral harmonics¹²), defined, respectively, by the real and imaginary parts of the ordinary spherical harmonic of degree ℓ and order m , $Y_{\ell,m}(\theta, \phi)$. An alternative form of the time-dependent multipole expansion in which the time-dependent multipole moments are expressed as a weighted radial-temporal average of the source distribution is also derived.

The main contribution of the paper is contained in Sec. III. There we extend the theory developed in Sec. II for scalar fields to the electromagnetic field case, with the vector spherical harmonics in the latter case playing a role analogous to that of the ordinary spherical harmonics in the scalar case. In Sec. III A we derive, in analytic signal representation, a new time-dependent plane wave representation of the electromagnetic field generated by a time-varying, localized, volume charge-current distribution. The time-dependent plane wave spectra are given in terms of a fourfold Radon transform representation of the transverse current distribution. Later, we derive, following lines analogous to those of the scalar field treatment in Sec. II, a new form of time-dependent multipole expansion of the electromagnetic field; the spherical-wave expansion functions used are the vector analogs of the real spherical harmonics. However, unlike in Sec. II, our electromagnetic field case manipulations in Sec. III are carried out in the time domain. Two alternative expressions for the time-dependent multipole moments (i.e., the expansion coefficients) are derived; the first gives them in terms of the spectral amplitude vectors of the corresponding time-dependent plane wave representation while the second gives them in terms of a weighted radial-temporal average of the current distribution. Emphasis is placed on real spherical harmonics, hence our expansion involves only real time-dependent multipole wave functions and moments.

Notation: We define the temporal and spatial-temporal Fourier transforms, $q(\mathbf{r}, \omega)$ and $\tilde{q}(\mathbf{K}, \omega)$, respectively, of a scalar source $Q(\mathbf{r}, t)$ of finite energy [i.e., square-integrable in (\mathbf{r}, t)] via

$$q(\mathbf{r}, \omega) = \int_{-\infty}^{\infty} dt' e^{i\omega t'} Q(\mathbf{r}, t') \quad (1)$$

where

$$Q(\mathbf{r}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega' e^{-i\omega' t} q(\mathbf{r}, \omega'), \quad (2)$$

and

$$\tilde{q}(\mathbf{K}, \omega) = \int_{-\infty}^{\infty} dt' e^{i\omega t'} \int d\mathbf{r}' e^{-i\mathbf{K} \cdot \mathbf{r}'} Q(\mathbf{r}', t') \quad (3)$$

where

$$Q(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \int_{-\infty}^{\infty} d\omega' e^{-i\omega' t} \int d\mathbf{K}' e^{i\mathbf{K}' \cdot \mathbf{r}} \tilde{q}(\mathbf{K}', \omega'). \quad (4)$$

Thus the spatial-temporal Fourier transform $\tilde{q}(\mathbf{K}, \omega)$ of $Q(\mathbf{r}, t)$ is simply the spatial Fourier transform of $q(\mathbf{r}, \omega)$. We shall refer to $q(\mathbf{r}, \omega)$ as being the source spectrum.

Unless otherwise stated, we shall subsequently use lowercase and uppercase letters to denote frequency and time domain quantities, respectively. Unless otherwise stated, a tilde (\sim) over a constituent will denote a spatial-frequency domain constituent. A caret ($\hat{}$) over a vector will denote a unit vector, e.g., $\hat{\mathbf{r}} \equiv \mathbf{r}/|\mathbf{r}|$.

We shall denote the Cartesian coordinates of \mathbf{r} (configuration space) and \mathbf{K} (momentum space) as (x, y, z) and (K_x, K_y, K_z) , respectively. The corresponding spherical coordinates will be denoted as (r, θ, ϕ) and (K, α, β) , respectively. We also define the \mathbf{K} -space unit vector $\mathbf{s} \equiv \mathbf{K}/K$.

In order to simplify the manipulations time-dependent quantities will be manipulated first in analytic signal representation. For example, we define the analytic source distribution $Q^+(\mathbf{r}, t)$ corresponding to a real, square-integrable source distribution $Q(\mathbf{r}, t)$ by

$$Q^+(\mathbf{r}, t) = \frac{1}{2\pi} \int_0^{\infty} d\omega' e^{-i\omega' t} q(\mathbf{r}, \omega') \quad (5)$$

so that, for real t ,

$$Q(\mathbf{r}, t) = 2\Re\{Q^+(\mathbf{r}, t)\}, \quad (6)$$

where \Re denotes the real part. The quantity $Q^+(\mathbf{r}, t)$ defined by Eq. (5) is an analytic function in the lower-half plane of the complex t plane.¹³ Of particular interest to us is the real- t limit of $Q^+(\mathbf{r}, t)$, which defines the real, physical source distribution $Q(\mathbf{r}, t)$ via Eq. (6). Subsequently a plus sign (+) in the upper-right of a constituent will denote its analytic extension.

To simplify notation we also define a temporal integration operator ∂_t^{-1} such that

$$\partial_t^{-1} Q(\mathbf{r}, t) \equiv \int_{-\infty}^t dt' Q(\mathbf{r}, t'). \quad (7)$$

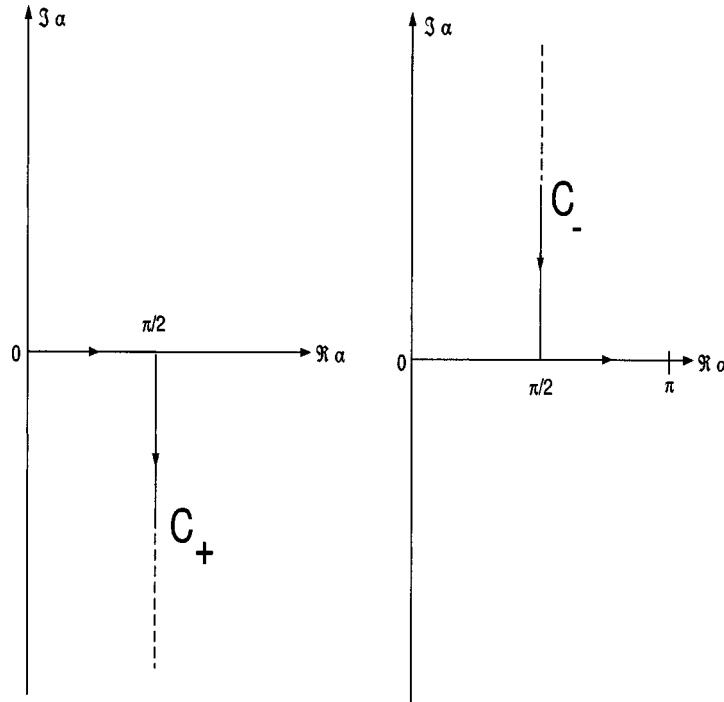
The corresponding n th-order temporal integration will be denoted as ∂_t^{-n} .

II. REVIEW OF THE THEORY FOR SCALAR FIELDS

This section reviews time-dependent plane wave and multipole expansions for scalar fields $U(\mathbf{r}, t)$ radiated by scalar sources $Q(\mathbf{r}, t)$ to the inhomogeneous wave equation

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) U(\mathbf{r}, t) = -4\pi Q(\mathbf{r}, t). \quad (8)$$

We will assume that the source $Q(\mathbf{r}, t)$ is a square-integrable, continuous and continuously differentiable function of (\mathbf{r}, t) , is confined to within a spherical volume about the origin [such that $Q(\mathbf{r}, t) = 0$ for $r > a$], and is turned on at some initial time, say $t = 0$ [such that $Q(\mathbf{r}, t) = 0$ for $t < 0$]. The radiated field $U(\mathbf{r}, t)$ is the causal solution to Eq. (8), given by¹⁴

FIG. 1. The α -contours of integration C_+ and C_- in complex α space.

$$U(\mathbf{r}, t) = \int_0^\infty dt' \int_{r' \leq a} d\mathbf{r}' \frac{Q(\mathbf{r}', t') \delta\left(t' - t + \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right)}{|\mathbf{r} - \mathbf{r}'|}, \quad (9)$$

where $\delta(\cdot)$ is Dirac's delta function.

A. Plane wave expansion of scalar fields in the time domain

It is well known that the radiated field $U(\mathbf{r}, t)$ at points \mathbf{r} within any half-space $|z| \geq a$ can be written as a superposition of monochromatic homogeneous and inhomogeneous (evanescent) plane waves; in particular,¹⁵

$$U(\mathbf{r}, t) = 2\Re\{U^+(\mathbf{r}, t)\} = 2\Re\left\{\frac{ic}{(2\pi)^2} \int_0^\infty \frac{\omega}{c} d\left(\frac{\omega}{c}\right) \int_{-\pi}^\pi d\beta \int_{C_\pm} d\alpha \sin \alpha f(\mathbf{s}, \omega) e^{[i(\omega/c)(\mathbf{s} \cdot \mathbf{r} - ct)]}\right\} \quad (10)$$

where \mathbf{s} is the unit vector having Cartesian components (in general complex) $(\sin \alpha \cos \beta, \sin \alpha \sin \beta, \cos \alpha)$ while the contours of integration C_\pm in the complex α plane are those shown in Fig. 1, where C_+ is used if $z \geq a$ and C_- if $z \leq -a$. The real-axis portion of the contours C_\pm contributes to the propagating spectrum, describing homogeneous plane waves propagating into the half-space $z \geq a$ or $z \leq -a$. The complex portion of the contours C_\pm corresponds, on the other hand, to the evanescent spectrum, describing evanescent plane waves that decay exponentially in amplitude along the z axis (i.e., in $\hat{\mathbf{z}}$ if $z \geq a$ and in $-\hat{\mathbf{z}}$ if $z \leq -a$). The spectral amplitude function $f(\mathbf{s}, \omega)$ in (10) is related to the source spectrum $q(\mathbf{r}, \omega)$ by¹⁵

$$f(\mathbf{s}, \omega) = \tilde{q}(\mathbf{K}, \omega)|_{\mathbf{K}=(\omega/c)\mathbf{s}} = \int_{r' \leq a} d\mathbf{r}' e^{-i(\omega/c)\mathbf{s} \cdot \mathbf{r}'} q(\mathbf{r}', \omega). \quad (11)$$

It is noted that since we require $q(\mathbf{r}, \omega)$ to be continuous and compactly supported in space, it follows that $\tilde{q}(\mathbf{K}, \omega)$ is, for fixed frequency ω , the boundary value, on the real K_x, K_y, K_z axes, of an entire analytic function of three complex variables K_x, K_y, K_z . Thus, for fixed frequency ω ,

$f(\mathbf{s}, \omega)$ is given by the value of the spatial Fourier transform $\tilde{q}(\mathbf{K}, \omega)$, and its analytic continuation into the complex \mathbf{K} plane, of the source spectrum $q(\mathbf{r}, \omega)$ evaluated at $\mathbf{K} = (\omega/c)\mathbf{s}$.

By performing the ω integration in (10) and using (5) we obtain

$$U^+(\mathbf{r}, t) = -\frac{1}{2\pi c} \frac{\partial}{\partial t} \int_{-\pi}^{\pi} d\beta \int_{C_{\pm}} d\alpha \sin \alpha F^+(\mathbf{s}, t - \mathbf{s} \cdot \mathbf{r}/c), \quad (12)$$

where from Eq. (11)

$$F^+(\mathbf{s}, t) \equiv \frac{1}{2\pi} \int_0^{\infty} d\omega' e^{-i\omega' t} f(\mathbf{s}, \omega') = \int_{r' \leq a} d\mathbf{r}' Q^+(\mathbf{r}', t + \mathbf{s} \cdot \mathbf{r}'/c). \quad (13)$$

Equations (12) and (13) define the time-dependent plane wave representation of the analytic signal $U^+(\mathbf{r}, t)$ corresponding to the real, physical field $U(\mathbf{r}, t)$.

Equation (13) is the time domain analog of (11). The operation defined by the integral on the right-hand side of Eq. (13), sometimes referred to as “slant-stack transform” of the analytic source distribution $Q^+(\mathbf{r}, t)$, has been studied in recent years by Heyman.¹⁶ It involves spatial Radon projection of $Q^+(\mathbf{r}, t)$ evaluated at planes normal to the spectral propagation direction \mathbf{s} and stacking them with a progressive time delay $\mathbf{s} \cdot \mathbf{r}/c$ associated with a wave front propagating at the speed of light along this direction. Mathematically, $F^+(\mathbf{s}, t)$ is given by the value of the fourfold Radon transform, and its analytic continuation for complex \mathbf{s} ,¹⁷ of the analytic source distribution evaluated at the hyperplane $ct - \mathbf{s} \cdot \mathbf{r} = 0$ in a four-dimensional space-time domain (ct, \mathbf{r}) .

The real, physical field $U(\mathbf{r}, t)$ is recovered from $U^+(\mathbf{r}, t)$ via Eqs. (6) and (5). In the far radiation zone¹⁴

$$U(r\hat{\mathbf{r}}, t)|_{r \rightarrow \infty} \sim \frac{1}{r} F(\mathbf{s}, \tau)|_{\mathbf{s}=\hat{\mathbf{r}}} \quad (14)$$

with $\tau = t - r/c$. Thus the time domain radiation pattern is from Eq. (14), the propagating portion of the time-dependent plane-wave spectrum. The evanescent spectrum can be recovered by analytic continuation of the time domain radiation pattern to complex-valued observation directions.¹⁷ The individual contributions of the propagating and evanescent spectra to the total field have been examined in detail by Heyman,¹⁶ who has shown that the contribution from the evanescent spectrum vanishes after a certain time t_E that may be determined *a priori* for a given field point; in regions of space where t_E occurs prior to the causal arrival time of the field, the latter is describable only by its propagating spectrum within the causal time window while noncausal contributions due to both propagating and evanescent spectra cancel out prior to the causal arrival time.

B. Multipole theory in the frequency domain

The multipole expansion of the radiated field spectrum $u(\mathbf{r}, \omega)$ is readily obtained by expanding the spectral amplitudes $f(\mathbf{s}, \omega)$ of the (frequency domain) angular spectrum expansion

$$u(\mathbf{r}, \omega) = \frac{i\omega/c}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_{\pm}} d\alpha \sin \alpha f(\mathbf{s}, \omega) e^{i(\omega/c)\mathbf{s} \cdot \mathbf{r}} \quad (15)$$

into a series of spherical harmonics $Y_{\ell, m}(\alpha, \beta)$ (see Appendix A for a review) so that

$$f(\mathbf{s}, \omega) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} (-i)^{\ell} \tilde{a}_{\ell, m}(\omega) Y_{\ell, m}(\alpha, \beta) \quad (16)$$

where

$$\tilde{a}_{\ell, m}(\omega) = i^{\ell} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha f(\mathbf{s}, \omega) Y_{\ell, m}^*(\alpha, \beta). \quad (17)$$

By substituting from Eqs. (16) and (17) into Eq. (15) while interchanging orders of integration and summation one readily finds

$$u(\mathbf{r}, \omega) = \frac{i\omega}{c} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \tilde{a}_{\ell,m}(\omega) \Pi_{\ell,m}(\mathbf{r}, \omega), \quad (18)$$

where

$$\Pi_{\ell,m}(\mathbf{r}, \omega) = (-i)^{\ell} \frac{1}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_{\pm}} d\alpha Y_{\ell,m}(\alpha, \beta) e^{i(\omega/c)\mathbf{s} \cdot \mathbf{r}} \quad (19)$$

is precisely the angular spectrum representation of the scalar multipole field of degree ℓ and order m , i.e.,⁴

$$\Pi_{\ell,m}(\mathbf{r}, \omega) = h_{\ell}^{(1)}\left(\frac{\omega}{c} r\right) Y_{\ell,m}(\theta, \phi). \quad (20)$$

In Eq. (20), $h_{\ell}^{(1)}(\cdot)$ is the spherical Hankel function of the first kind of order ℓ (as defined in Ref. 14). Even though Eqs. (18) and (20) were derived for $\omega > 0$, they apply to all real ω . This fact is implicit in our time-dependent multipole treatment (Sec. II C). For example, the operator $\mathcal{L}_{\ell}^{(r,t)}$ in Eq. (24) applies to both analytic and real signals.

The multipole expansion Eq. (18) is valid for $r \geq a$ while the plane wave expansion Eq. (15) is valid for $|z| \geq a$. However, it is obvious that as long as \mathbf{r} is outside the source, one can always decompose the radiated field spectrum $u(\mathbf{r}, \omega)$ in the form Eq. (15) by a suitable choice of coordinates.¹⁶

An alternative expression for the multipole moments $\tilde{a}_{\ell,m}(\omega)$ is obtained by substituting (11) on (17) and interchanging orders of integration so as to yield⁴

$$\tilde{a}_{\ell,m}(\omega) = 4\pi \int_{r' \leq a} d\mathbf{r}' q(\mathbf{r}', \omega) \Lambda_{\ell,m}^*(\mathbf{r}', \omega) \quad (21)$$

with

$$\Lambda_{\ell,m}(\mathbf{r}, \omega) \equiv j_{\ell}\left(\frac{\omega}{c} r\right) Y_{\ell,m}(\theta, \phi) = (-i)^{\ell} \frac{1}{4\pi} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha Y_{\ell,m}(\alpha, \beta) e^{i(\omega/c)\mathbf{s} \cdot \mathbf{r}}, \quad (22)$$

where $j_{\ell}(\cdot)$ is the spherical Bessel function of the first kind and order ℓ (as defined in Ref. 4).

The multipole expansion enjoys several advantages over the angular spectrum expansion. First of all, we see from Eq. (17) that the multipole moments $\tilde{a}_{\ell,m}(\omega)$ are defined only by the real \mathbf{s} (propagating) spectral amplitudes $f(\mathbf{s}, \omega)$. Thus, in contrast to the angular spectrum expansion—which requires analytic continuation—the multipole expansion requires no analytic continuation of $f(\mathbf{s}, \omega)$ for complex values of the polar angle α . In addition, unlike the plane waves $e^{i(\omega/c)\mathbf{s} \cdot \mathbf{r}}$, the multipole fields $\Pi_{\ell,m}(\mathbf{r}, \omega)$ obey Sommerfeld's radiation condition. Hence, the multipole expansion is, unlike the angular spectrum expansion, a true mode expansion of the radiation field spectrum. This can be of interest if one wishes to approximate the radiated field either by truncating the series (18) or by limiting the (α, β) region of integration in (15). In the former case the approximation will obey Sommerfeld's radiation condition while in the latter it will not. These advantages are obtained, however, at the expense of the increased complexity of the spherical wave fields $\Pi_{\ell,m}(\mathbf{r}, \omega)$ as compared to the plane waves $e^{i(\omega/c)\mathbf{s} \cdot \mathbf{r}}$.

C. Time-dependent multipole expansion of the scalar field

Next we obtain a time domain multipole expansion for the scalar field $U(\mathbf{r}, t)$ following the lines of Ref. 3. Our starting point is the result [see Ref. 18, Eq. (10.1.16)]

$$\frac{1}{2\pi} \int_0^\infty d\omega e^{-i\omega t} h_\ell^{(1)}\left(\frac{\omega}{c} r\right) \tilde{a}_{\ell,m}(\omega) = -(-i)^\ell \frac{c}{r} \partial_t^{-1} \mathcal{L}_{\ell}^{(r,t)} a_{\ell,m}^+(t), \quad (23)$$

where

$$a_{\ell,m}^+(t) = \frac{1}{2\pi} \int_0^\infty d\omega e^{-i\omega t} \tilde{a}_{\ell,m}(\omega)$$

and

$$\mathcal{L}_{\ell}^{(r,t)} a_{\ell,m}^+(t) = \sum_{n=0}^{\ell} \frac{(\ell+n)!}{n!(\ell-n)!} \left(\frac{2r}{c}\right)^{-n} \partial_t^{-n} a_{\ell,m}^+(\tau). \quad (24)$$

In Eqs. (23) and (24) $\tau = t - r/c$ (as before) and ∂_t^{-n} denotes n th-order temporal integration [recall (7)].

By Fourier inverting Eq. (18) over positive frequencies while using Eqs. (20) and (23) we obtain

$$U^+(\mathbf{r}, t) = \frac{1}{r} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} (-i)^\ell Y_{\ell,m}(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} a_{\ell,m}^+(t). \quad (25)$$

The analytic time-dependent multipole moments $a_{\ell,m}^+(t)$ are obtained by Fourier inverting over positive frequencies both sides of Eq. (17), thus giving

$$a_{\ell,m}^+(t) = i^\ell \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha F^+(\mathbf{s}, t) Y_{\ell,m}^*(\alpha, \beta). \quad (26)$$

Thus, for fixed t , the analytic time-dependent multipole moments $a_{\ell,m}^+(t)$ are, apart from the factor i^ℓ , given by projections of the analytic time-dependent plane wave spectrum $F^+(\mathbf{s}, t)$ [see Eq. (13)] onto the set of spherical harmonics $Y_{\ell,m}(\alpha, \beta)$. Note, however, that—as expected from our discussion in Eq. (17)—only the real- \mathbf{s} , propagating plane wave spectral components enter into play in Eq. (26).

An alternative form of Eq. (25) due to Hansen and Norris¹⁹ is obtained using (see Ref. 19 and references therein)

$$h_\ell^{(1)}\left(\frac{\omega}{c} r\right) Y_{\ell,m}(\theta, \phi) = i^{-\ell} \tilde{P}_{\ell,m}^\omega \frac{e^{i(\omega/c)r}}{i(\omega/c)r}, \quad (27)$$

where

$$\tilde{P}_{\ell,m}^\omega = (-1)^m \sqrt{\frac{(2\ell+1)(\ell-m)!}{4\pi(\ell+m)!}} \left(\frac{1}{i(\omega/c)} \frac{\partial}{\partial x} + i \frac{1}{\omega} \frac{\partial}{\partial y} \right)^m P_\ell^{(m)} \left(\frac{1}{i} \frac{\omega}{c} \frac{\partial}{\partial z} \right), \quad m \geq 0, \quad (28)$$

where $P_\ell^{(m)}(\cdot)$ is the m th derivative of the Legendre polynomial and for $m < 0$, $\tilde{P}_{\ell,-m}^\omega = (-1)^m \tilde{P}_{\ell,m}^{\omega*}$. On substituting from Eqs. (27) and (28) into Eqs. (18) and (20) while using the complex-point source field expansion of the real-point source field [see Eq. (16) in Ref. 19]

$$\frac{e^{i(\omega/c)r}}{4\pi r} = \frac{1}{4\pi j_0(i(\omega/c)\rho)} \int_{-\pi}^{\pi} d\phi' \int_0^{\pi} d\theta' \sin \theta' \frac{e^{i(\omega/c)\xi(\theta', \phi')}}{4\pi \xi(\theta', \phi')}, \quad (29)$$

where $\xi = |\mathbf{r} - i\rho \hat{\mathbf{r}}'| = \xi_r + i\xi_i$ is the complex length having real and imaginary parts ξ_r and ξ_i , respectively, where $\rho \leq a$ is a positive constant (such that $\xi_r \geq 0$ and $-\rho \leq \xi_i \leq \rho$), and transforming to the time domain one obtains

$$U^+(\mathbf{r}, t) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} (-i)^{\ell} \int_{-\pi}^{\pi} d\phi' \int_0^{\pi} d\theta' \sin \theta' \tilde{L}_0 \tilde{P}_{\ell m}^t \frac{a_{\ell, m}^+(t - \xi(\theta', \phi')/c)}{4\pi \xi(\theta', \phi')}, \quad (30)$$

where \tilde{L}_0 is a time domain operator whose frequency domain representation is $1/j_0((i\omega/c)\rho)$ [see Eq. (37) in Ref. 19] while $\tilde{P}_{\ell m}^t$ is the time domain analog of $\tilde{P}_{\ell m}^{\omega}$. Equation (30) is the result Eq. (59) in Ref. 19; it is an analytic field expansion in terms of time-dependent multipoles with complex-source locations. Unlike in Eq. (25), in Eq. (30) the basis functions are directionally localized; hence they are of interest in well-collimated wave radiation and propagation (see, e.g., Heyman and Felsen¹). However, the real-signal counterpart of the analytic complex-source multipole expansion Eq. (30) is complicated and involves convolutions; in contrast, the real-signal version of Eq. (25) is straightforward and can be written in terms of the operator $\mathcal{L}_{\ell}^{(r, t)}$. This is done below.

We use (25) to recover the real, physical field $U(\mathbf{r}, t) = 2\Re\{U^+(\mathbf{r}, t)\}$. In order to evaluate the real part of expression (25) we consider the real spherical harmonics $S_{\ell, m}^{(j)}$, with $\ell = 0, 1, \dots, m = 0, 1, \dots, \ell$ (see Appendix A). They obey the following orthogonality and completeness properties:³

$$\int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha S_{\ell, m}^{(j)}(\alpha, \beta) S_{\ell', m'}^{(j')}(\alpha, \beta) = \delta_{\ell, \ell'} \delta_{m, m'} \delta_{j, j'} \frac{1}{c_m}, \quad (31)$$

where $c_m = 1$ or 2 for $m = 0$ or $m \geq 1$, respectively, and

$$\sum_{\ell=1}^{\infty} \sum_{m=0}^{\ell} \sum_{j=1,2} c_m S_{\ell, m}^{(j)}(\alpha, \beta) S_{\ell', m'}^{(j')}(\alpha', \beta') = \frac{1}{\sin \alpha} \delta(\beta - \beta') \delta(\alpha - \alpha') \equiv \delta(\mathbf{s} - \mathbf{s}'). \quad (32)$$

[In Eq. (32), α' and β' are, respectively, the polar and azimuthal coordinates of the unit vector \mathbf{s}' .]

By taking the real part to both sides of Eq. (25) while using Eqs. (26), (32), and (31) one obtains (see derivation in Appendix B)

$$U(\mathbf{r}, t) = \frac{1}{r} \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} \sum_{j=1,2} c_m S_{\ell, m}^{(j)}(\theta, \phi) \mathcal{L}_{\ell}^{(r, t)} q_{\ell, m}^{(j)}(t), \quad (33)$$

where we have introduced the new (real) multipole moments $q_{\ell, m}^{(j)}(t)$, defined via

$$q_{\ell, m}^{(j)}(t) = \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha F(\mathbf{s}, t) S_{\ell, m}^{(j)}(\alpha, \beta). \quad (34)$$

On comparing Eqs. (14) and (33) with $\mathcal{L}_{\ell}^{(r, t)}$ given by Eq. (24), the far field is seen to correspond to the lowest order term (i.e., $n = 0$) in the series expansion Eq. (24) corresponding to $\mathcal{L}_{\ell}^{(r, t)}$. Thus from Eq. (24) $\mathcal{L}_{\ell}^{(r, t)} q_{\ell, m}^{(j)}(t)|_{r \rightarrow \infty} \sim q_{\ell, m}^{(j)}(\tau)$. Using this result in Eq. (33) yields

$$U(\mathbf{r}, t)|_{r \rightarrow \infty} \sim \frac{1}{r} \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} \sum_{j=1,2} c_m S_{\ell, m}^{(j)}(\theta, \phi) q_{\ell, m}^{(j)}(\tau). \quad (35)$$

Also, from (14) and (35), the time domain radiation pattern reduces to

$$F(\hat{\mathbf{r}}, t) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} \sum_{j=1,2} c_m S_{\ell, m}^{(j)}(\theta, \phi) q_{\ell, m}^{(j)}(t). \quad (36)$$

Since the multipole moments $q_{\ell, m}^{(j)}(t)$ define via Eq. (33) the field everywhere outside the source, it follows from Eq. (34) that the time domain radiation pattern $F(\hat{\mathbf{r}}, t)$ also determines the field everywhere outside the source (problem of inverse diffraction from far field data^{3,4}).

Equation (34) defines the real time-dependent multipole moments $q_{\ell, m}^{(j)}(t)$ as projections of the time-dependent plane wave spectrum onto the set of real spherical harmonics. An alternative

formula for $q_{\ell,m}^{(j)}(t)$ based on a radially dependent weighted time average of $Q(\mathbf{r},t)$ follows readily from the time domain analog of Eq. (21). The remainder of this section is based on Eq. (21) and the following result:³

$$j_{\ell}(z) = \frac{(z/2)^{\ell}}{2^{\ell}\ell!} \int_{-1}^1 d\eta (1-\eta^2)^{\ell} e^{iz\eta}. \quad (37)$$

We begin by substituting from Eq. (37) into Eq. (21), thus obtaining

$$\tilde{a}_{\ell,m}(\omega) = \frac{4\pi c^{-\ell} \omega^{\ell}}{\ell! 2^{\ell+1}} \int_{r \leq a} d\mathbf{r} q(\mathbf{r}, \omega) Y_{\ell,m}^*(\theta, \phi) r^{\ell} \int_{-1}^1 d\eta (1-\eta^2)^{\ell} e^{-i\omega r \eta/c}. \quad (38)$$

The analytic time-dependent multipole moments $a_{\ell,m}^{+}(t)$ are obtained by Fourier inverting $\tilde{a}_{\ell,m}(\omega)$ as given by Eq. (38) over positive frequencies. This procedure yields

$$a_{\ell,m}^{+}(t) = \frac{1}{2\pi} \int_0^{\infty} d\omega e^{-i\omega t} \tilde{a}_{\ell,m}(\omega) = i^{\ell} \int_{-\pi}^{\pi} d\phi \int_0^{\pi} d\theta \sin \theta Y_{\ell,m}^*(\theta, \phi) F_{\ell}^{+}(\hat{\mathbf{r}}, t), \quad (39)$$

where

$$F_{\ell}^{+}(\hat{\mathbf{r}}, t) = \frac{4\pi c^{-\ell}}{\ell! 2^{\ell+1}} \int_{r \leq a} dr r^{\ell} \int_{-1}^1 d\eta (1-\eta^2)^{\ell} \frac{\partial^{\ell}}{\partial t^{\ell}} Q^{+}(\mathbf{r}, t + r\eta/c). \quad (40)$$

The real time-dependent multipole moments $q_{\ell,m}^{(j)}(t)$ follow from Eqs. (39) and (40) with $q_{\ell,m}^{+}(t) = (-i)^{\ell} a_{\ell,m}^{+}(t)$:

$$q_{\ell,m}^{(j)}(t) = \int_{-\pi}^{\pi} d\phi \int_0^{\pi} d\theta \sin \theta S_{\ell,m}^{(j)}(\theta, \phi) F_{\ell}(\hat{\mathbf{r}}, t), \quad (41)$$

where

$$F_{\ell}(\hat{\mathbf{r}}, t) = \frac{4\pi c^{-\ell}}{\ell! 2^{\ell+1}} \int_{r \leq a} dr r^{\ell} \int_{-1}^1 d\eta (1-\eta^2)^{\ell} \frac{\partial^{\ell}}{\partial t^{\ell}} Q(\mathbf{r}, t + r\eta/c). \quad (42)$$

The time-dependent real multipole moments $q_{\ell,m}^{(j)}(t)$ are now given from Eqs. (41) and (42) by the projection of an ℓ -dependent (effective) time domain radiation pattern $F_{\ell}(\hat{\mathbf{r}}, t)$ onto the set of real spherical harmonics. This latter quantity is computed using a two-step procedure described below.

The first step is that of η integration in (42). It involves radially dependent weighted time averaging of the ℓ th time derivative of the source distribution, $(\partial^{\ell}/\partial t^{\ell})Q(\mathbf{r}, t)$ (η integration plays the role of temporal integration). The term $(1-\eta^2)^{\ell}$ in Eq. (42) acts as weighting kernel. The η -integration limits in Eq. (42) define the time window $t \pm r/c$ within which weighted time averaging of $(\partial^{\ell}/\partial t^{\ell})Q(\mathbf{r}, t)$ is carried out. This first step acts on both the temporal and radial dependence of the source. The temporal nature of the source enters into play here through both time derivation and time averaging. On the other hand, both ℓ and r (and, therefore, also the source's radius a) control the effective time averaging window.

The second step is that of radial integration. It acts only on the radial dependence of the source. Thus it plays a role in controlling the relative relevance of the outer versus the inner source's structure in defining $q_{\ell,m}^{(j)}(t)$.

III. TIME-DEPENDENT PLANE WAVE AND MULTIPOLE EXPANSIONS OF ELECTROMAGNETIC FIELDS

In Gaussian systems of units, the time-dependent electric and magnetic field vectors $\mathcal{E}(\mathbf{r}, t)$ and $\mathcal{H}(\mathbf{r}, t)$, respectively, generated in free space by a time-dependent current distribution $\mathcal{J}(\mathbf{r}, t)$, are the causal solutions to the vector wave equations¹⁴

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) \mathcal{E}(\mathbf{r}, t) = -4\pi \left[-\frac{1}{c^2} \frac{\partial}{\partial t} + \partial_t^{-1} \nabla \cdot \right] \mathcal{T}(\mathbf{r}, t), \quad (43)$$

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) \mathcal{H}(\mathbf{r}, t) = -4\pi \frac{1}{c} \nabla \times \mathcal{T}(\mathbf{r}, t),$$

where, again, ∂_t^{-1} denotes integration over time [see Eq. (7)]. Here and henceforth it is assumed that $\mathcal{T}(\mathbf{r}, t)$ is a square-integrable, continuous and continuously differentiable function of (\mathbf{r}, t) , is confined to a finite region of radius a around the origin, and is turned on at some initial time $t = 0$. In Secs. III A and III B we derive new time-dependent plane wave and multipole expansions for $\mathcal{E}(\mathbf{r}, t)$ and $\mathcal{H}(\mathbf{r}, t)$ in terms of $\mathcal{T}(\mathbf{r}, t)$.

A. Plane wave expansion of electromagnetic fields in the time domain

It follows from (43) that each Cartesian component of the electromagnetic field vectors $\mathcal{E}(\mathbf{r}, t)$, $\mathcal{H}(\mathbf{r}, t)$ satisfies an inhomogeneous scalar wave equation of the form Eq. (8). Therefore, the results of Sec. II are applicable to each Cartesian component of $\mathcal{E}(\mathbf{r}, t)$ and $\mathcal{H}(\mathbf{r}, t)$. Thus, we introduce the analytic field vectors $\mathcal{E}^+(\mathbf{r}, t) = (1/2\pi) \int_0^\infty d\omega e^{-i\omega t} \mathbf{E}(\mathbf{r}, \omega)$, $\mathcal{H}^+(\mathbf{r}, t) = (1/2\pi) \int_0^\infty d\omega e^{-i\omega t} \mathbf{H}(\mathbf{r}, \omega)$, where $\mathbf{E}(\mathbf{r}, \omega)$ and $\mathbf{H}(\mathbf{r}, \omega)$ are, respectively, the electric and magnetic field spectrum. They are related to the analytic current distribution $\mathcal{T}^+(\mathbf{r}, t) = (1/2\pi) \int_0^\infty d\omega e^{-i\omega t} \mathbf{J}(\mathbf{r}, \omega)$ by Eq. (43) with \mathcal{E} , \mathcal{H} and \mathcal{T} substituted by \mathcal{E}^+ , \mathcal{H}^+ and \mathcal{T}^+ , respectively.

By applying to Eq. (43) the scalar case result Eq. (12) one obtains

$$\mathcal{E}^+(\mathbf{r}, t) = -\frac{1}{2\pi c} \frac{\partial}{\partial t} \int_{-\pi}^{\pi} d\beta \int_{C_{\pm}} d\alpha \sin \alpha \mathbf{F}_E^+(\mathbf{s}, t - \mathbf{s} \cdot \mathbf{r}/c), \quad (44)$$

$$\mathcal{H}^+(\mathbf{r}, t) = -\frac{1}{2\pi c} \frac{\partial}{\partial t} \int_{-\pi}^{\pi} d\beta \int_{C_{\pm}} d\alpha \sin \alpha \mathbf{F}_H^+(\mathbf{s}, t - \mathbf{s} \cdot \mathbf{r}/c),$$

where from Eq. (13)

$$\mathbf{F}_E^+(\mathbf{s}, t) = \int_{r' \leq a} d\mathbf{r}' \left[-\frac{1}{c^2} \frac{\partial}{\partial t} + \partial_t^{-1} \nabla' \cdot \right] \mathcal{T}^+(\mathbf{r}', t + \mathbf{s} \cdot \mathbf{r}'/c), \quad (45)$$

$$\mathbf{F}_H^+(\mathbf{s}, t) = \frac{1}{c} \int_{r' \leq a} d\mathbf{r}' \nabla' \times \mathcal{T}^+(\mathbf{r}', t + \mathbf{s} \cdot \mathbf{r}'/c),$$

with ∇' operating on \mathbf{r}' (∇' does not act on $t + \mathbf{s} \cdot \mathbf{r}'/c$). The results Eqs. (44) and (45) above define a new time-dependent plane wave representation for electromagnetic fields whose interpretation is analogous to that of its scalar counterpart Eqs. (12) and (13).

Next we would like to express Eq. (45) more compactly in terms of the slant-stack transform

$$\hat{\mathcal{T}}^+(\mathbf{s}, t) = \int_{r' \leq a} d\mathbf{r}' \mathcal{T}^+(\mathbf{r}', t + \mathbf{s} \cdot \mathbf{r}'/c) \quad (46)$$

of the analytic time-dependent current distribution \mathcal{T}^+ . By manipulating the integrals of Eq. (45) while using (46) we obtain (see derivation in Appendix C)

$$\mathbf{F}_E^+(\mathbf{s}, t) = \frac{1}{c^2} \frac{\partial}{\partial t} \{ \mathbf{s} \times [\mathbf{s} \times \hat{\mathcal{T}}^+(\mathbf{s}, t)] \}, \quad (47)$$

$$\mathbf{F}_H^+(\mathbf{s}, t) = -\frac{1}{c^2} \frac{\partial}{\partial t} [\mathbf{s} \times \hat{\mathcal{T}}^+(\mathbf{s}, t)].$$

We see from Eq. (47) that $\mathbf{F}_E^+(\mathbf{s}, t)$ and $\mathbf{F}_H^+(\mathbf{s}, t)$ are defined by the transverse part (relative to \mathbf{s})—and its analytic continuation to complex \mathbf{s} —of $\hat{\mathcal{T}}^+(\mathbf{s}, t)$. We see also from Eq. (47) that

$$\mathbf{s} \cdot \mathbf{F}_E^+(\mathbf{s}, t) = \mathbf{s} \cdot \mathbf{F}_H^+(\mathbf{s}, t) = 0, \quad (48)$$

$$\mathbf{F}_E^+(\mathbf{s}, t) = -\mathbf{s} \times \mathbf{F}_H^+(\mathbf{s}, t).$$

Hence, the terms in the integrands of (44) represent—for each \mathbf{s} —electromagnetic plane waves, as expected.

We conclude this section by applying the scalar case result Eq. (14) to each Cartesian component of the field vectors, thus obtaining from (47) the far zone approximations

$$\mathcal{E}(r\hat{\mathbf{r}}, t)|_{r \rightarrow \infty} \sim \frac{1}{r} \mathbf{F}_E(\hat{\mathbf{r}}, \tau)|_{r \rightarrow \infty} = \frac{1}{c^2 r} \frac{\partial}{\partial t} \{ \hat{\mathbf{r}} \times [\hat{\mathbf{r}} \times \hat{\mathcal{T}}(\hat{\mathbf{r}}, \tau)] \}, \quad (49)$$

$$\mathcal{H}(r\hat{\mathbf{r}}, t)|_{r \rightarrow \infty} \sim \frac{1}{r} \mathbf{F}_H(\hat{\mathbf{r}}, \tau)|_{r \rightarrow \infty} = -\frac{1}{c^2 r} \frac{\partial}{\partial t} [\hat{\mathbf{r}} \times \hat{\mathcal{T}}(\hat{\mathbf{r}}, \tau)]$$

where, as before, $\tau = t - r/c$. Expressions (49) have been derived before in connection with questions of realizability of time domain radiation patterns¹⁷ and in time domain antenna characterization.²⁰

B. Time-dependent multipole expansion of the electromagnetic field

Next we will obtain an expansion of the analytic time-dependent electric and magnetic field vectors $\mathcal{E}^+(\mathbf{r}, t)$ and $\mathcal{H}^+(\mathbf{r}, t)$ in terms of vector spherical wave functions. Methodologically, first we will expand the analytic time-dependent plane wave spectra $[\mathbf{F}_E^+(\mathbf{s}, t)$ and $\mathbf{F}_H^+(\mathbf{s}, t)]$ into a series of the two sets of vector functions $\mathbf{Y}_{\ell, m}(\alpha, \beta)$ and $\mathbf{s} \times \mathbf{Y}_{\ell, m}(\alpha, \beta)$; later we will substitute the vector spherical harmonics expansions for $\mathbf{F}_E^+(\mathbf{s}, t)$ and $\mathbf{F}_H^+(\mathbf{s}, t)$ into the time-dependent plane wave expansions—Eq. (44). The first step above yields

$$\begin{aligned} \mathbf{F}_E^+(\mathbf{s}, t) &= \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} (-i)^{\ell} [a_{\ell, m}^+(t) \mathbf{s} \times \mathbf{Y}_{\ell, m}(\alpha, \beta) + b_{\ell, m}^+(t) \mathbf{Y}_{\ell, m}(\alpha, \beta)], \\ \mathbf{F}_H^+(\mathbf{s}, t) &= \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} (-i)^{\ell} [-a_{\ell, m}^+(t) \mathbf{Y}_{\ell, m}(\alpha, \beta) + b_{\ell, m}^+(t) \mathbf{s} \times \mathbf{Y}_{\ell, m}(\alpha, \beta)]. \end{aligned} \quad (50)$$

It is noted that the forms of the expansions above are consistent with Eq. (48).

By making use of the orthogonality condition of the vector spherical harmonics, Eq. (A8), we obtain from Eq. (50)

$$\begin{aligned} a_{\ell, m}^+(t) &= -\frac{i^{\ell}}{\ell(\ell+1)} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha \mathbf{F}_H^+(\mathbf{s}, t) \cdot \mathbf{Y}_{\ell, m}^*(\alpha, \beta), \\ b_{\ell, m}^+(t) &= \frac{i^{\ell}}{\ell(\ell+1)} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha \mathbf{F}_E^+(\mathbf{s}, t) \cdot \mathbf{Y}_{\ell, m}^*(\alpha, \beta). \end{aligned} \quad (51)$$

Thus the analytic time-dependent multipole moments $a_{\ell, m}^+(t)$ and $b_{\ell, m}^+(t)$ are defined by the projections of $\mathbf{F}_H^+(\mathbf{s}, t)$ and $\mathbf{F}_E^+(\mathbf{s}, t)$, respectively, onto the set of vector spherical harmonics.

The time-dependent multipoles can also be expressed directly in terms of the slant-stack transform of the current distribution by substituting from Eq. (47) into Eq. (51), thus giving

$$a_{\ell,m}^+(t) = \frac{i^\ell c^{-2}}{\ell(\ell+1)} \frac{\partial}{\partial t} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha [\mathbf{s} \times \hat{\mathcal{T}}^+(\mathbf{s}, t)] \cdot \mathbf{Y}_{\ell,m}^*(\alpha, \beta), \quad (52)$$

$$b_{\ell,m}^+(t) = \frac{i^\ell c^{-2}}{\ell(\ell+1)} \frac{\partial}{\partial t} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha \{ \mathbf{s} \times [\mathbf{s} \times \hat{\mathcal{T}}^+(\mathbf{s}, t)] \} \cdot \mathbf{Y}_{\ell,m}^*(\alpha, \beta).$$

Next we insert the series expansions Eq. (50) into the time-dependent plane wave expansions of the field vectors $\mathcal{E}(\mathbf{r}, t)$ and $\mathcal{H}(\mathbf{r}, t)$ [refer to Eq. (44)] and interchange orders of summation and integration, thus finding

$$\begin{aligned} \mathcal{E}^+(\mathbf{r}, t) = & -\frac{1}{2\pi c} \frac{\partial}{\partial t} \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} \int_{-\pi}^{\pi} d\beta \int_{C_{\pm}} d\alpha \sin \alpha (-i)^\ell [a_{\ell,m}^+(t - \mathbf{s} \cdot \mathbf{r}/c) \mathbf{s} \times \mathbf{Y}_{\ell,m}(\alpha, \beta) \\ & + b_{\ell,m}^+(t - \mathbf{s} \cdot \mathbf{r}/c) \mathbf{Y}_{\ell,m}(\alpha, \beta)], \quad (53) \\ \mathcal{H}^+(\mathbf{r}, t) = & -\frac{1}{2\pi c} \frac{\partial}{\partial t} \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} \int_{-\pi}^{\pi} d\beta \int_{C_{\pm}} d\alpha \sin \alpha (-i)^\ell [-a_{\ell,m}^+(t - \mathbf{s} \cdot \mathbf{r}/c) \mathbf{Y}_{\ell,m}(\alpha, \beta) \\ & + b_{\ell,m}^+(t - \mathbf{s} \cdot \mathbf{r}/c) \mathbf{s} \times \mathbf{Y}_{\ell,m}(\alpha, \beta)]. \end{aligned}$$

Next we shall solve the integrals over \mathbf{s} above so as to obtain the sought-after time-dependent multipole expansion of the electromagnetic field. For this purpose we will use the vector counterpart of Eqs. (19) and (20), given by⁴

$$\frac{(-i)^\ell}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_{\pm}} d\alpha \sin \alpha \mathbf{Y}_{\ell,m}(\alpha, \beta) e^{i(\omega/c)\mathbf{s} \cdot \mathbf{r}} = h_{\ell}^{(1)}\left(\frac{\omega}{c} r\right) \mathbf{Y}_{\ell,m}(\theta, \phi). \quad (54)$$

Furthermore, by applying the curl operator $\nabla \times$ to both sides of (54) while using $\nabla e^{i(\omega/c)\mathbf{s} \cdot \mathbf{r}} = i(\omega/c) \mathbf{s} e^{i(\omega/c)\mathbf{s} \cdot \mathbf{r}}$ and $\nabla \times \mathbf{Y}_{\ell,m}(\alpha, \beta) = 0$, one readily obtains⁴

$$\frac{(-i)^\ell}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_{\pm}} d\alpha \sin \alpha \mathbf{s} \times \mathbf{Y}_{\ell,m}(\alpha, \beta) e^{i(\omega/c)\mathbf{s} \cdot \mathbf{r}} = \frac{-ic}{\omega} \nabla \times \left[h_{\ell}^{(1)}\left(\frac{\omega}{c} r\right) \mathbf{Y}_{\ell,m}(\theta, \phi) \right]. \quad (55)$$

The following results follow readily from Eqs. (54) and (55) and Eqs. (23) and (24):

$$\begin{aligned} & \frac{(-i)^\ell}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_{\pm}} d\alpha \sin \alpha a_{\ell,m}^+(t - \mathbf{s} \cdot \mathbf{r}/c) \mathbf{Y}_{\ell,m}(\alpha, \beta) \\ &= \frac{1}{2\pi} \int_0^{\infty} d\omega e^{-i\omega t} \tilde{a}_{\ell,m}(\omega) h_{\ell}^{(1)}\left(\frac{\omega}{c} r\right) \mathbf{Y}_{\ell,m}(\theta, \phi) \\ &= -(-i)^\ell \frac{c}{r} \mathbf{Y}_{\ell,m}(\theta, \phi) \partial_t^{-1} \mathcal{L}_{\ell}^{(r,t)} a_{\ell,m}^+(t) \end{aligned} \quad (56)$$

and, similarly,

$$\begin{aligned} & \frac{(-i)^\ell}{2\pi} \int_{-\pi}^{\pi} d\beta \int_{C_{\pm}} d\alpha \sin \alpha a_{\ell,m}^+(t - \mathbf{s} \cdot \mathbf{r}/c) \mathbf{s} \times \mathbf{Y}_{\ell,m}(\alpha, \beta) = \frac{1}{2\pi} \int_0^{\infty} d\omega e^{-i\omega t} \tilde{a}_{\ell,m}(\omega) \left\{ \frac{-ic}{\omega} \right. \\ & \quad \left. \times \nabla \left[h_{\ell}^{(1)}\left(\frac{\omega}{c} r\right) \mathbf{Y}_{\ell,m}(\theta, \phi) \right] \right\} = (-i)^\ell \frac{c^2}{r} \partial_t^{-2} \nabla \times [\mathbf{Y}_{\ell,m}(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} a_{\ell,m}^+(t)]. \end{aligned} \quad (57)$$

By substituting from Eqs. (56) and (57) into Eq. (53) we obtain

$$\begin{aligned}
\mathcal{E}^+(\mathbf{r}, t) &= \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} \frac{1}{r} (-i)^{\ell} \{ -c \partial_t^{-1} \nabla \times [\mathbf{Y}_{\ell,m}(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} a_{\ell,m}^+(t)] \\
&\quad + \mathbf{Y}_{\ell,m}(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} b_{\ell,m}^+(t) \} \\
\mathcal{H}^+(\mathbf{r}, t) &= \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} \frac{1}{r} (-i)^{\ell} \{ -\mathbf{Y}_{\ell,m}(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} a_{\ell,m}^+(t) \\
&\quad - c \partial_t^{-1} \nabla [\mathbf{Y}_{\ell,m}(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} b_{\ell,m}^+(t)] \}.
\end{aligned} \tag{58}$$

Furthermore, we can now express the real electric and magnetic fields using the real vector spherical harmonics (refer to Appendix A for a review) following lines analogous to those used for scalar fields in Sec. II C. By analogy with our treatment of the scalar field we begin by introducing the new (real) multipole moments

$$\begin{aligned}
c_{\ell,m}^{(j)}(t) &= -\frac{1}{\ell(\ell+1)} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha \mathbf{F}_H(\mathbf{s}, t) \cdot \mathbf{S}_{\ell,m}^{(j)}(\alpha, \beta), \\
d_{\ell,m}^{(j)}(t) &= \frac{1}{\ell(\ell+1)} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha \mathbf{F}_E(\mathbf{s}, t) \cdot \mathbf{S}_{\ell,m}^{(j)}(\alpha, \beta).
\end{aligned} \tag{59}$$

We can also express $c_{\ell,m}^{(j)}(t)$ and $d_{\ell,m}^{(j)}(t)$ in terms of the slant-stack transform of the current distribution by substituting from Eq. (47) into Eq. (59). This procedure yields

$$\begin{aligned}
c_{\ell,m}^{(j)}(t) &= \frac{c^{-2}}{\ell(\ell+1)} \frac{\partial}{\partial t} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha [\mathbf{s} \times \hat{\mathcal{T}}(\mathbf{s}, t)] \cdot \mathbf{S}_{\ell,m}^{(j)}(\alpha, \beta), \\
d_{\ell,m}^{(j)}(t) &= \frac{c^{-2}}{\ell(\ell+1)} \frac{\partial}{\partial t} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha \{ \mathbf{s} \times [\mathbf{s} \times \hat{\mathcal{T}}(\mathbf{s}, t)] \} \cdot \mathbf{S}_{\ell,m}^{(j)}(\alpha, \beta),
\end{aligned} \tag{60}$$

in analogy to its analytic signal counterpart Eq. (52). Both forms of time-dependent multipole moments defined above [i.e., Eqs. (59) and (60)] appear to be new.

By employing the vector counterpart of the procedure used to derive Eq. (33) from Eq. (25) (see Appendix B), we obtain from Eq. (58) the sought-after time-dependent multipole representation for the real, physical fields:

$$\begin{aligned}
\mathcal{E}(\mathbf{r}, t) &= \sum_{\ell=1}^{\infty} \sum_{m=0}^{\ell} \sum_{j=1,2} c_m \frac{1}{r} \{ -c \partial_t^{-1} \nabla \times [\mathbf{S}_{\ell,m}^{(j)}(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} c_{\ell,m}^{(j)}(t)] + \mathbf{S}_{\ell,m}^{(j)}(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} d_{\ell,m}^{(j)}(t) \}, \\
\mathcal{H}(\mathbf{r}, t) &= \sum_{\ell=1}^{\infty} \sum_{m=0}^{\ell} \sum_{j=1,2} c_m \frac{1}{r} \{ -\mathbf{S}_{\ell,m}^{(j)}(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} c_{\ell,m}^{(j)}(t) - c \partial_t^{-1} \nabla \times [\mathbf{S}_{\ell,m}^{(j)}(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} d_{\ell,m}^{(j)}(t)] \},
\end{aligned} \tag{61}$$

where, as before, $c_m = 1$ or 2 if $m = 0$ or $m \geq 1$, respectively. Equation (61) is, to the authors' knowledge, also new. Note also that Eq. (59) gives the real electric and magnetic multipole moments in terms of the far electric and magnetic fields while Eq. (60) gives them in terms of the transverse part of the slant-stack transform of the current distribution. Thus, Eq. (61) yields solutions to the radiation problem as well as to the inverse diffraction problem from time domain far field data. The theory above also corroborates, in the time domain, the well-known result that the radiation pattern uniquely determines the radiated field everywhere outside the source, as is evident from Eqs. (61) and (59).

In analogy to our treatment of the scalar field in Sec. II, we conclude this section with alternative expressions for the real time-dependent multipole moments $c_{\ell,m}^{(j)}(t)$ and $d_{\ell,m}^{(j)}(t)$. The corresponding derivation follows by analogy with the procedure used in Sec. II.

We conveniently rewrite $a_{\ell,m}^+(t)$ as defined in Eq. (52) as

$$a_{\ell,m}^+(t) = -\frac{i^\ell c^{-1}}{\ell(\ell+1)} \int_{r \leq a} d\mathbf{r} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha [\nabla \times \mathcal{T}^+(\mathbf{r}, t + \mathbf{s} \cdot \mathbf{r}/c)] \cdot \mathbf{Y}_{\ell,m}^*(\alpha, \beta). \quad (62)$$

The result above was obtained by using (see Appendix C)

$$\mathbf{s} \times \hat{\mathcal{T}}^+(\mathbf{s}, t) = -c \partial_t^{-1} \int d\mathbf{r} \nabla \times \mathcal{T}^+(\mathbf{r}, t + \mathbf{s} \cdot \mathbf{r}/c). \quad (63)$$

A convenient expression for $b_{\ell,m}^+(t)$ is obtained, on the other hand, by substituting from Eq. (46) into Eq. (52) and interchanging orders of \mathbf{r} and \mathbf{s} integration, thus giving

$$b_{\ell,m}^+(t) = -\frac{i^\ell c^{-2}}{\ell(\ell+1)} \frac{\partial}{\partial t} \int_{r \leq a} d\mathbf{r} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha \mathcal{T}^+(\mathbf{r}, t + \mathbf{s} \cdot \mathbf{r}/c) \cdot \mathbf{Y}_{\ell,m}(\alpha, \beta), \quad (64)$$

where we have made use of $\mathbf{s} \times \mathbf{s} \times \mathbf{S}_{\ell,m}^{(j)}(\alpha, \beta) = -\mathbf{S}_{\ell,m}^{(j)}(\alpha, \beta)$ and the well-known vector identity $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})$.

Next we make use of the vector counterpart of Eq. (22):⁴

$$\frac{(-i)^\ell}{4\pi} \int_{-\pi}^{\pi} d\beta \int_0^{\pi} d\alpha \sin \alpha \mathbf{Y}_{\ell,m}(\alpha, \beta) e^{i(\omega/c)\mathbf{s} \cdot \mathbf{r}} = j_\ell \left(\frac{\omega}{c} r \right) \mathbf{Y}_{\ell,m}(\theta, \phi). \quad (65)$$

By substituting from (65) into Eqs. (63) and (64) we obtain after some manipulations

$$\begin{aligned} a_{\ell,m}^+(t) &= -\frac{i^\ell c^{-1}}{\ell(\ell+1)} \frac{4\pi c^{-\ell}}{\ell! 2^{\ell+1}} \int_{-\pi}^{\pi} d\phi \int_0^{\pi} d\theta \sin \theta \mathbf{Y}_{\ell,m}^*(\theta, \phi) \cdot \left\{ \int_{r \leq a} dr r^\ell [\nabla \times \mathbf{I}^+(\mathbf{r}, t)] \right\}, \\ b_{\ell,m}^+(t) &= -\frac{i^\ell c^{-2}}{\ell(\ell+1)} \frac{4\pi c^{-\ell}}{\ell! 2^{\ell+1}} \frac{\partial}{\partial t} \int_{-\pi}^{\pi} d\phi \int_0^{\pi} d\theta \sin \theta \mathbf{Y}_{\ell,m}^*(\theta, \phi) \cdot \left[\int_{r \leq a} dr r^\ell \mathbf{I}^+(\mathbf{r}, t) \right], \end{aligned} \quad (66)$$

with

$$\mathbf{I}^+(\mathbf{r}, t) = \int_{-1}^1 d\eta (1 - \eta^2)^\ell \frac{\partial^\ell}{\partial t^\ell} \mathcal{T}^+(\mathbf{r}, t + r\eta/c). \quad (67)$$

Equation (66) has been derived before from the Debye representation;^{6,7} here it has been obtained from the time-dependent plane wave representation.

Finally, the real time-dependent electric and magnetic multipole moments $c_{\ell,m}^{(j)}(t)$ and $d_{\ell,m}^{(j)}(t)$ are readily obtained from Eqs. (66) and (67) by recalling that $c_{\ell,m}^+(t) = (-i)^\ell a_{\ell,m}^+(t)$ and $d_{\ell,m}^+(t) = (-i)^\ell b_{\ell,m}^+(t)$. We obtain

$$\begin{aligned} c_{\ell,m}^{(j)}(t) &= -\frac{c^{-1}}{\ell(\ell+1)} \frac{4\pi c^{-\ell}}{\ell! 2^{\ell+1}} \int_{-\pi}^{\pi} d\phi \int_0^{\pi} d\theta \sin \theta \mathbf{S}_{\ell,m}^{(j)}(\theta, \phi) \cdot \left\{ \int_{r \leq a} dr r^\ell [\nabla \times \mathbf{I}(\mathbf{r}, t)] \right\}, \\ d_{\ell,m}^{(j)}(t) &= -\frac{c^{-2}}{\ell(\ell+1)} \frac{4\pi c^{-\ell}}{\ell! 2^{\ell+1}} \frac{\partial}{\partial t} \int_{-\pi}^{\pi} d\phi \int_0^{\pi} d\theta \sin \theta \mathbf{S}_{\ell,m}^{(j)}(\theta, \phi) \cdot \left[\int_{r \leq a} dr r^\ell \mathbf{I}(\mathbf{r}, t) \right], \end{aligned} \quad (68)$$

with

$$\mathbf{I}(\mathbf{r}, t) = \int_{-1}^1 d\eta (1 - \eta^2)^\ell \frac{\partial^\ell}{\partial t^\ell} \mathcal{T}(\mathbf{r}, t + r\eta/c). \quad (69)$$

Equations (68) and (69) are the electromagnetic case counterparts of Eqs. (41) and (42) and their interpretation is analogous to that of Eqs. (41) and (42).

ACKNOWLEDGMENTS

The authors acknowledge support from the U.S. Air Force Office of Scientific Research, under Grant No. F49620-93-1-0093 and the Center for Electromagnetics Research at Northeastern University.

APPENDIX A: ORDINARY AND VECTOR SPHERICAL HARMONICS

We use the standard definition of the spherical harmonics:¹⁴

$$Y_{\ell,m}(\theta, \phi) = y_{\ell,m} P_{\ell}^m(\cos \theta) e^{im\phi} \quad (\text{A1})$$

with

$$y_{\ell,m} = \sqrt{\frac{(2\ell+1)(\ell-m)!}{4\pi(\ell+m)!}}, \quad (\text{A2})$$

where $\ell = 0, 1, 2, \dots$, $m = 0, \pm 1, \dots, \pm \ell$, and $P_{\ell}^m(x)$ are the associated Legendre polynomials, defined by means of the Rodrigues formula

$$P_{\ell}^m(x) = \frac{(-1)^m}{2^{\ell} \ell!} (1-x^2)^{m/2} \left(\frac{d}{dx} \right)^{\ell+m} (x^2-1)^{\ell}. \quad (\text{A3})$$

The spherical harmonics obey the orthogonality and closure properties¹⁴

$$\int_{-\pi}^{\pi} d\phi \int_0^{\pi} d\theta \sin \theta Y_{\ell,m}^*(\theta, \phi) Y_{\ell',m'}(\theta, \phi) = \delta_{\ell,\ell'} \delta_{m,m'}, \quad (\text{A4})$$

$$\sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y_{\ell,m}(\theta, \phi) Y_{\ell,m}^*(\theta', \phi') = \frac{1}{\sin \theta} \delta(\theta - \theta') \delta(\phi - \phi') \equiv \delta(\hat{\mathbf{r}} - \hat{\mathbf{r}}'). \quad (\text{A5})$$

The real spherical harmonics $S_{\ell,m}^{(1)}(\theta, \phi) \equiv \Re\{Y_{\ell,m}(\theta, \phi)\}$ and $S_{\ell,m}^{(2)}(\theta, \phi) \equiv \Im\{Y_{\ell,m}(\theta, \phi)\}$, where $\ell = 0, 1, 2, \dots$, and $m = 0, 1, \dots, \ell$, obey, on the other hand, the orthogonality and closure properties of Eqs. (31) and (32).³

The vector spherical harmonic $\mathbf{Y}_{\ell,m}(\theta, \phi)$ is defined by^{14,21}

$$\mathbf{Y}_{\ell,m}(\theta, \phi) = \mathbf{L} Y_{\ell,m}(\theta, \phi), \quad (\text{A6})$$

where

$$\mathbf{L} = -i\mathbf{r} \times \nabla = -i \left(\mathbf{u}_{\phi} \frac{\partial}{\partial \theta} - \frac{1}{\sin \theta} \mathbf{u}_{\theta} \frac{\partial}{\partial \phi} \right). \quad (\text{A7})$$

In Eq. (A7), \mathbf{u}_{θ} and \mathbf{u}_{ϕ} are, respectively, the unit vectors in the positive θ and ϕ directions. The $\mathbf{Y}_{\ell,m}(\theta, \phi)$'s are defined for $\ell = 1, 2, \dots$, and $m = 0, \pm 1, \dots, \pm \ell$. The value $\ell = 0$ is excluded since $Y_{0,0} = 1/\sqrt{4\pi}$, hence from Eqs. (A6) and (A7) $\mathbf{Y}_{0,0} = 0$.

The vector spherical harmonics $\mathbf{Y}_{\ell,m}(\theta, \phi)$ are everywhere tangential to the unit sphere [such that $\hat{\mathbf{r}} \cdot \mathbf{Y}_{\ell,m}(\theta, \phi) = 0$]. They satisfy the orthogonality condition^{14,21}

$$\int_{-\pi}^{\pi} d\phi \int_0^{\pi} d\theta \sin \theta \mathbf{Y}_{\ell,m}^*(\theta, \phi) \cdot \mathbf{Y}_{\ell',m'}(\theta, \phi) = \ell(\ell+1) \delta_{\ell,\ell'} \delta_{m,m'}. \quad (\text{A8})$$

They also form, together with the associated vector functions $\hat{\mathbf{r}} \times \mathbf{Y}_{\ell,m}(\theta, \phi)$, a complete orthogonal set in terms of which all well-behaved vector functions $\mathbf{F}(\hat{\mathbf{r}})$ with $\hat{\mathbf{r}} \cdot \mathbf{F}(\hat{\mathbf{r}}) = 0$ can be expanded.^{14,21}

We define the real vector spherical harmonics $\mathbf{S}_{\ell,m}^{(j)}$ via

$$\mathbf{S}_{\ell,m}^{(1)}(\theta, \phi) = \mathcal{I}\{\mathbf{Y}_{\ell,m}(\theta, \phi)\} = -i\mathbf{L}S_{\ell,m}^{(1)}, \quad (\text{A9})$$

$$\mathbf{S}_{\ell,m}^{(2)}(\theta, \phi) = \mathcal{R}\{\mathbf{Y}_{\ell,m}(\theta, \phi)\} = i\mathbf{L}S_{\ell,m}^{(2)},$$

so that, from (31) and $L^2 Y_{\ell,m} = \ell(\ell+1)Y_{\ell,m}$ (see Refs. 14 and 21),

$$\int_{-\pi}^{\pi} d\phi \int_0^{\pi} d\theta \sin \theta \mathbf{S}_{\ell,m}^{(j)}(\theta, \phi) \cdot \mathbf{S}_{\ell',m'}^{(j')}(\theta, \phi) = \ell(\ell+1) \delta_{\ell,\ell'} \delta_{m,m'} \delta_{j,j'} \frac{1}{c_m}. \quad (\text{A10})$$

They also form, together with the vector functions $\hat{\mathbf{r}} \times \mathbf{S}_{\ell,m}^{(j)}(\theta, \phi)$ a complete orthogonal set in terms of which all well-behaved vector functions $\mathbf{F}(\hat{\mathbf{r}})$ with $\hat{\mathbf{r}} \cdot \mathbf{F}(\hat{\mathbf{r}}) = 0$ can be expanded.

APPENDIX B: DERIVATION OF EQ. (33)

A key step in deriving (33) will consist in showing that

$$2\Re \left\{ \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} (-i)^{\ell} Y_{\ell,m}^m(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} a_{\ell,m}^+(t) \right\} = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} \sum_{j=1,2} c_m S_{\ell,m}^{(j)}(\theta, \phi) \mathcal{L}_{\ell}^{(r,t)} q_{\ell,m}^{(j)}(t) \quad (\text{B1})$$

with $a_{\ell,m}^+(t)$ and $q_{\ell,m}^{(j)}(t)$ given by (26) and (34), respectively, and where $c_m = 1$ or 2 for $m = 0$ or $m \geq 1$, respectively. To simplify the manipulations we will use Dirac bra-ket notation.²¹ Thus we represent the spherical and real spherical harmonics $Y_{\ell,m}$ and $S_{\ell,m}^{(j)}$, respectively, by the vectors (kets) $|\ell, m\rangle$ and $|\ell, m, j\rangle$, where²¹

$$Y_{\ell,m}(\theta, \phi) \equiv \langle \mathbf{r} | \ell, m \rangle, \quad (\text{B2})$$

$$S_{\ell,m}^{(j)}(\theta, \phi) \equiv \langle \mathbf{r} | \ell, m, j \rangle. \quad (\text{B3})$$

In Dirac bra-ket notation, the orthogonality and closure properties defined in Eqs. (A4) and (A5) and Eqs. (31) and (32) translate into

$$\langle \ell, m | \ell', m' \rangle = \delta_{\ell,\ell'} \delta_{m,m'}, \quad (\text{B4})$$

$$\sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} |\ell, m\rangle \langle \ell, m| = I, \quad (\text{B5})$$

$$\langle \ell, m, j | \ell', m', j' \rangle = \frac{1}{c_m} \delta_{\ell,\ell'} \delta_{m,m'} \delta_{j,j'}, \quad (\text{B6})$$

$$\sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} \sum_{j=1,2} c_m |\ell, m, j\rangle \langle \ell, m, j| = I, \quad (\text{B7})$$

where I is the identity operator. We introduce, conveniently, the new analytic multipole moments

$$q_{\ell,m}^+(t) \equiv (-i)^{\ell} a_{\ell,m}^+(t), \quad (\text{B8})$$

which are the ordinary spherical harmonic counterpart of $q_{\ell,m}^{(j)}(t)$. By substituting from (B8) into (25) and (26) while using Dirac bra-ket notation we obtain

$$U(\mathbf{r}, t) = \frac{1}{r} 2\Re \left\{ \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \langle \mathbf{r} | \ell, m \rangle \mathcal{L}_{\ell}^{(r,t)} \langle \ell, m | F^+(t) \rangle \right\}. \quad (\text{B9})$$

By inserting the identity operator right after the term $\langle \ell, m |$ in Eq. (B9) while making use of Eqs. (B5) and (B7) and interchanging orders of summation one readily finds

$$U(\mathbf{r}, t) = \frac{1}{r} \sum_{\ell'=0}^{\infty} \sum_{m'=0}^{\ell'} \sum_{j'=1,2} c_{m'} \langle \mathbf{r} | \ell', m', j' \rangle \mathcal{L}_{\ell'}^{(r,t)} \langle \ell', m', j' | F(t) \rangle. \quad (\text{B10})$$

Finally, we complete our derivation of Eq. (33) by substituting expression (B3) and $q_{\ell', m'}^{(j')}(t) = \langle \ell', m', j' | F(t) \rangle$ [this follows from Eq. (34)] in Eq. (B10).

APPENDIX C: DERIVATION OF EQS. (47)

Consider a four-dimensional (4D) space-time with Cartesian coordinates ($x_0 = ct$, $x_1 = x$, $x_2 = y$, $x_3 = z$) and denote the coordinate vector in space-time as \mathbf{x} . The Radon transform $R\mathbf{f}(\mathbf{v}, \xi)$ of a rather arbitrary vector function $\mathbf{f}(\mathbf{x})$ is defined as

$$R\mathbf{f}(\mathbf{v}, \xi) = \int d\mathbf{x} \mathbf{f}(\mathbf{x}) \delta(\xi - \mathbf{v} \cdot \mathbf{x}), \quad (\text{C1})$$

where \mathbf{v} is a unit vector in 4D Radon domain and ξ is a scalar. We will use the property [see Ref. 22, Eq. (6.7)]

$$R(\mathcal{H}\mathbf{f}) = \mathcal{H} \left(v_1 \frac{\partial}{\partial \xi}, \dots, v_n \frac{\partial}{\partial \xi} \right) R\mathbf{f}, \quad (\text{C2})$$

where x_k is the k th Cartesian component of \mathbf{x} , v_k is the k th Cartesian component of the unit vector \mathbf{v} and $\mathcal{H}(\partial/\partial x_1, \dots, \partial/\partial x_n)$ is a linear operator with constant coefficients.

By putting $\mathbf{f}^+(\mathbf{x}) \equiv \mathcal{F}^+(\mathbf{r}, t = x_0/c)$, Eq. (46) and the various definitions above yield

$$\hat{\mathcal{F}}^+(\mathbf{s}, t) = \frac{1}{\sqrt{2}} R\mathbf{f}^+(\mathbf{v}_s, ct/\sqrt{2}), \quad (\text{C3})$$

where $\mathbf{v}_s = (1/\sqrt{2})(1, -\mathbf{s})$ is the unit vector whose Cartesian components are ($x_0 = (1/\sqrt{2})$, $x_1 = -(1/\sqrt{2})$, $x_2 = -(1/\sqrt{2})$, $x_3 = -(1/\sqrt{2})$).

By substituting into Eq. (45) from Eqs. (C1)–(C3) we obtain

$$\mathbf{F}_E^+(\mathbf{s}, t) = \frac{1}{\sqrt{2}} R \left\{ \left[-\frac{1}{c^2} \frac{\partial}{\partial t} + \partial_t^{-1} \nabla \nabla \cdot \right] \mathbf{f}^+ \right\} (\mathbf{v}_s, ct/\sqrt{2}) \quad (\text{C4})$$

and

$$\mathbf{F}_H^+(\mathbf{s}, t) = \frac{1}{c\sqrt{2}} R \{ \nabla \times \mathbf{f}^+ \} (\mathbf{v}_s, ct/\sqrt{2}). \quad (\text{C5})$$

Using Eqs. (C2) and (C3) one can easily show that

$$\begin{aligned} R \{ \nabla \nabla \cdot \mathbf{f} \} (\mathbf{v}_s, ct/\sqrt{2}) &= \frac{1}{c^2} \frac{\partial^2}{\partial t^2} [\mathbf{s} \mathbf{s} \cdot R\mathbf{f}(\mathbf{v}_s, ct/\sqrt{2})], \\ R \{ \nabla \times \mathbf{f} \} (\mathbf{v}_s, ct/\sqrt{2}) &= -\frac{1}{c} \frac{\partial}{\partial t} [\mathbf{s} \times R\mathbf{f}(\mathbf{v}_s, ct/\sqrt{2})]. \end{aligned} \quad (\text{C6})$$

The final step consists of applying the results (C6)–(C3) to Eqs. (C4)–(C5) and substituting into the result thus obtained from Eq. (C3).

¹E. Heyman and L. B. Felsen, "Complex-Source Pulsed Beam Fields," J. Opt. Soc. Am. A **6**, 806–817 (1989).

²R. W. Ziolkowski, I. M. Besieris, and A. M. Shaarawi, "Localized Wave Representations of Acoustic and Electromagnetic Radiation," Proc. IEEE **79**, 1371–1377 (1991).

³E. Heyman and A. J. Devaney, "Time-Dependent Multipoles and Their Application for Radiation from Volume Source Distributions," J. Math. Phys. **37**, 682–692 (1996).

- ⁴A. J. Devaney and E. Wolf, "Multipole Expansions and Plane-Wave Representations of the Electromagnetic Field," J. Math. Phys. **15**, 234–244 (1974).
- ⁵K. D. Granzow, "Multipole Theory in the Time Domain," J. Math. Phys. **7**, 634–640 (1966).
- ⁶W. C. Davidon, "Time-Dependent Multipole Analysis," J. Phys. A **6**, 1635–1646 (1973).
- ⁷W. B. Campbell, J. Macek, and T. A. Morgan, "Relativistic Time-Dependent Multipole Analysis for Scalar, Electromagnetic and Gravitational Fields," Phys. Rev. D **15**, 2156–2164 (1977).
- ⁸T. B. Hansen, "Formulation of Spherical Near-Field Scanning for Electromagnetic Fields in the Time Domain," IEEE Trans. Antennas Propag. **45**, 620–630 (1997).
- ⁹T. B. Hansen and A. D. Yaghjian, "Planar Near-Field Scanning in the Time Domain. I. Formulation," IEEE Trans. Antennas Propag. **42**, 1268–1279 (1994).
- ¹⁰R. F. Harrington, "On the Gain and Beamwidth of Directional Antennas," IRE Trans. Antennas Propag. **6**, 219–225 (1958).
- ¹¹D. R. Rhodes, "On a Fundamental Principle in the Theory of Planar Antennas," Proc. IEEE **52**, 1013–1021 (1964).
- ¹²J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941).
- ¹³L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, New York, 1995).
- ¹⁴J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962).
- ¹⁵A. J. Devaney and G. C. Sherman, "Plane-Wave Representations for Scalar Wave Fields," SIAM (Soc. Ind. Appl. Math.) Rev. **15**, 765–786 (1973).
- ¹⁶E. Heyman, "Time-Dependent Plane-Wave Spectrum Representations for Radiation from Volume Source Distributions," J. Math. Phys. **37**, 658–681 (1996).
- ¹⁷A. D. Yaghjian and T. B. Hansen, "Time-Domain Far Fields," J. Appl. Phys. **79**, 2822–2830 (1996).
- ¹⁸M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, 9th ed. (Dover, New York, 1970).
- ¹⁹T. B. Hansen and A. N. Norris, "Exact Complex Source Representations of Transient Radiation," Wave Motion **26**, 101–115 (1997).
- ²⁰A. Shlivinski, E. Heyman, and R. Kastner, "Antenna Characterization in the Time Domain," IEEE Trans. Antennas Propag. **45**, 1140–1149 (1997).
- ²¹C. Cohen-Tannoudji, B. Diu, and F. Laloe, *Quantum Mechanics* (Wiley, New York, 1997).
- ²²S. R. Deans, *The Radon Transform and Some of its Applications* (Wiley, New York, 1983).

- spaces," *IEEE Trans. Antennas Propagat.*, vol. 41, pp. 806–812, June 1993.
- [7] J. He, N. Geng, T. Yu, and L. Carin, "Electromagnetic scattering from a general dielectric target embedded in a multilayered medium," *Radio Sci.* to be published.
- [8] L. Carin, R. Kapoor, and C. E. Baum, "A Polarimetric SAR imaging of buried landmines," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 1985–1988, Nov. 1998.
- [9] J. R. Mautz and R. F. Harrington, "Electromagnetic scattering from a homogeneous material body of revolution," *AEU*, vol. 33, pp. 71–80, 1979.
- [10] J. M. Bourgeois and G. S. Smith, "A fully three-dimensional simulation of a ground-penetrating radar: FDTD theory compared with measurements," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 36–44, Jan. 1996.
- [11] T. Yu and L. Carin, "Extended-Born analysis of the electromagnetic-induction response of a dielectric target embedded in a lossy layered medium," *Radio Sci.*, to be published.

Buried Object Detection and Location Estimation from Electromagnetic Field Measurements

George A. Tsihrintzis, Peter Meincke Johansen,
and Anthony J. Devaney

Abstract—A translation property is derived describing the field scattered from a known buried object placed at distinct locations. The result is used to derive the optimum algorithm for detecting the known buried object and estimating its location from noisy scattered electromagnetic field measurements.

Index Terms—Ground penetrating radar, Lippmann–Schwinger equation, object detection, wave scattering.

I. INTRODUCTION

The problem of detecting known buried objects and estimating their location from electromagnetic field measurements is relevant in many technological areas such as demining, buried waste clean up, excavation planning, and archaeological investigations. In all of the above applications, serious challenges arise, mainly due to physical limitations such as: 1) significant losses due to moist soil that limit the signal-to-noise ratio; 2) presence of a large number of randomly distributed unwanted objects, returns from which obscure the return from the object of interest (volume clutter); and 3) random roughness of the air/soil interface that results in incoherent (random) returns (surface clutter). As a result of these limitations, standard threshold detection algorithms [1] may not successfully address the buried object detection problem, especially when the object is small and the return signal weak.

Manuscript received August 27, 1998; revised August 16, 1999. The work of G. A. Tsihrintzis and A. J. Devaney were supported in part by ARO Grant DAAG55-97-1-0013 and by AFOSR Grant F49620-96-1-0028. The work of P. M. Johansen was supported by the Danish Technical Research Council.

G. A. Tsihrintzis is with the Department of Informatics, University of Piraeus, Piraeus, 185 34 Greece.

P. M. Johansen is with the Department of Electromagnetic Systems, Technical University of Denmark, Lyngby, DK-2800 Denmark.

A. J. Devaney is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA.

Publisher Item Identifier S 0018-926X(99)09952-4.

The signal processing strategy with the highest potential to address the buried object detection problem is the one that utilizes field models for the air/soil/object environment. Due to the complicated nature of this environment, however, certain simplifications need to be made, namely the assumption of a planar air/soil interface and of no interactions between the object of interest and the distribution of unwanted objects. These assumptions allow, as illustrated in this letter, the solution to the buried object detection problem within the framework of exact electromagnetic field scattering theory and the derivation of a computationally efficient optimum detection algorithm. More specifically, the letter is organized as follows. Section II is devoted to statement of the basic field equations governing the interaction of probing fields with the air/soil/object environment and derivation of a field translation property. Section III addresses the buried object detection problem and presents the maximum likelihood algorithm for the solution to it. Finally, Section IV is a discussion of the results of the letter with suggestions for possible future research avenues.

II. CONFIGURATION AND SCATTERING EQUATIONS

Consider the configuration in Fig. 1 in which a planar interface separates air (medium 1) from soil (medium 2). The orthogonal coordinate system $\mathbf{r} = \mathbf{x} + \hat{z}z$ is defined, with \mathbf{x} indicating a two-dimensional (2-D) coordinate on the interface and the z axis directed so that $z > 0$ is air and $z < 0$ is soil. Buried in the soil is a known target object, but its coordinates $\mathbf{r}_c = \mathbf{x}_c + \hat{z}z_c$ with $z_c \leq 0$ are unknown. A monochromatic plane wave with electric field $\mathbf{E}^i(\mathbf{r}) = \mathbf{E}_0 e^{i\mathbf{k}_1 \cdot \mathbf{r}}$ and wavevector \mathbf{k}_1 is incident upon the planar interface. As the incident field reaches the interface, it partially reflects back into air and partially refracts into the soil where it interacts with the buried object. The interaction produces a scattered field, part of which refracts back into air where it is measured.

The total field $\mathbf{E}(\mathbf{r}; \mathbf{r}_c)$ at position \mathbf{r} is given by

$$\mathbf{E}(\mathbf{r}; \mathbf{r}_c) = \mathbf{E}^b(\mathbf{r}) + \mathbf{E}^s(\mathbf{r}; \mathbf{r}_c) \quad (1)$$

where $\mathbf{E}^b(\mathbf{r})$ is the electric field in the absence of the buried object (background field) consisting of the incident and the reflected field in air and the refracted field in soil and $\mathbf{E}^s(\mathbf{r}; \mathbf{r}_c)$ is the scattered field due to the presence of the buried object at the unknown location \mathbf{r}_c .

$$\mathbf{E}^s(\mathbf{r}; \mathbf{r}_c) = \begin{cases} \mathbf{E}_2^s(\mathbf{r}; \mathbf{r}_c) \equiv \int_{z' < 0} \bar{\mathbf{G}}_{22}(\mathbf{r}, \mathbf{r}') \cdot [\mathbf{E}^b(\mathbf{r}') + \mathbf{E}_2^s(\mathbf{r}'; \mathbf{r}_c)] O(\mathbf{r}'; \mathbf{r}_c) d^3 r', & z < 0 \text{ (soil)} \\ \mathbf{E}_1^s(\mathbf{r}; \mathbf{r}_c) \equiv \int_{z' < 0} \bar{\mathbf{G}}_{12}(\mathbf{r}, \mathbf{r}') \cdot [\mathbf{E}^b(\mathbf{r}') + \mathbf{E}_2^s(\mathbf{r}'; \mathbf{r}_c)] O(\mathbf{r}'; \mathbf{r}_c) d^3 r', & z > 0 \text{ (air)}. \end{cases} \quad (2)$$

In (2), the object function $O(\mathbf{r}; \mathbf{r}_c)$ is defined as

$$O(\mathbf{r}; \mathbf{r}_c) = k^2(\mathbf{r}; \mathbf{r}_c) - k_2^2 \quad (3)$$

where $k(\mathbf{r}; \mathbf{r}_c)$ is the complex wavenumber at point \mathbf{r} and k_2 is the complex wavenumber of the soil. Equation (2) is the Lippmann–Schwinger equation for the scattering problem and maps the object function to the corresponding scattered field nonlinearly.

The dyadic Green functions $\bar{\mathbf{G}}_{12}$ and $\bar{\mathbf{G}}_{22}$ are defined for the source point in soil and the observation point in air and soil, respectively. Explicit expressions for them can be found in [2]; however, the critical information for this context is that both Green

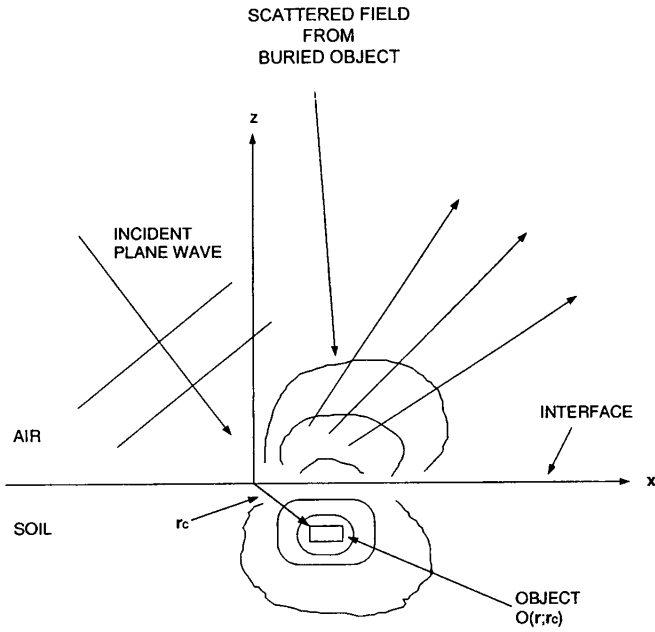


Fig. 1. Configuration of the buried object detection problem in which a plane wave is impinging upon the air/soil interface. From scattered field measurements above the interface, an estimate of the location of the known object is searched for.

functions satisfy the property $\bar{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \bar{\mathbf{G}}(\mathbf{x} - \mathbf{x}', z, z')$. Express now the object function $O(\mathbf{r}; \mathbf{r}_c)$ as

$$O(\mathbf{r}; \mathbf{r}_c) = O_0(\mathbf{r} - \mathbf{r}_c) \quad (4)$$

describing the known buried object $O_0(\mathbf{r})$ located at $\mathbf{r}_c = \mathbf{x}_c + \hat{z}z_c$. The scattered field in soil satisfies the upper branch in (2) or, equivalently, as algebraic manipulation shows

$$\begin{aligned} & e^{-i\mathbf{K} \cdot \mathbf{x}_c} \mathbf{E}_2^s(\mathbf{r} + \mathbf{x}_c; \mathbf{r}_c) \\ &= \int_{z' < 0} \bar{\mathbf{G}}_{22}(\mathbf{x} - \mathbf{x}', z, z') \cdot [\mathbf{E}^b(\mathbf{r}') \\ &+ e^{-i\mathbf{K} \cdot \mathbf{x}_c} \mathbf{E}_2^s(\mathbf{r}' + \mathbf{x}_c; \mathbf{r}_c)] O_0(\mathbf{r}' - \hat{z}z_c) d^3 r' \end{aligned} \quad (5)$$

where $\mathbf{K} = \mathbf{k}_1 - \hat{z}\hat{z} \cdot \mathbf{k}_1$ is the horizontal component of \mathbf{k}_1 .

The key step is now to observe that the scattered field $\mathbf{E}_2^s(\mathbf{r}; \hat{z}z_c)$ due to the object centered at $\hat{z}z_c$, satisfies the same equation as does the quantity $e^{-i\mathbf{K} \cdot \mathbf{x}_c} \mathbf{E}_2^s(\mathbf{r} + \mathbf{x}_c; \mathbf{r}_c)$ in (5). Hence, from the uniqueness of the solution to the Lippmann-Schwinger equation (2), it is readily concluded that

$$\mathbf{E}_2^s(\mathbf{r}; \mathbf{r}_c) = e^{i\mathbf{K} \cdot \mathbf{x}_c} \mathbf{E}_2^s(\mathbf{r} - \mathbf{x}_c; \hat{z}z_c); \quad z < 0. \quad (6)$$

Similarly, from (6) and the lower branch in (2), it is seen that

$$\mathbf{E}_1^s(\mathbf{r}; \mathbf{r}_c) = e^{i\mathbf{K} \cdot \mathbf{x}_c} \mathbf{E}_1^s(\mathbf{r} - \mathbf{x}_c; \hat{z}z_c); \quad z > 0. \quad (7)$$

Equations (6) and (7) state that the scattered field at position \mathbf{r} due to an object located at \mathbf{r}_c is related through a phase factor to the scattered field at position $\mathbf{r} - \mathbf{x}_c$ due to the same object located at $\hat{z}z_c$. Therefore, these equations are the two-layer background medium equivalent to the homogeneous background medium field translation property first derived by Tsirintzis and Devaney in [3]. The main difference is that in the homogeneous background case, the field translation property also applies to vertical translations of the object. In the two-layer background case, however, vertical object translations do not result in simple, closed-form field translations such as (6) and (7), due to the nonlinearity associated with the multiple interactions between the object and the air/soil interface.

III. OBJECT LOCATION ESTIMATION

Consider now noisy signal measurements V of the scattered electric field \mathbf{E}_1^s on a plane $\mathbf{r}_p = \mathbf{x}_p + \hat{z}l$ parallel to the interface at a height l above the ground

$$\begin{aligned} V(\mathbf{x}_p) &= \alpha(\mathbf{x}_p; \mathbf{r}_c) + n(\mathbf{x}_p) \\ &\equiv \mathbf{h}(\mathbf{x}_p) \otimes_{\mathbf{x}_p} \mathbf{E}_1^s(\mathbf{x}_p + \hat{z}l; \mathbf{r}_c) + n(\mathbf{x}_p). \end{aligned} \quad (8)$$

Herein, $\mathbf{h}(\mathbf{x}_p)$ is a measurement filter comprising dot multiplication and convolution and $n(\mathbf{x}_p)$ is additive white Gaussian measurement noise. The problem is to estimate the unknown location \mathbf{r}_c of the buried object from the noisy measurements $V(\mathbf{x}_p)$. The desired estimate $\hat{\mathbf{r}}_c$ is computed by minimizing the squared error between the measured data and synthetic data computed for the object at a test location \mathbf{r}_t , i.e.,

$$\hat{\mathbf{r}}_c = \arg \min_{\mathbf{r}_t} \int |V(\mathbf{x}_p) - \alpha(\mathbf{x}_p; \mathbf{r}_t)|^2 d^2 x_p. \quad (9)$$

Under the additive white Gaussian noise assumption, the estimate thus returned is the maximum likelihood estimate [1]. The parameter \mathbf{r}_c of interest enters the signal model in a nonlinear manner; therefore, no unbiased estimate exists that achieves the lower (Cramér-Rao) bound¹ on error variance [1]. However, the likelihood function to be presented below in (11) and (13) contains the same information as the full set of measurements V for the estimation of the parameter \mathbf{r}_c and, therefore, constitutes a sufficient statistic for this problem. Additionally, the maximum likelihood estimate is asymptotically unbiased and optimum in the sense that it becomes unbiased and of minimum error variance as the data measurement procedure is repeated a large number of times [1].

To derive a computationally more efficient procedure for calculating $\hat{\mathbf{r}}_c$, (9) is rewritten in two steps. First, after expanding the square in (9), it is found that the estimate $\hat{\mathbf{r}}_c$ is equivalently obtained via maximization as

$$\hat{\mathbf{r}}_c = \arg \max_{\mathbf{r}_t} L(\mathbf{r}_t) \quad (10)$$

where $L(\mathbf{r}_t)$ is the log likelihood function [1] defined as

$$\begin{aligned} L(\mathbf{r}_t) &= \Re \left\{ \int V(\mathbf{x}_p) \alpha^*(\mathbf{x}_p; \mathbf{r}_t) d^2 x_p \right\} \\ &- \frac{1}{2} \int |\alpha(\mathbf{x}_p; \mathbf{r}_t)|^2 d^2 x_p. \end{aligned} \quad (11)$$

In (11), $*$ denotes the complex conjugate and $\Re\{\cdot\}$ the real part. The second term in (11) is independent of the data and, thus, can be precomputed as a function (image) of the variable \mathbf{r}_t and stored. Moreover, as can be seen from (7), it is independent of \mathbf{x}_t , the component of \mathbf{r}_t on the interface and, thus, needs only be computed for the range of values of interest of z_t . Second, using the field translation property (7) derived in Section II, Parseval's theorem, and the shift property of the Fourier transform, the first term in (11) becomes

$$\begin{aligned} & \int V(\mathbf{x}_p) \alpha^*(\mathbf{x}_p; \mathbf{r}_t) d^2 x_p \\ &= e^{-i\mathbf{K} \cdot \mathbf{x}_t} \int V(\mathbf{x}_p) \alpha^*(\mathbf{x}_p - \mathbf{x}_t; \hat{z}z_t) d^2 x_p \\ &= \frac{e^{-i\mathbf{K} \cdot \mathbf{x}_t}}{(2\pi)^2} \int e^{i\mathbf{K}_p \cdot \mathbf{x}_t} \tilde{V}(\mathbf{K}_p) \tilde{\alpha}^*(\mathbf{K}_p; \hat{z}z_t) d^2 K_p \end{aligned} \quad (12)$$

¹The Cramér-Rao bound for the estimation problem can be computed following a procedure similar to the one outlined in [4].

where $\hat{V}(\mathbf{K}_p) = \int e^{-i\mathbf{K}_p \cdot \mathbf{x}_p} V(\mathbf{x}_p) d^2x_p$ is the 2-D spatial Fourier transform of the signal measurements $V(\mathbf{x}_p)$. Substituting (12) into (11), the log likelihood function becomes

$$L(\mathbf{r}_t) = \frac{1}{(2\pi)^2} \cdot \Re \left\{ e^{-i\mathbf{K} \cdot \mathbf{x}_t} \int e^{i\mathbf{K}_p \cdot \mathbf{x}_t} \hat{V}(\mathbf{K}_p) \hat{\alpha}^*(\mathbf{K}_p; \hat{z}_t) d^2K_p \right\} - \frac{1}{2} \int |\alpha(\mathbf{x}_p; \mathbf{r}_t)|^2 d^2x_p. \quad (13)$$

Equations (10) and (13) define an algorithm that employs a bank of parallel procedures with each branch corresponding to a different test depth. In each branch of the bank, the data are convolutionally filtered (using for example a fast Fourier transform algorithm) and multiplied by the exponential factor $e^{-i\mathbf{K} \cdot \mathbf{x}_t}$. The filter in each branch is the one matched to signal measurements of the scattered field due to the object at zero horizontal location and at the corresponding depth. The algorithm returns a bank of images, the highest maximum of which gives the horizontal location estimate and the branch at which this occurs specifies the depth estimate. The algorithm of (10) and (13) can be readily generalized to the case where a multiplicity of scattering experiments are performed in each of which the wavevector \mathbf{k}_1 of the probing wave is varied in direction and/or temporal frequency. The maximum likelihood estimate is returned via maximization of the likelihood function obtained as the superposition of the expressions (13) computed for all probing wavevectors \mathbf{k}_1 .

IV. DISCUSSION AND FUTURE WORK

First, in this letter, a translation property was derived describing the field scattered by a known buried object placed at distinct locations in the soil. Next, the property was used to derive the maximum likelihood algorithm for detection and estimation of the location of a known buried object from noisy scattered electromagnetic field

measurements. It was shown that the algorithm attained the form (13), which could be efficiently implemented with use of a fast Fourier transform algorithm. Even though the case of a two-layer background medium was considered, the critical information for the derivations in the letter is that the Green functions in the Lippmann-Schwinger equation (2) satisfy the property $\overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') = \overline{\mathbf{G}}(\mathbf{x} - \mathbf{x}', z, z')$. The same property is satisfied by the Green functions for arbitrary background media that vary only in the vertical direction. Therefore, both the field translation property (7) and the expression in (13) for the likelihood function are valid for such background media.

Further research in this area seems appropriate. Issues to be addressed in the future include the derivation of proper detection and location estimation algorithms for the cases of probing with arbitrary (nonplane) waves and arbitrary measurement configurations. Another avenue of future research is in the direction of efficient time-domain algorithms for underground probing with electromagnetic pulses of very short duration. For such pulses, frequency-domain algorithms such as the one proposed in this letter may not exploit the entire data spectrum efficiently and direct time-domain may lead to more efficient implementations. This and related research is currently pursued and its results will be announced shortly.

REFERENCES

- [1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory; Part I*. New York: Wiley, 1968.
- [2] W. C. Chew, *Waves and Fields in Inhomogeneous Media*. New York: Van Nostrand Reinhold, 1990, sec. 7.4.1.
- [3] G. A. Tsihrintzis and A. J. Devaney, "Maximum likelihood estimation of object location in diffraction tomography—Part II: Strongly scattering objects," *IEEE Trans. Signal Processing*, vol. 39, pp. 1466–1470, June 1991.
- [4] A. J. Devaney and G. A. Tsihrintzis, "Maximum likelihood estimation of object location in diffraction tomography," *IEEE Trans. Signal Processing*, vol. 39, pp. 672–682, Mar. 1991.

Higher-Order (Nonlinear) Diffraction Tomography: Reconstruction Algorithms and Computer Simulation

George A. Tsihrintzis, *Member, IEEE*, and Anthony J. Devaney, *Member, IEEE*

Abstract—The usual propagation transform of diffraction tomography is generalized into higher-order (nonlinear) propagation transforms via use of the Born series as the data-generating model in scattering experiments. Nonlinear tomographic reconstruction algorithms are developed for inversion of scattered field data modeled up to an arbitrarily large (possibly infinite) number of terms in the Born series. A computer simulation study is included to illustrate the performance of the algorithms for the case of scattering objects with cylindrical symmetry.

Index Terms—Born series, diffraction tomography, image reconstruction, Volterra series.

I. INTRODUCTION

A. Historical Development of Diffraction Tomography

IN INVERSE scattering, an unknown scattering object is probed with known waves with the goal to deduce (reconstruct) a quantitative estimate of the object structure from measurements of the wavefields diffracted by it [7], [9], [42]. Applications from a number of different scientific disciplines, such as crystal structure determination [36], medical ultrasound tomography [25], acoustic and electromagnetic underground surveying [17], [75]–[77], optical and coherent X-ray microscopy [39], and elastic wave inverse scattering [20] can be addressed within the same unified mathematical theory of inverse scattering.

The structure determination objective of inverse scattering usually consists of an attempt to estimate the spatial distribution of the complex-valued index of refraction of the object by inverting the mathematical mapping relating the probing wave, the refraction index, and the measurable total wave. This objective is nontrivial to achieve due to the inherent nonuniqueness and nonlinearity of the mapping from index of refraction to scattered wave in any single scattering experiment [13]. The nonuniqueness issue can be partially addressed by employing a multiplicity of experiments, where the object is probed from several incident wave directions, and the full scattering data set is then available for the inversion. However, the issue of nonlinearity is significantly harder to address and, to date, research has

only produced mathematical results or computationally intensive iterative algorithms [2], [5], [6], [11], [12], [26], [28], [31], [32], [43], [70], [81] as opposed to practically implementable reconstruction algorithms.

Over the past 20 years, an alternative approach to the Inverse Scattering problem has been employed based on certain linearizing approximations [15], [18], [79], [80]. This approach has led to an expanded discipline within the regime of tomography, known as *diffraction tomography (DT)*. The first application of linearized inverse scattering seems to date back in 1912, when von Laue suggested that Friedrich and Knipping try diffracting X-rays by crystals in order to test the hypothesis that X-rays had wavelengths on the order of 10^{-10} m. The experiment was successful and led, within less than a year, to the first structure determination by X-ray methods (sodium chloride by Bragg) [3], [23]. Since then, X-ray probes are typically used to determine the structure of crystals using reconstruction algorithms based on the Born scattering model and measurement of far field intensity distributions [23], [36]. Indeed, the foundation of modern linearized DT lies in the generalized projection-slice theorem of (1.2), which forms the core of X-ray crystal structure determination and the basis of Wolf's pioneering work in 1969 [79].

In [79], Wolf showed how near field measurements can be employed to generate reconstructions within the Born model. Wolf's formulation was extended in 1974 by Iwata and Nagata [27] to determine the structure of a less restrictive class of scatterers satisfying the Rytov rather than the Born approximation. In 1979, Mueller *et al.* [41] employed the same concepts of the Born and Rytov approximations and presented Fourier interpolation-based algorithms for the inverse problem of ultrasound tomography, while in 1982 Devaney [15] derived an inversion algorithm, named "*the filtered backpropagation algorithm of DT*," for the inversion of full view, scattered field data under the Born or Rytov approximations. When scattering experiments are done at a wavelength λ , the filtered backpropagation algorithm returns an estimate of the unknown index of refraction distribution whose frequency content is the same as of the true distribution over a circular disk in Fourier space of radius $(2\pi/\lambda)\sqrt{2}$ and zero elsewhere [19].

The filtered backpropagation algorithm has been recognized as the one providing highest quality in the reconstructed images [45] and modifications to it have been presented by Devaney [17] in 1984, Deming and Devaney [10] in 1996, and Hansen and Johansen in 2000 to adjust it to the configurations employed in geophysical tomographic surveys. Tsihrintzis and Devaney addressed the reconstruction problem of linearized DT

Manuscript received January 20, 1999; revised March 23, 2000. This work was supported by ARO Grant DAAG55-97-1-0013 and by the Center for Subsurface Sensing and Imaging Systems (CenSSIS) of Northeastern University. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jeffrey Fessler.

G. A. Tsihrintzis is with the Department of Informatics, University of Piraeus, Piraeus, 18534, Greece (e-mail: geoatsi@unipi.gr).

A. J. Devaney is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: tonydev2@aol.com).

Publisher Item Identifier S 1057-7149(00)06912-8.

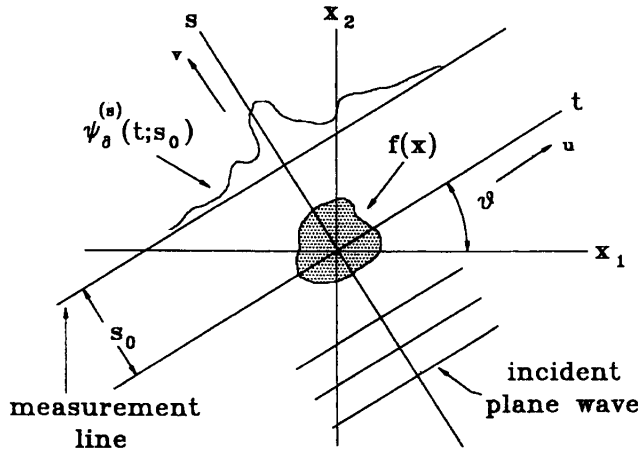


Fig. 1. Classical scan configuration of diffraction tomography.

from noisy scattered field data and showed that the optimum (Wiener) estimation filter attains again the form of a filtered backpropagation algorithm [66], [67]. Recently, a class of DT reconstruction algorithms with noise control was presented in [46]. Finally, iterative algorithms for inversion of an angularly limited set of noise-free linearized scattering data were presented by Ladas and Devaney in [33] and [34].

B. Mathematical Formulation of Diffraction Tomography

For two-dimensional (2-D) objects, the reconstruction problem of linearized DT consists of obtaining an estimate $\hat{f} \in L^2(R^2)$ of an unknown “object function” $f \in L^2(R^2)$ from its generalized projections (“propagations”) [19]

$$P_\theta f(t_0; s_0) = f(x) *_{\theta} \Gamma_\theta(x_0 - x); \quad \theta \in [0, 2\pi) \quad (1.1)$$

where $x_0 = t_0 u + s_0 v$, $t_0 \in R^1$ and s_0 : fixed, is a point on a “measurement line” that lies totally outside the support volume of the object. Here, θ is the angle between the t axis of a rotated Cartesian coordinate system (t, s) and the x_1 axis of a fixed Cartesian coordinate system (x_1, x_2) , and u, v denote unit vectors in the t, s directions respectively, as in Fig. 1. The transform kernel Γ_θ is such that the “generalized projection-slice theorem” holds [19]

$$\tilde{P}_\theta f(p; s_0) = \begin{cases} \frac{k}{m} e^{i(m-k)s_0} \tilde{f}[pu + (m-k)v], \\ \text{with } m = \sqrt{k^2 - p^2}, & \text{if } |p| \leq k \\ 0, & \text{otherwise} \end{cases} \quad (1.2)$$

where $k = (2\pi/\lambda)$ is the wavenumber of the probing wavefields and the tilde \sim denotes the Fourier transform

$$\tilde{P}_\theta f(p; s_0) = \int_{-\infty}^{\infty} dt_0 e^{-ipt_0} P_\theta f(t_0; s_0).$$

The generalized projection-slice theorem is illustrated in Fig. 2 and relates the one-dimensional (1-D) Fourier transform of a

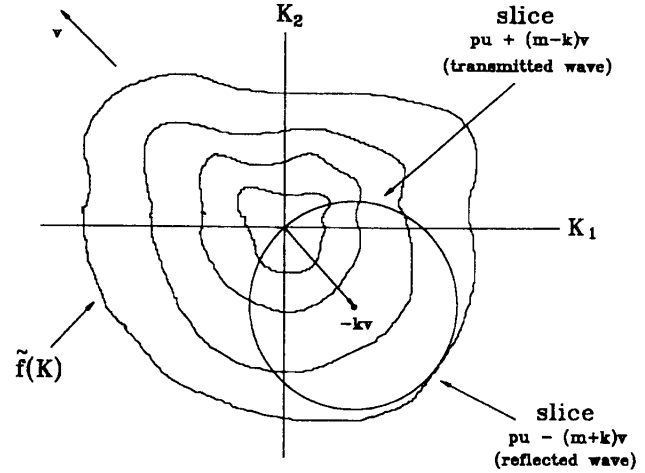


Fig. 2. Generalized projection-slice theorem of diffraction tomography.

generalized projection to a semicircular slice through the 2-D Fourier transform of the object function f itself.¹

The above formulation of the inverse problem of DT is based on the assumption that the object being probed interacts weakly with the incident waves, so that either the Born or the Rytov approximations [15] hold. In that case, the object function f is related to the tomographic data $y(t_0, \theta)$ through the linear transform (1.1), provided that the data are defined as [15]

$$\begin{aligned} P_\theta f(t_0; s_0) &= y(t_0, \theta) \\ &= \begin{cases} \frac{e^{-iks_0}}{ik} \psi_\theta^{(s)}(x_0) & \text{within the Born approximation} \\ \frac{1}{ik} \delta\Phi_\theta(x_0) & \text{within the Rytov approximation.} \end{cases} \end{aligned} \quad (1.3)$$

In (1.3), $\psi_\theta^{(s)}(x_0)$ is the scattered wavefield measured at the point $x_0 = t_0 u + s_0 v$, $t_0 \in R^1$, s_0 : fixed, while $\delta\Phi_\theta(x_0)$ is the deviation of the unwrapped complex phase (i.e., the unwrapped complex log) of the total wavefield measured at x_0 from the complex phase $ik s_0$ of the incident wavefield at x_0 .

The inverse problem of DT, as formulated above, has been extensively studied in the literature and reconstruction algorithms have been devised of the Fourier interpolation type [41], or the filtered backpropagation type [15], [18], and, recently, of an iterative nature [33], [34]. Of particular importance to this paper is the filtered backpropagation algorithm [15], [18], which provides a reconstruction \hat{f} that corresponds to a lowpass filtered version of the unknown object function f . In particular, the filtered backpropagation algorithm estimates \hat{f} as [15]

$$\hat{f}(x) = \frac{1}{2} \frac{1}{(2\pi)^2} \int_0^{2\pi} d\theta \int_{-k}^k dp |p| e^{ipt} e^{i(m-k)(s-s_0)} \tilde{P}_\theta \hat{f}(p; s_0) \quad (1.4)$$

¹The remainder of the circular slice is associated with the reflected field, i.e., the scattered field along the measurement line $s = -s_0$.

where $x = tu + sv$. It is readily shown that the estimate \hat{f} has the same spatial frequency content as the unknown f over a disk \mathcal{D} in Fourier space of radius $k\sqrt{2}$ and zero frequency content outside that disk. Since its first appearance [15], the filtered back-propagation algorithm has been modified and extended to other data acquisition geometries [17] and tested on computer simulated [14], [45] and experimental data [60], [65].

C. State of the Art and Present Contribution

Linearized DT has reached today the stage of being implemented in prototype commercial tomographic scanners for ultrasonic [60], [65], underground [48], [75], and optical [39] imaging systems. Particularly successful have been geophysical DT algorithms when applied to a range of underground imaging problems such as oil field prospecting and reservoir monitoring [78], locating underground tunnels between North and South Korea [75], and searching for dinosaur bones in the New Mexico desert [74]. The success of the linearized DT algorithms depends critically, however, on the two assumptions of linearity and availability of multiple experiments and in many cases, the linearity assumption fails, while different constraints (economic, safety, operating, geometric, or physical) limit the number of scattering experiments that can be performed and/or provide low signal-to-noise ratio data. Even though algebraic reconstruction techniques reduce the effect of availability of only a small number of scattering experiments, the effects of nonlinearity are much harder to combat and remain an issue of current research. To overcome these limitations a more modest inverse scattering problem was addressed by Tsihrintzis and Devaney, originally within the framework of linearized [21], [63], [65], [68], [69] and later exact [64], [65], [69] scattering theory. The goal of this more modest inverse problem, motivated in part by earlier work on a related problem in X-ray computed tomography [52], was to estimate the location of a *known* scatterer having unknown central location from noisy scattered field data. It was found that for monochromatic plane-wave probing the optimum (in the maximum likelihood sense) location estimate could be obtained via a filtered backpropagation algorithm, in which partial images formed by filtering and backpropagating scattered field data for different probing directions were coherently summed. The algorithm yields an *image* of the log likelihood function of the object's location and can be used for target detection and classification as well as target location estimation. The detection/estimation/classification procedure is optimum (in the maximum likelihood sense) for any given number of scattering experiments and returns good estimates even from a single experiment as long as the wavelength of the probing radiation is comparable with the typical dimensions of the target [21].

One practically important imaging situation arises when the object consists of a number of distinct scatterers. As pointed out by Azimi and Kak [1] and Slaney, Kak, and Larsen [54], even though each scatterer individually may be weak enough for validity of the Born approximate model, multiple scattering interactions among several scatterers degrade the performance of linearized DT reconstruction algorithms. The situation can be partially ameliorated if the reconstruction algorithms are based on higher-order (nonlinear) scattering models and, indeed, formal

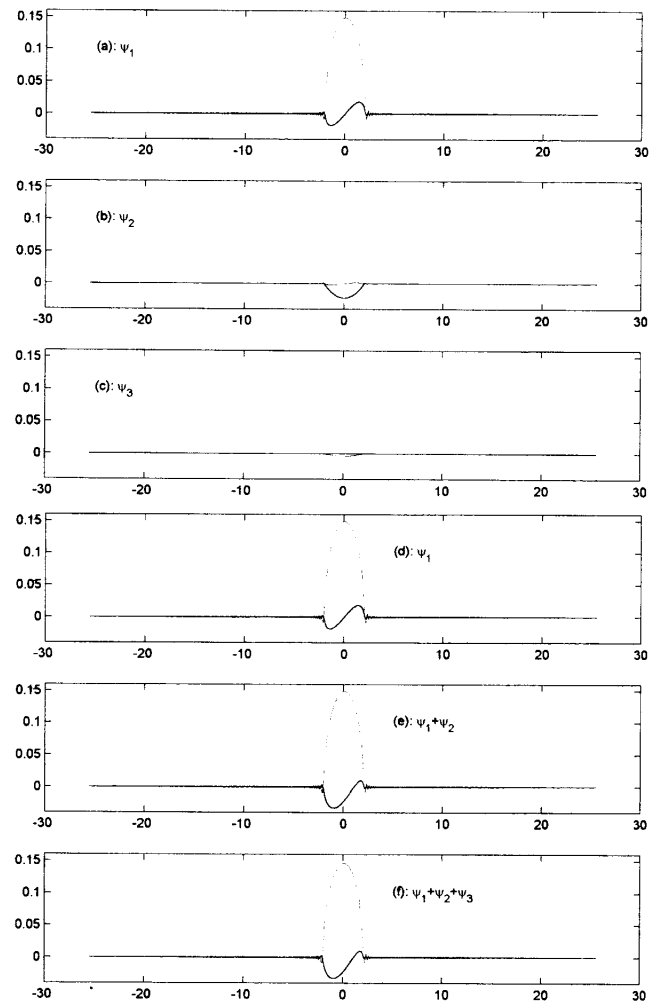


Fig. 3. Scattered fields in the first computer simulation: (a) the first Born term, (b) the second Born term, (c) the third Born term, (d) the first Born approximation, (e) the second Born approximation, and (f) the third Born approximation. The real (imaginary) part of the fields is plotted in solid (dashed) line

series solutions to the inverse scattering problem have been presented in the literature [22], [37], [38], [49]–[51], [55]–[57], [71], [72]. In [22], [37], and [38] more specifically, perturbative expansions of the scattering object's Fourier transform were utilized to develop DT reconstruction algorithms of arbitrary order, which contained linear reconstruction algorithms as special cases and effectively attained the form of nonlinear data filtering followed by a linear operation. The algorithms in [22], [37], [38] attempt to estimate samples of the Fourier transform of the object function and use them to obtain an estimate of the object function itself. On the other hand, recent attempts (see [4] and [47] and references therein) to invert a second-order scattering model have resulted in algorithms of the form of iterative numerical solutions of systems of quadratic equations and revealed significantly higher fidelity than their linear counterparts. In this paper, we address the reconstruction problem of DT within the framework of higher-order (nonlinear) scattering approximations and derive analytical reconstruction algorithms for the, so-called, “*classical scan configuration of DT*” of Fig. 1, that can be implemented via use of a fast Fourier transform algorithm and, in fact, make extensive use

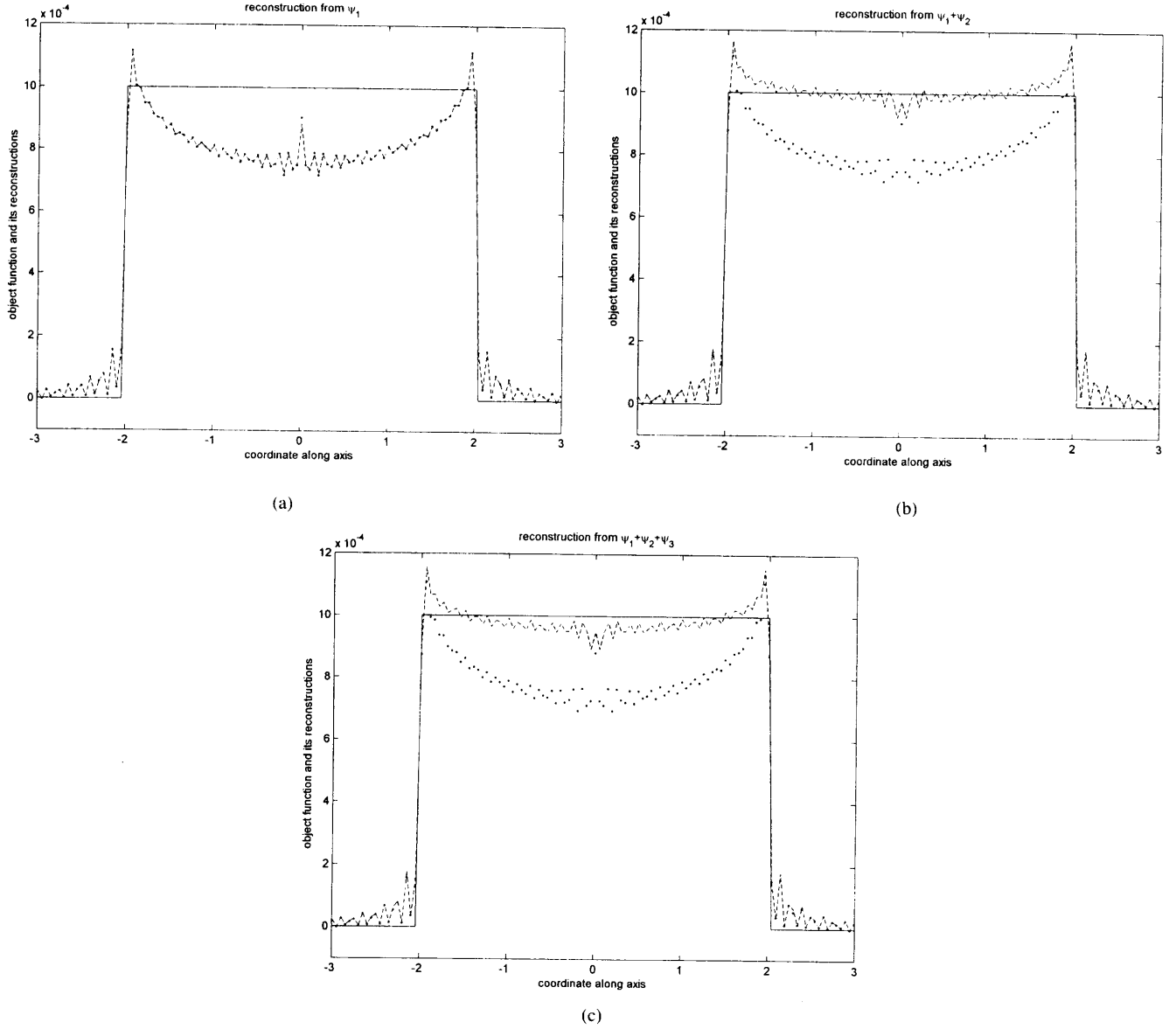


Fig. 4. Original object function (solid line) and reconstructions returned from the filtered backpropagation (dotted line) and the second-order algorithm (dashed line) in the first computer simulation: (a) reconstruction from the first-order Born approximation, (b) reconstruction from the second-order Born approximation, and (c) reconstruction from the third-order Born approximation

of the filtered backpropagation algorithm. The algorithms are illustrated in a computer simulation study.

More specifically, the paper is organized as follows: Section II defines nonlinear propagation transforms as the mathematical mapping from the object structure to the scattered wave data that arises from higher-order Born approximations. Section III is devoted to study of the inversion of nonlinear propagation transforms and a computer implementation, while Section IV summarizes the paper, draws conclusions, and points to future related research.

II. NONLINEAR PROPAGATION TRANSFORMS FOR DIFFRACTION TOMOGRAPHY

A. Configuration and Scattering Equations

Consider the data collection configuration illustrated in Fig. 1, where a known monochromatic plane wave

$\psi_0(x) = \exp(ik\langle v, x \rangle)$ of wavenumber k propagates in the direction of the unit vector v and is incident on a scattering object of support \mathcal{V} and the total wave is measured over a line located outside the object and perpendicular to the unit vector v . A point on the measurement line is denoted with the coordinate vector $x_0 = t_0 u + s_0 v$ where s_0 is the distance of the measurement line from the origin. The object is assumed to be embedded in a nonattenuating, homogeneous medium of wave velocity c_0 and wavenumber $k = (\omega/c_0)$ and characterized by its distribution of complex-valued index of refraction $n(x) = c_0/c(x)$, where $c(x)$ is the wave velocity distribution inside the scatterer.

The interaction of the incident wave with the object results in the formation of a wavefield $\psi(x)$ that satisfies the time-independent inhomogeneous Helmholtz equation

$$(\nabla^2 + k^2)\psi(x) = -2k^2 f(x)\psi(x) \quad (2.1)$$

where $f(x) = (1/2)[n^2(x) - 1]$ is the *object function*, the quantity whose determination is the ultimate goal of Inverse Scattering theory and DT. The solution of the Helmholtz equation (2.1) can be decomposed into the sum of the incident field and a scattered field in the form $\psi(x) = \psi_0(x) + \psi^{(s)}(x)$ where the scattered field component is related to the object function and total field via the integral equation

$$\psi^{(s)}(x) = -2k^2 \int_{\mathcal{V}} dx' f(x') \psi(x') G(x - x') \quad (2.2)$$

with $G(x - x') = -(i/4)H_0^{(1)}(k|x - x'|)$ the outgoing wave Green function to the Helmholtz equation [40].

Equation (2.2) provides a link between source and scattering problems in wave theory and shows that the scattered field *outside* the scattering region \mathcal{V} is uniquely determined from the total field *inside* \mathcal{V} . As a result, approximate scattering models, such as the Born models of the following section, seek to replace the “*induced source*” $f\psi$ in (2.2) with an approximate expression that leads to a tractable inverse problem. The first Born and Rytov approximations lead to linear models that are inverted via DT as outlined in the Introduction. The objective of this paper is to improve these linear models by including higher-order Born terms.

B. Born Series

Equation (2.2) implies a nonlinear mapping from the object function f to the scattered field $\psi^{(s)}$, which under certain weak scattering conditions can be expanded into a convergent Liouville–Neumann expansion, known as the Born series. The Born series is, thus, a perturbational expansion for the field scattered by an object f , obtained by introducing a smallness parameter ϵ and replacing the object function f in (2.1) and (2.2) by ϵf . The total field $\psi(x; \epsilon)$ inside the scattering region \mathcal{V} is assumed an analytic function of the parameter ϵ around the point $\epsilon = 0$, so that the Taylor series expansion around $\epsilon = 0$

$$\psi(x; \epsilon) = \psi_0(x) + \epsilon \psi_1(x) + \epsilon^2 \psi_2(x) + \dots \quad (2.3)$$

converges for sufficiently small ϵ . The terms $\epsilon \psi_1(x)$, $\epsilon^2 \psi_2(x)$, \dots are the first, second, \dots order corrections made to the incident field $\psi_0(x)$ to yield the total field $\psi(x; \epsilon)$. Clearly, this expansion is compatible with the fact that the scattered field $\psi^{(s)}(x; \epsilon) = \epsilon \psi_1(x) + \epsilon^2 \psi_2(x) + \dots$ is solely due to the presence of the scatterer with object function ϵf and in the absence of it (i.e., for $\epsilon = 0$) the total field $\psi(x; \epsilon = 0)$ reduces simply to the known incident field $\psi_0(x)$.

Inserting (2.3) into (2.2) and equating the coefficients of equal powers of ϵ on the two sides of the resulting expression, one obtains

$$\begin{aligned} \psi_1(x_0) &= -2k^2 \int_{\mathcal{V}} dx' G(x_0 - x') f(x') \psi_0(x') \\ \psi_2(x_0) &= -2k^2 \int_{\mathcal{V}} dx' G(x_0 - x') f(x') \psi_1(x') \\ &\vdots \\ \psi_n(x_0) &= -2k^2 \int_{\mathcal{V}} dx' G(x_0 - x') f(x') \psi_{n-1}(x') \\ &\vdots \end{aligned} \quad (2.4)$$

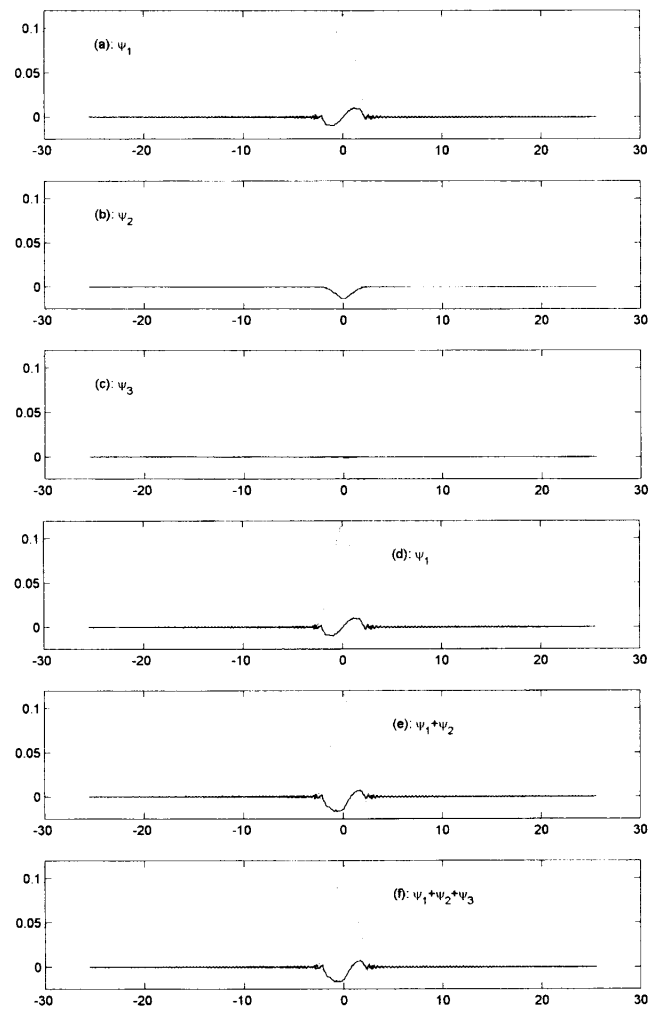


Fig. 5. Scattered fields in the second computer simulation: (a) the first Born term, (b) the second Born term, (c) the third Born term, (d) the first Born approximation, (e) the second Born approximation, and (f) the third Born approximation. The real (imaginary) part of the fields is plotted in solid (dashed) line

Clearly, the Born term of order n can be computed from the Born term of order $n - 1$, $n = 1, 2, 3, \dots$, by replacing the induced source $f\psi$ with the product $f\psi_{n-1}$. After computation of the Born terms up to some desired order N , the total field ψ resulting from the interaction of the incident wave ψ_0 with the scatterer f is computed as $\psi(x_0) = \psi_0(x_0) + \psi^{(s)}(x_0)$ with the scattered field approximated as

$$\psi^{(s)}(x_0) \approx \psi_1(x_0) + \psi_2(x_0) + \dots + \psi_N(x_0). \quad (2.5)$$

Of course, the assumption is made that the expansion (2.3) is valid for $\epsilon = 1$.²

The first-order Born approximation, simply referred to as the Born approximation [42], has been extensively employed in quantum mechanical scattering theory [24], [42]. From (2.4), the Born approximation estimates the scattered field $\psi^{(s)}$ by substituting the incident field ψ_0 for the total field ψ in the integrand in the right hand side of (2.2) and therefore requires that 1) $|\psi^{(s)}| \ll |\psi_0|$ inside the scatterer support \mathcal{V} and 2) the volume of scatterer support \mathcal{V} itself be small [30].

²For conditions under which the Born series converges, see [7], [9].

It is seen that the Born approximation is quite restrictive, since it requires both the size and the strength of the scattering object to be “small.” Additionally, it is a low frequency approximation, since it requires the frequency of the incident wave to be low. The Born approximation has been extensively used for predictions in direct [8], [62] and inverse [16] scattering theory. Its domain of validity, has been extensively explored in the literature [29], [30], [35], [44], [58], [59], [73] and compared with experimental results [39], [61].

C. Nonlinear Propagation Transforms of 2-D DT

Consider again the classical scan configuration of Fig. 1, but assume that the scattered field along the line $x_0 = t_0 u + s_0 v$, $t_0 \in \mathbb{R}^1$, with s_0 fixed, is well approximated by the first N terms of the Born series, as in (2.5). Let $G(x - x')$ denote the outgoing wave Green function to the Helmholtz operator and consider its Weyl expansion [40]

$$\begin{aligned} G[(t - t')u + (s - s')v] \\ &= -\frac{i}{4\pi} \int_{-\infty}^{\infty} dp e^{ip(t-t')} \frac{e^{im|s-s'|}}{m} \\ &\approx -\frac{i}{4\pi} \int_{-k}^k dp e^{ip(t-t')} \frac{e^{im|s-s'|}}{m} \\ &\equiv G_L[(t - t')u + (s - s')v] \end{aligned} \quad (2.6)$$

where

$$m = \begin{cases} \sqrt{k^2 - p^2}, & \text{if } |p| \leq k \\ i\sqrt{p^2 - k^2}, & \text{if } |p| > k. \end{cases}$$

The Weyl expansion in (2.6) decomposes the Green function $G(x - x')$ into a superposition of both homogeneous (corresponding to $|p| \leq k$) and evanescent (corresponding to $|p| > k$) plane waves. The evanescent plane waves decay exponentially with $|s - s'|$ and the approximation G_L on the far right hand side of (2.6) is a lowpass approximation to G that consists only of homogeneous plane wave components and is valid for $|s - s'| \gg \lambda$.

Next, define the Hilbert spaces \mathcal{F} , \mathcal{F}_L , and \mathcal{P} as follows: \mathcal{F} is $L^2(\mathcal{D}_a)$, where \mathcal{D}_a is a disk of radius a in \mathbb{R}^2 that includes the scattering region \mathcal{V} . \mathcal{F}_L is the subspace of \mathcal{F} consisting of the functions in \mathcal{F} lowpass filtered down to a disk of radius $k\sqrt{2}$ in Fourier space, i.e.,³

$$\mathcal{F}_L = \left\{ (f_L|_{\mathcal{D}_a}): \exists (f|_{\mathcal{D}_a}) \in \mathcal{F} \text{ such that } \begin{aligned} \tilde{f}_L(K) &= \begin{cases} \tilde{f}(K), & \text{for } |K| \leq k\sqrt{2} \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \right\}.$$

Finally, \mathcal{P} is the space of functions defined on $\mathbb{R}^1 \times [0, 2\pi)$ whose frequency content with respect to the first variable is zero outside the $[-k, k]$ interval.

³Since the functions f in \mathcal{F} are compactly supported, their Fourier transform is an entire function. Therefore, for given f_L , the corresponding f is unique and can, in principle, be obtained by analytically continuing the Fourier transform of f_L . The lowpass filtering is essentially a regularization procedure that makes the inverse of the linear propagation transform continuous.

Define the nonlinear operators

$$\mathcal{G}_n: \mathcal{F}_L \rightarrow \mathcal{P} \quad (2.7)$$

where

$$(\mathcal{G}_n f_L)(t_0, \theta) = \frac{e^{-ik s_0}}{ik} \psi_n^L(t_0 u + s_0 v, \theta) \quad (2.8)$$

and $\psi_n^L(t_0 u + s_0 v, \theta)$ is the n th term of the Born series to the scattered field along the measurement line $x_0 = t_0 u + s_0 v$, lowpass filtered with respect to t_0 down to the frequency interval $[-k, k]$. From (2.4), it is seen that \mathcal{G}_n is a homogeneous Volterra operator of order n

$$\begin{aligned} (\mathcal{G}_n f_L)(t_0, \theta) &= \frac{e^{-ik s_0} (-2k^2)^n}{ik} \cdot \int_{\mathcal{V}} dx^{(1)} G_L(x_0 - x^{(1)}) \\ &\quad \cdot \int_{\mathcal{V}} dx^{(2)} G(x^{(1)} - x^{(2)}) \cdots \int_{\mathcal{V}} dx^{(n)} \\ &\quad \cdot G(x^{(n-1)} - x^{(n)}) e^{ik \langle v, x^{(n)} \rangle} f(x^{(1)}) \\ &\quad \cdot f(x^{(2)}) \cdots f(x^{(n)}). \end{aligned} \quad (2.9)$$

In (2.9), f is the function in \mathcal{F} obtained by analytic continuation of the Fourier transform of f_L . The choice of G_L (as opposed to G) in the first integration is practically valid as long as the measurement offset s_0 is on the order of at least several wavelengths for attenuation of evanescent wave components in the scattered fields outside of the object ($s_0 \gg \lambda$). However, evanescent wave components inside the object will, in general, give rise to homogeneous wave components of higher-order in multiple scattering interactions; therefore, the full G needs to be utilized in the remaining integrations.

To write the nonlinear propagation transforms in (2.9) in a computationally attractive form, use is made of the Weyl expansion of (2.6). We have

$$\begin{aligned} (\mathcal{G}_1 f_L)(t_0, \theta) &= \frac{e^{-ik s_0}}{ik} \frac{2ik^2}{4\pi} \iint_{\mathcal{V}} dt^{(1)} ds^{(1)} \int_{-k}^k dp \\ &\quad \cdot e^{ip(t_0 - t^{(1)})} \frac{e^{im(s_0 - s^{(1)})}}{m} f(t^{(1)}u + s^{(1)}v) e^{iks^{(1)}} \\ &= \frac{ke^{-ik s_0}}{2\pi} \int_{-k}^k dp e^{ipt_0} \frac{e^{im s_0}}{m} \iint_{\mathcal{V}} dt^{(1)} ds^{(1)} \\ &\quad \cdot e^{-ipt^{(1)}} e^{-i(m-k)s^{(1)}} f(t^{(1)}u + s^{(1)}v) \\ &= \frac{1}{2\pi} \int_{-k}^k dp e^{ipt_0} \frac{ke^{i(m-k)s_0}}{m} \tilde{f}[pu + (m-k)v] \end{aligned} \quad (2.10)$$

where \tilde{f} denotes the 2-D Fourier transform of f . Equation (2.10) is simply a statement of the generalized projection-slice theorem of (1.2) and verifies that the proposed (nonlinear) propagation transforms include the usual linear propagation transform as a special case. It is seen that the linear propagation transform (2.10) of an object function f is directly computable via use of a fast Fourier transform algorithm (FFT). The nonlinear propagation transforms $\mathcal{G}_n f$ (2.9) are not directly computable

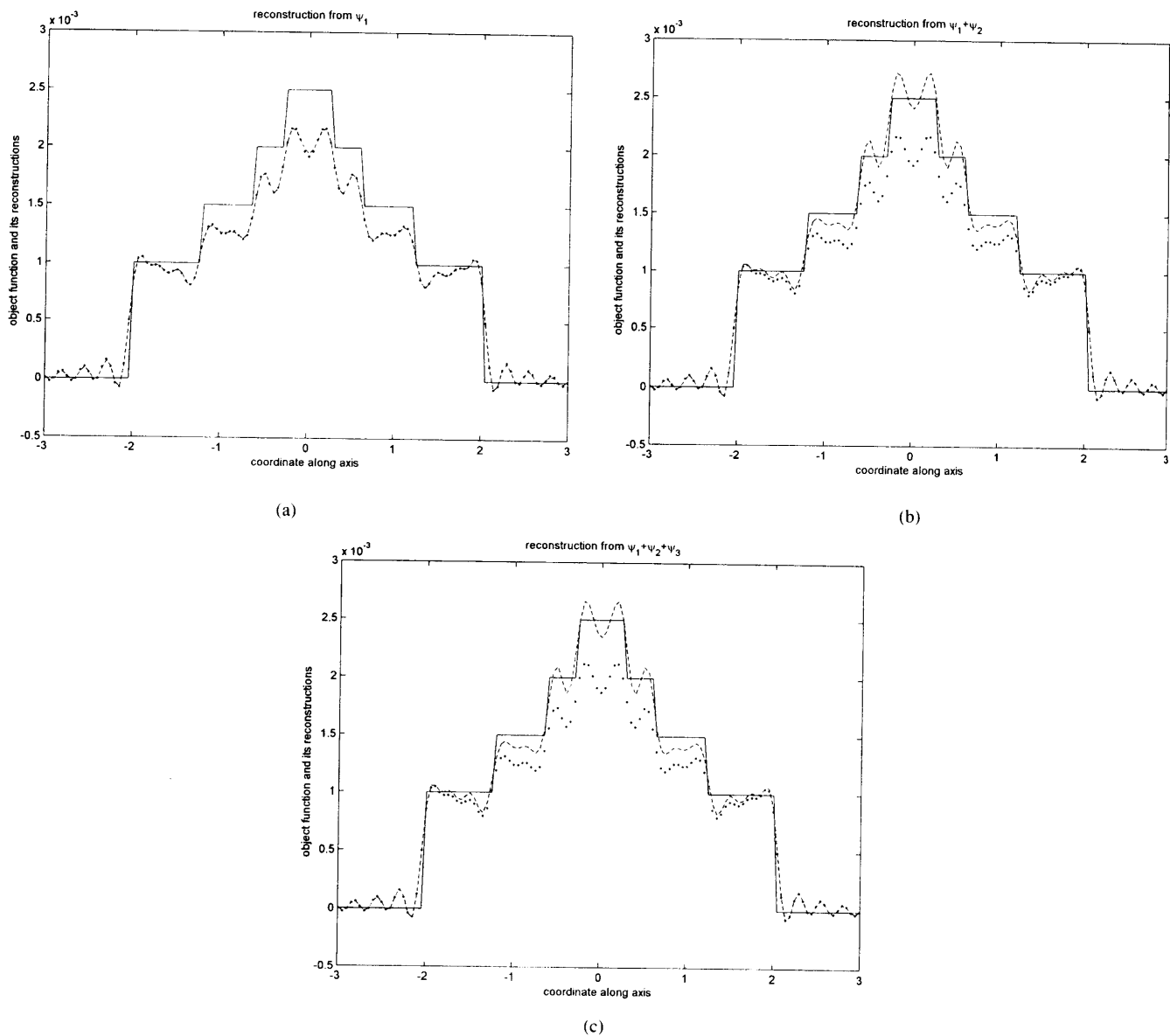


Fig. 6. Original object function (solid line) and reconstructions returned from the filtered backpropagation (dotted line) and the second-order algorithm (dashed line) in the second computer simulation: (a) reconstruction from the first-order Born approximation, (b) reconstruction from the second-order Born approximation, and (c) reconstruction from the third-order Born approximation

using a FFT, since they require computation of the Fourier transform of the object function f over a set of complex frequencies. However, if the lowpass approximations to the Weyl expansion in (2.6) is used, a valid approximation for sufficiently high wavenumber, the nonlinear propagation transforms can also be computed approximately by FFT-based algorithms.

Define now the linear operator

$$\mathcal{B}: \mathcal{P} \rightarrow \mathcal{F}_L \quad (2.11)$$

with

$$(\mathcal{B}d)(x = tu + sv) = \frac{1}{2} \frac{1}{(2\pi)^2} \int_0^{2\pi} d\theta \int_{-k}^k dp |p| \cdot e^{ipt} e^{i(m-k)(s-s_0)} \tilde{d}(p, \theta). \quad (2.12)$$

The operator $(\mathcal{B}|\mathcal{P})$ is the inverse of $(\mathcal{G}_1|\mathcal{F}_L)$, as stated in (1.4).

D. Cylindrically Symmetric Case

A particularly important imaging situation is that of scattering objects with cylindrical symmetry, that is scatterers whose properties vary only with the radial distance from the scatterer center and, thus, their object function is of the form

$$f(x) = f_r(r), \quad r = |x|. \quad (2.13)$$

This important situation arises in practical applications such as reconstruction of the structure of optical fibers or large molecules, but also provides a paradigm for both analytical and computer simulation study of DT reconstruction algorithms. Both the filtered backpropagation algorithm (1.4) and the nonlinear propagation transforms attain a simpler form that makes use of 1-D Fourier-Bessel transforms.

Indeed, let

$$\tilde{f}_r(|K|) = 2\pi \int_0^\infty dr r J_0(|K|r) f_r(r). \quad (2.14)$$

Then, the Fourier transform of the object function f becomes

$$\tilde{f}(K) = \tilde{f}_r(|K|) \quad (2.15)$$

and the generalized projection-slice theorem (1.2) reads⁴

$$\tilde{P}_\theta f(p; s_0) = \begin{cases} \frac{k}{m} e^{i(m-k)s_0} \tilde{f}_r[\sqrt{2k(k-m)}], & \text{if } |p| \leq k, \\ 0, & \text{otherwise} \end{cases} \quad (2.16)$$

where the generalized projections $P_\theta f$ are independent of the view angle θ . The filtered backpropagation algorithm is now put in the form

$$\hat{f}(x) = \frac{1}{2\pi} \int_0^k dp p e^{i(k-m)s_0} J_0[r\sqrt{2k(k-m)}] \cdot \tilde{P}_\theta f(p; s_0), \quad r = |x| \quad (2.17)$$

where the integration with respect to view angle θ has been eliminated.⁵ Finally, the nonlinear propagation transforms can be computed in a similar yet more complicated, manner.

III. INVERSION OF NONLINEAR PROPAGATION TRANSFORMS

A. Theory

Define the inhomogeneous Volterra operator of order N

$$\mathcal{G}^N = \mathcal{G}_1 + \mathcal{G}_2 + \cdots + \mathcal{G}_N: \mathcal{F}_L \rightarrow \mathcal{P} \quad (3.1)$$

where

$$(\mathcal{G}^N f_L)(t_0, \theta) = \frac{e^{-ik s_0}}{ik} (\psi_1^L + \psi_2^L + \cdots \psi_N^L)(t_0 u + s_0 v, \theta) \quad (3.2)$$

and $\psi_n^L(t_0 u + s_0 v, \theta)$, $n = 1, 2, \dots$, is the n th term of the Born series to the scattered field along the measurement line $x_0 = t_0 u + s_0 v$, lowpass filtered with respect to t_0 down to the frequency interval $[-k, k]$. The existence of a local inverse of the operator \mathcal{G}^N and its representation by a convergent series is guaranteed by the existence of the inverse $(\mathcal{B}|\mathcal{P})$ of the linear term $(\mathcal{G}_1|\mathcal{F}_L)$ [55].⁶ Specifically, there exists a region in the data space \mathcal{P} , on which the operator

$$\mathcal{B}_1 + \mathcal{B}_2 + \cdots + \mathcal{B}_n + \cdots \quad (3.3)$$

is the inverse of \mathcal{G}^N , where \mathcal{B}_n is a homogeneous Volterra operator of order n in the data.

⁴More generally

$$\tilde{f}_1(pu + s_0 v) = \begin{cases} \frac{ik^2}{m} e^{im s_0} \tilde{f}_r[\sqrt{2k(k-m)}], & \text{if } |p| \leq k, \\ \frac{ik^2}{m} e^{im s_0} \tilde{f}_r(|p|\sqrt{2}), & \text{if } |p| > k \end{cases}$$

which includes evanescent wavemodes in the scattered field data.

⁵In the case of noisy data, the measurements at each view angle are generally distinct and, thus, the angular integration cannot be eliminated. Equation (2.17) is no longer the statistically optimal strategy, but one needs to use the more general algorithms in [48] and [68].

⁶The stability (continuity) of the inverse operator series is guaranteed by the boundedness of the operators \mathcal{B} and \mathcal{G}_n , [7], [9].

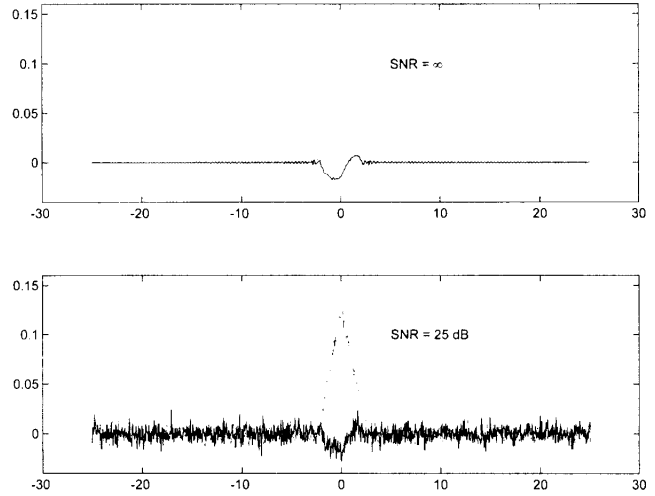


Fig. 7. Scattered fields in the second computer simulation: (top plot) noise-free third Born approximation and (bottom plot) typical noisy measurement at a signal-to-noise ratio of 25 dB. The real (imaginary) part of the fields is plotted in solid (dashed) line

More specifically

$$\mathcal{B}_1 = \mathcal{B} \quad (3.4)$$

$$\mathcal{B}_2 = -\mathcal{B}\mathcal{G}_2\mathcal{B} \quad (3.5)$$

$$\mathcal{B}_3 = \mathcal{B}\mathcal{G}_2\mathcal{B}\mathcal{G}_2\mathcal{B} - \mathcal{B}\mathcal{G}_3\mathcal{B} \quad (3.6)$$

\vdots

It is observed that the terms $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3$, etc. are *universal* in the sense that they do not depend on the order N of the forward model \mathcal{G}^N . Thus, the operator series in (3.3) can be truncated to desired order n and return DT inversion algorithms of order n in the data. For example, the usual filtered backpropagation algorithm provides a universal DT inversion algorithm of first-order (linear) in the data. The general Volterra theory of nonlinear systems can be applied to provide bounds on the error resulting from truncation of the series (3.3) at an arbitrary term. However, these bounds are functionals of the object function and cannot be computed beforehand. A practical approach to determine the term at which to terminate the series (3.3) is to monitor the norm of each new term in (3.3) and stop when it has become sufficiently small.

A second observation is that if the data are sufficiently modeled by the first Born term, i.e., $\psi^{(s)} = \psi_1$ and $\psi_2 = \psi_3 = \cdots = 0$, then each higher-order term in series (3.3) returns an identically zero contribution to image reconstruction, as close examination of the expressions in (3.4)–(3.6) reveals. This fact is also verified in the simulations of this section. A final observation is that the resulting algorithms are effectively of the form of *nonlinear* data filtering followed by backpropagation. Indeed, the series (3.3) can be written as

$$\mathcal{B}(\mathcal{I} - \mathcal{G}_2\mathcal{B} + \mathcal{G}_2\mathcal{B}\mathcal{G}_2\mathcal{B} - \mathcal{G}_3\mathcal{B} - \cdots) \quad (3.7)$$

where $\mathcal{I}: \mathcal{P} \rightarrow \mathcal{P}$ is the identity operator. The operator in brackets in (3.7) maps $\mathcal{P} \rightarrow \mathcal{P}$ and, thus, this term is a higher-order nonlinear Volterra filter applied on the data. Therefore, the series (3.3) gives rise to an entire class of *nonlinear filtered backpropagation* algorithms for inversion of DT data.

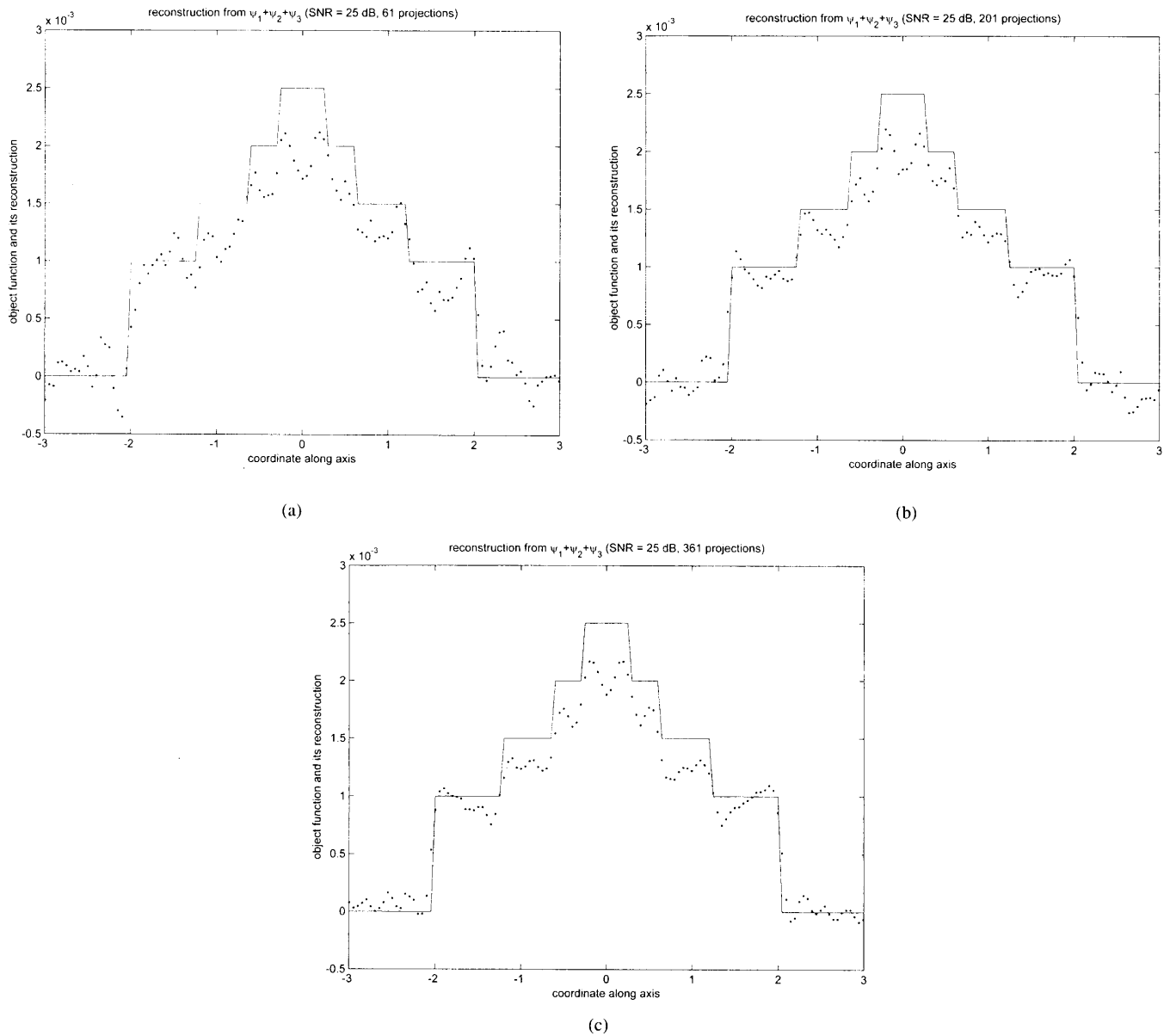


Fig. 8. Original object function (solid line) and reconstructions returned from the filtered backpropagation algorithm (dash-dotted line) as applied to 61 [Fig. 8(a)], 201 [Fig. 8(b)], and 361 [Fig. 8(c)] noisy projections measured at angles equally spaced over 360° .

B. Computer Simulation of Higher-Order DT

The proposed nonlinear diffraction tomographic reconstruction algorithms were implemented and studied in two different computer simulations. In both simulations, the scattering object exhibited cylindrical symmetry and, in the case of noiseless data, could be reconstructed from a single scattering experiment.

Computer Simulation 1: The scattering object consisted of a single disk of radius 2, within which the object function was constant and equal to 0.001, i.e.,

$$f(x) = \begin{cases} 0.001, & \text{if } |x| \leq 2 \\ 0, & \text{elsewhere.} \end{cases} \quad (3.8)$$

The wavenumber was set to $k = 12\pi$, corresponding to a wavelength $\lambda = (1/6)$, so that the object was 12 wavelengths in di-

ameter. Fig. 3 show plots of the real (solid line) and imaginary (dashed line) parts of the first [Fig. 3(a)], second [Fig. 3(b)], and third [Fig. 3(c)] Born terms and the corresponding first- [Fig. 3(d)], second- [Fig. 3(e)], and third-order Born approximations to the scattered field for a measurement distance of $s_0 = 0$. Fig. 4 show the object function reconstructions returned by the filtered backpropagation (dotted line) and the second-order (dashed line) and compare it to the original object function (solid line). Specifically, the data consisted of 1) the first Born approximation [Fig. 4(a)], 2) the second Born approximation [Fig. 4(b)], and 3) the third Born approximation [Fig. 4(c)] to the scattered field. Clearly, the second-order algorithm returns the same reconstruction as the filtered backpropagation algorithm when the data consists of only the first Born term, but outperforms the filtered backpropagation algorithm when the data consists of the second- or third-order Born approximation.

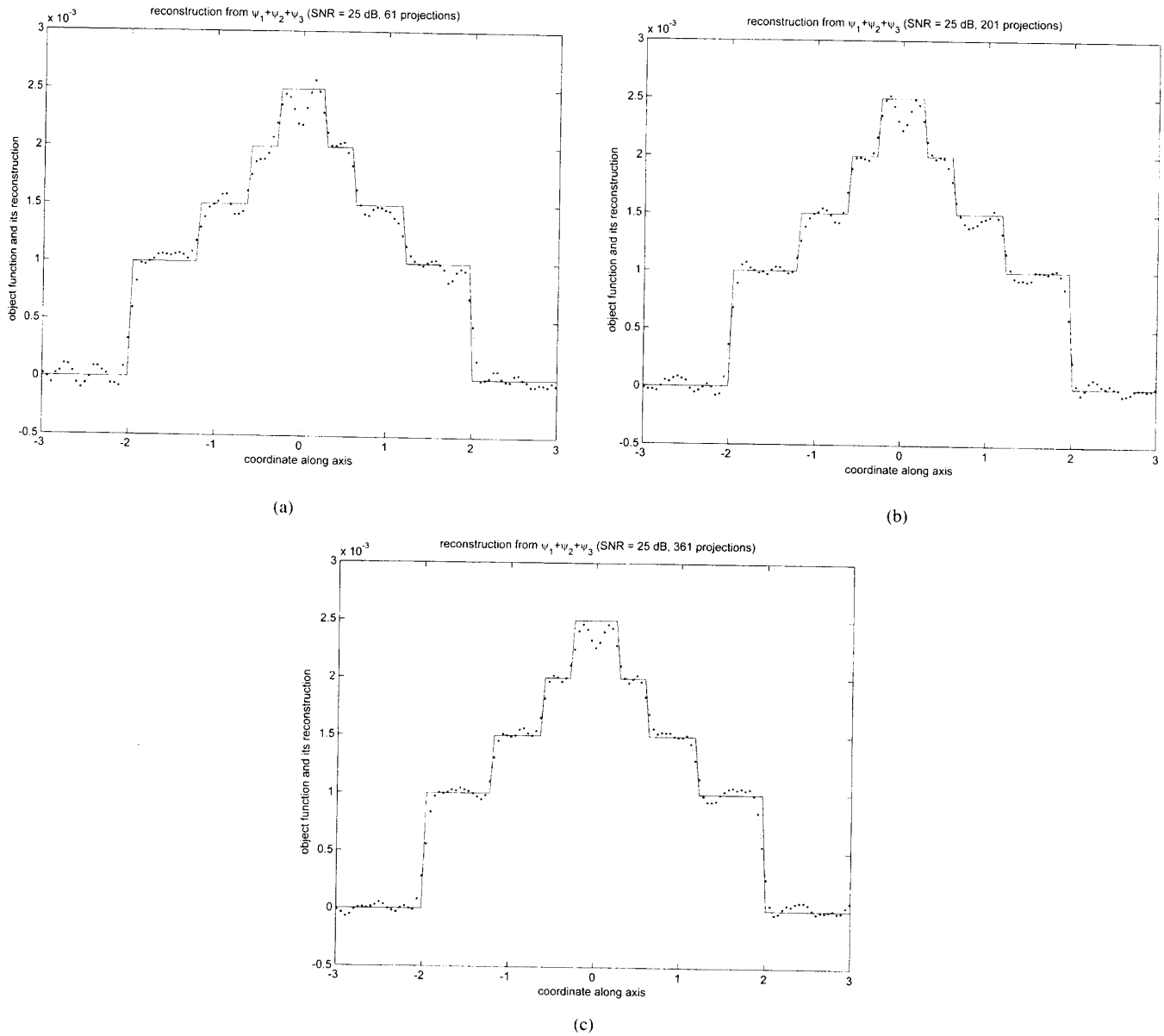


Fig. 9. Original object function (solid line) and reconstructions returned from the second-order inversion algorithm (dash-dotted line) as applied to 61 [Fig. 9(a)], 201 [Fig. 9(b)], and 361 [Fig. 9(c)] noisy projections measured at angles equally spaced over 360° .

Computer Simulation #2: The scattering object consisted of a circular core and three concentric circular coatings and constitutes a good model for small optical fibers. Specifically

$$f(x) = \begin{cases} 0.0025, & \text{if } |x| \leq 0.25 \\ 0.002, & \text{if } 0.25 \leq |x| \leq 0.6 \\ 0.0015, & \text{if } 0.6 \leq |x| \leq 1.2 \\ 0.001, & \text{if } 1.2 \leq |x| \leq 2 \\ 0, & \text{elsewhere.} \end{cases} \quad (3.9)$$

The wavenumber was equal to $k = 6\pi$, corresponding to a wavelength $\lambda = (1/3)$ and the measurement distance was set to $s_0 = 0$. Fig. 5 show plots of the first three Born terms and approximations to the scattered field, in correspondence with Fig. 3. Clearly, the contribution of the third Born term is negligible, when compared to the first two Born terms; thus, termination of the Born series at its second term seems to be valid. Fig. 6 show the object function reconstructions returned by the filtered

backpropagation and the second-order algorithms, in correspondence with Fig. 4. Clearly, the second-order algorithm returns the same reconstruction as the filtered backpropagation algorithm in the case of data consisting of only the first Born term. In the other cases, however, the filtered backpropagation algorithm misses the core of the object, while the second-order algorithm returns a more accurate reconstruction and reveals the core.

At this point, we need to stress that the oscillations observed in the object function reconstructions in areas where the original object function is uniform are due to the fact that the reconstruction algorithms are effectively sharp lowpass filters. Therefore, the observed oscillations are Gibbs phenomena that can be eliminated via use of a smoother lowpass filter. Such an approach has been followed successfully in [14]; however, the resulting reconstructions, even though smoother, are no longer optimum in the least-squares sense.

Next, we study the performance of the proposed algorithms when the input data are limited in both angular coverage and signal-to-noise ratio. In Fig. 7, we show the third-order Born approximation $\psi_1 + \psi_2 + \psi_3$ to the scattered field $\psi^{(s)}$ (top plot) and a typical noisy measurement of it (bottom plot). In the bottom plot, the noise was an additive white Gaussian process of variance σ^2 . The variance was chosen so as to set the signal-to-noise ratio, defined as $(\int_{-\infty}^{\infty} |\psi^{(s)}(t_0 u + s_0 v)|^2 dt_0 / \sigma^2)$, to 25 dB, probably an unrealistically low value given published experimental studies on DT.

Fig. 8 compare the original object function and its reconstruction returned by the filtered backpropagation algorithm⁷ as applied to 61 [Fig. 8(a)], 201 [Fig. 8(b)], and 361 [Fig. 8(c)] projections measured at angles equally spaced over 360 degrees. Clearly, the reconstructions are "close" to the reconstruction obtained from noise-free, full-view data in Fig. 6(c). In fact, the signal-to-noise ratio in the reconstructions has improved because of the lowpass filtering effected on the data. As the number of tomographic projections increases, the signal-to-noise ratio improves further as can be seen in Fig. 8(c). The robustness to noise and limited angular coverage of the filtered backpropagation algorithm is not unknown in the relevant literature (e.g., see [21, Figs. 12 and 13, p. 680] for a simulation study or the cited literature for experimental performance evaluations).

Fig. 9 compare the original object function and its reconstruction returned by the second-order inversion algorithm, in which only 61 [Fig. 9(a)], 201 [Fig. 9(b)], and 361 [Fig. 9(c)] projections are measured. Clearly, the reconstructions are "close" to the reconstruction obtained from noise-free, full-view data in Fig. 6(c) and outperform the filtered backpropagation algorithm in the corresponding simulations.

IV. SUMMARY, CONCLUSIONS, AND FUTURE RESEARCH

In this paper, the usual propagation transform of linearized diffraction tomography was generalized into higher-order (nonlinear) propagation transforms via use of the Born series as the data-generating model in scattering experiments. Tomographic reconstruction algorithms were developed for inversion of scattered field data and image formation. A computer simulation study was included to illustrate the performance of the algorithms in the case of imaging objects with cylindrical symmetry. From this study, it was found that a second-order algorithm returned essentially the same reconstruction as the usual filtered backpropagation algorithm when applied to first-order Born data, but clearly outperformed the filtered backpropagation algorithm when applied to data consisting of the second- or third-order Born approximation. Similar results were found when we processed measured projection data that were limited both in angular coverage and by noise.

It should be noted that the proposed nonlinear reconstruction algorithms are easily extended to a stochastic framework

in which the object function is considered as a realization from a wide sense stationary second-order process, the measured data are corrupted by noise, and an inversion algorithm is sought that is optimum in a mean squared error sense. In our previous work [68], [69], we addressed this problem within the framework of linearizing approximations and showed that the optimum linear inversion algorithm (Wiener filter) is a filtered backpropagation algorithm with a properly modified filter. It seems, therefore, that further investigation in that direction is due. Future relevant research may also follow the avenues of inversion of higher-order Rytov data (see [71]) or higher-order Diffraction Tomography with alternative data measurement configurations (e.g., ground-penetrating radar tomography). These and other research avenues are currently being pursued and the findings will be reported elsewhere.

REFERENCES

- [1] M. Azimi and A. C. Kak, "Distortion in diffraction tomography caused by multiple scattering," *IEEE Trans. Med. Imag.*, vol. MI-2, pp. 176–195, 1983.
- [2] S. Barkeshli and R. G. Lautzenheizer, "An iterative method for inverse scattering problems based on an exact gradient search," *Radio Sci.*, vol. 29, pp. 1119–1130, 1994.
- [3] W. H. Bragg and W. L. Bragg, *X-rays and Crystal Structure*. London, U.K.: Bell, 1915.
- [4] A. Brancaccio and G. R. Pierri, "Information content of Born scattered fields: Results in the circular cylindrical case," *J. Opt. Soc. Amer. A*, vol. 15, pp. 1909–1917, 1998.
- [5] S. Caorsi, G. L. Gragnani, S. Medicina, M. Pastorino, and A. Pinto, "A Gibbs random fields-based active electromagnetic method for non-invasive diagnostics in biomedical applications," *Radio Sci.*, vol. 30, pp. 291–301, 1995.
- [6] S. Caorsi, G. L. Gragnani, and M. Pastorino, "Two-dimensional microwave imaging by a numerical inverse scattering solution," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-38, pp. 981–989, 1990.
- [7] K. Chadon, D. Colton, L. Paivarinta, and W. Rundell, *An Introduction to Inverse Scattering and Inverse Spectral Problems*. Philadelphia, PA: SIAM, 1997.
- [8] L. A. Chernov, *Wave Propagation in a Random Medium*. New York: Dover, 1967.
- [9] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering*, 2nd ed. Berlin, Germany: Springer-Verlag, 1998.
- [10] R. W. Deming and A. J. Devaney, "Diffraction tomography for multi-monostatic ground penetrating radar imaging," *Inv. Prob.*, vol. 13, pp. 29–45, 1997.
- [11] P. M. Van der Berg and R. E. Kleinman, "A total variation enhanced modified gradient algorithm for profile reconstruction," *Inv. Prob.*, vol. 11, pp. L5–L10, 1995.
- [12] —, "A contrast source inversion method," *Inv. Prob.*, vol. 13, pp. 1607–1620, 1998.
- [13] A. J. Devaney, "Nonuniqueness in the inverse scattering problem," *J. Math. Phys.*, vol. 19, p. 1526, 1978.
- [14] —, "A computer simulation study of diffraction tomography," *IEEE Trans. Biomed. Eng.*, vol. BME-30, p. 377, 1982.
- [15] —, "A filtered backpropagation algorithm for diffraction tomography," *Ultrason. Imag.*, vol. 4, p. 336, 1982.
- [16] —, "Inversion formula for inverse scattering within the Born approximation," *Opt. Lett.*, vol. 7, p. 111, 1982.
- [17] —, "Geophysical diffraction tomography," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-22, p. 3, Jan. 1984.
- [18] —, "Reconstructive tomography with diffracting wavefields," *Inv. Prob.*, vol. 2, p. 161, 1986.
- [19] —, "The limited-view problem in diffraction tomography," *Inv. Prob.*, vol. 5, p. 501, 1989.
- [20] —, "Elastic wave inverse scattering," in *Elastic Waves and Ultrasonic Nondestructive Evaluation*, S. K. Datta, J. D. Achenbach, and Y. S. Rajapakse, Eds. New York: Elsevier, 1990.
- [21] A. J. Devaney and G. A. Tsihrintzis, "Maximum likelihood estimation of object location in diffraction tomography," *IEEE Trans. Signal Processing*, vol. 39, p. 672, Mar. 1991.

⁷The filtered backpropagation algorithm of (1.4) is no longer optimum as the number of tomographic projections available is finite. The optimum algorithm is of the form of a filtered backpropagation algorithm, where, however, the filter is no longer the "rho" filter appropriate for inversion of full-view data. The appropriate filter is not known in closed form, but can be computed via optimization. The reader is referred to [19] for further details.

- [22] A. J. Devaney and E. Wolf, "A new perturbation expansion for inverse scattering from three-dimensional finite-range potentials," *Phys. Lett.*, vol. 89A, no. 6, pp. 269–272, 1982.
- [23] J. P. Glusker and K. N. Trueblood, *Crystal Structure Analysis: A Primer*. New York: Oxford Univ. Press, 1972.
- [24] K. Gottfried, *Quantum Mechanics, Volume I: Fundamentals*. Reading, MA: Addison-Wesley, 1989.
- [25] J. F. Greenleaf, "Computerized tomography with ultrasound," *Proc. IEEE*, vol. 71, p. 330, 1983.
- [26] T. B. Hansen and P. M. Johansen, "Inversion scheme for ground penetrating radar that takes into account the air–soil interface," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 496–506, Jan. 2000.
- [27] J. M. Harris, "Diffraction tomography with arrays of discrete sources and receivers," *IEEE Trans. Geosci. Remote Sensing*, vol. 25, no. 4, p. 448, 1987.
- [28] K. Iwata and R. Nagata, "Calculation of refractive index distribution from interferograms using the Born and Rytov's approximation," *Jpn. J. Appl. Phys.*, vol. 14, p. 379, 1974.
- [29] N. Joachimowicz, C. Pichot, and J.-P. Hugonin, "Inverse scattering: An iterative numerical method for electromagnetic imaging," *IEEE Trans. Antennas Propagat.*, vol. AP-39, pp. 1742–1752, 1991.
- [30] M. Kaveh, M. Soumekh, and R. K. Mueller, "A comparison of Born and Rytov approximation," in *Acoustical Imaging*, H. Lee and G. Wade, Eds. New York: Plenum, 1981, vol. 11, p. 325.
- [31] J. B. Keller, "Accuracy and validity of the Born and Rytov approximation," *J. Opt. Soc. Amer.*, vol. 59, p. 1003, 1969.
- [32] R. E. Kleinman and P. M. Van der Berg, "A modified gradient method for two-dimensional problems in tomography," *J. Comput. Appl. Math.*, vol. 42, pp. 17–35, 1992.
- [33] —, "An extended range modified gradient technique for profile inversion," *Radio Sci.*, vol. 28, pp. 877–884, 1993.
- [34] K. T. Ladas and A. J. Devaney, "Generalized ART algorithm for diffraction tomography," *Inv. Prob.*, vol. 7, p. 109, 1991.
- [35] —, "Iterative methods in geophysical diffraction tomography," *Inv. Prob.*, vol. 8, pp. 119–132, 1992.
- [36] F. C. Lin and M. A. Fiddy, "On the issue of the Born–Rytov controversy: I: Comparing analytical and approximate expressions for the one-dimensional deterministic case," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1102–1110, 1992.
- [37] H. Lipson and W. Cochran, *The Determination of Crystal Structures*. Ithaca, NY: Cornell Univ. Press, 1966.
- [38] Z.-Q. Lu, "JKM perturbation theory, relaxation perturbation theory, and their applications to inverse scattering: Theory and reconstruction algorithms," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 33, pp. 722–730, 1986.
- [39] Z.-Q. Lu and Y.-Y. Zhang, "Acoustical tomography based on the second-order transform perturbation approximation," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 43, pp. 296–302, 1996.
- [40] M. H. Maleki, A. J. Devaney, and A. Schatzberg, "Tomographic reconstruction from optical scattered intensities," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1356–1363, 1992.
- [41] J. E. Molyneux and A. J. Witten, "Diffraction tomographic imaging in a monostatic measurement geometry," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 507–511, 1993.
- [42] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics, Parts I and II*. New York: McGraw-Hill, 1953.
- [43] R. K. Mueller, M. Kaveh, and G. Wade, "Reconstructive tomography and applications to ultrasonics," *Proc. IEEE*, vol. 67, p. 567, 1979.
- [44] R. G. Newton, *Scattering Theory of Waves and Particles*. Berlin, Germany: Springer-Verlag, 1982.
- [45] M. K. Nguyen and A. Mohammad-Djafari, "Bayesian approach with maximum entropy priors in image reconstruction from microwave scattered fields data," *IEEE Trans. Med. Imag.*, vol. 13, pp. 254–261, 1994.
- [46] M. L. Oristaglio, "Accuracy of the Born and Rytov approximations for the reflection and refraction at a plane interface," *J. Opt. Soc. Amer.*, vol. 2, p. 1987, 1985.
- [47] S. X. Pan and A. C. Kak, "A computational study of reconstruction algorithms for diffraction tomography," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, p. 1262, 1982.
- [48] X. Pan, "Unified reconstruction theory for diffraction tomography, with consideration of noise control," *J. Opt. Soc. Amer. A*, vol. 15, pp. 2312–2326, 1998.
- [49] R. Pierri and A. Brancaccio, "Imaging of a rotationally symmetric dielectric cylinder by a quadratic approach," *J. Opt. Soc. Amer. A*, vol. 14, pp. 2777–2785, 1997.
- [50] R. B. Pratt and M. H. Worthington, "The application of diffraction tomography to cross-hole seismic data," *Geophysics*, vol. 53, p. 1284, 1988.
- [51] R. T. Prosser, "Formal solutions of inverse scattering problems," *J. Math. Phys.*, vol. 10, pp. 1819–1822, 1969.
- [52] —, "Formal solutions of inverse scattering problems, II," *J. Math. Phys.*, vol. 17, pp. 1775–1779, 1976.
- [53] —, "Formal solutions of inverse scattering problems, III," *J. Math. Phys.*, vol. 21, pp. 2648–2653, 1980.
- [54] D. J. Rossi and A. S. Willsky, "Reconstruction from projections based on detection and estimation of objects, Parts I and II: Performance analysis and robustness analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, p. 886, 1984.
- [55] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York: Wiley, 1980.
- [56] M. Slaney, A. C. Kak, and L. E. Larsen, "Limitations of imaging with first-order diffraction tomography," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-32, pp. 860–874, 1984.
- [57] R. Snieder, "The role of the Born approximation in nonlinear inversion," *Inv. Prob.*, vol. 6, pp. 247–266, 1990.
- [58] —, "An extension of Backus–Gilbert theory to nonlinear inverse problems," *Inv. Prob.*, vol. 7, pp. 409–433, 1991.
- [59] —, "The role of nonlinearity in inverse problems," *Inv. Prob.*, vol. 14, pp. 387–404, 1998.
- [60] M. Soumekh, "An improvement to the Rytov approximation in diffraction tomography," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. UFFC-33, no. 4, p. 394, 1986.
- [61] M. Soumekh and M. Kaveh, "A theoretical study of model approximation errors in diffraction tomography," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. UFFC-33, no. 1, p. 10, 1986.
- [62] N. Sponheim, L.-J. Gelius, I. Johansen, and J. J. Stamnes, "Quantitative results in ultrasonic tomography of large objects using line sources and curved detector arrays," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 38, p. 370, 1991.
- [63] N. Sponheim, I. Johansen, and A. J. Devaney, "Initial testing of a clinical ultrasound mammograph," in *Acoustical Imaging*, H. Lee and G. Wade, Eds. New York: Plenum, 1990, vol. 18.
- [64] V. I. Tatarski, *Wave Propagation in a Turbulent Medium*. New York: McGraw-Hill, 1961.
- [65] G. A. Tsihrintzis and A. J. Devaney, "Estimation of object location from diffraction tomographic intensity data," *IEEE Trans. Signal Processing*, vol. 39, no. 9, p. 2136, 1991.
- [66] —, "Maximum likelihood estimation of object location in diffraction tomography, Part II: Strongly scattering objects," *IEEE Trans. Signal Processing*, vol. 39, p. 1466, June 1991.
- [67] —, "Application of a maximum likelihood estimator in an experimental study of ultrasonic diffraction tomography," *IEEE Trans. Med. Imag.*, vol. 12, pp. 545–554, 1993.
- [68] —, "Stochastic diffraction tomography: Theory and computer simulation," *Signal Process.*, vol. 39, pp. 49–64, 1993.
- [69] —, "Stochastic geophysical diffraction tomography," *Int. J. Imag. Syst. Technol.*, vol. 5, pp. 239–242, 1994.
- [70] —, "Maximum likelihood techniques in ultrasonic diffraction tomography," in *Medical Imaging Techniques and Applications*, C. T. Leondes, Ed. Newark, NJ: Gordon & Breach, 1998, vol. 6, pp. 43–126.
- [71] —, "Higher-order (nonlinear) diffraction tomography: Inversion of the Rytov series," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1748–1761, Aug. 2000.
- [72] G. A. Tsihrintzis, A. J. Devaney, and E. Heyman, "Estimation of object location from wideband scattering data," *IEEE Trans. Image Processing*, vol. 8, pp. 996–1001, 1999.
- [73] Y. M. Wang and W. C. Chew, "An iterative solution of the two-dimensional electromagnetic inverse scattering problem," *Int. J. Imag. Sci. Syst.*, vol. 1, pp. 100–108, 1989.
- [74] A. B. Weglein, F. A. Casparotto, P. M. Carvalho, and R. H. Stolt, "An inverse-scattering series method for attenuating multiples in seismic reflection data," *Geophysics*, vol. 62, pp. 1975–1989, 1997.
- [75] A. B. Weglein and K. H. Mason, "Inverse scattering internal multiple attenuation: An analytic example and subevent interpretation," *Proc. SPIE*, pp. 1008–1017, 1998.
- [76] V. H. Weston, "On the convergence of the Rytov approximation for the reduced wave equation," *J. Math. Phys.*, vol. 26, p. 1979, 1985.
- [77] A. Witten, D. Gillette, W. C. King, and J. Sypniewski, "Geophysical diffraction tomography at a dinosaur site," *Geophys.*, vol. 57, pp. 187–195, 1992.
- [78] A. Witten and W. C. King, "Acoustical imaging of subsurface features," *J. Environ. Eng.*, vol. 116, p. 166, 1990.

- [79] A. Witten and E. Long, "Shallow applications of geophysical diffraction tomography," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-24, p. 654, 1986.
- [80] A. Witten, J. Tuggle, and R. C. Waag, "A practical approach to ultrasonic imaging using diffraction tomography," *J. Acoust. Soc. Amer.*, vol. 83, p. 1645, 1988.
- [81] A. J. Witten, J. E. Molyneux, and J. E. Nyquist, "Ground penetrating radar tomography: Algorithms and case studies," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-32, pp. 461-467, 1994.
- [82] E. J. Witterholt, J. L. Kretzschmar, and K. L. Jay, "The application of crosshole electromagnetic wave measurements to mapping of a steam flood," in *Proc. Petroleum Soc. CIM*, 1982.
- [83] E. Wolf, "Three dimensional structure determination of semi transparent objects from holographic data," *Opt. Commun.*, vol. 1, p. 153, 1969.
- [84] —, "Principles and development of diffraction tomography," in *Trends in Optics*, A. Consortini, Ed. New York: Academic, 1996.
- [85] R. J. Wombell and R. D. Murch, "The reconstruction of dielectric objects from scattered field data using the distorted-wave Born approximation," *J. Electromagn. Waves Applicat.*, vol. 7, pp. 687-702, 1993.

George A. Tsihrintzis (S'88-M'93) was born in Athens, Greece. He received the Diploma in electrical engineering (with honors) from the National Technical University of Athens in 1987 and the M.S. and Ph.D. degrees in electrical engineering from Northeastern University, Boston, MA, in 1988 and 1992, respectively.

He is currently an Associate Professor with the Department of Informatics, University of Piraeus, Greece, and a Visiting Professor in the Department of Electrical and Computer Engineering, Northeastern University. Previously, he was a Research Associate with the Signal and Image Processing Institute, University of Southern California, Los Angeles, and Assistant Professor of electrical engineering at the University of Virginia, Charlottesville. His expertise combines signal and image processing and electromagnetic wave theory and his research interests include pattern recognition, detection and estimation theory, statistical signal processing, and inverse problems in scattering with applications in imaging, radar, and communications. He has contributed to more than 100 publications in these areas and is the author of the article on radar applications in the *Wiley Encyclopedia of Electrical and Electronics Engineering* (New York: Wiley, 1999).

Anthony J. Devaney (M'87) received the B.S.E.E. degree in electrical engineering from Northeastern University, Boston, MA, in 1964, the M.S. degree in engineering and applied science from Yale University, New Haven, CT, in 1965, and the Ph.D. degree from the Institute of Optics, University of Rochester, Rochester, NY, in 1971.

He is currently Professor of electrical and computer engineering at Northeastern University, Boston, President of A. J. Devaney Associates, Boston, and Vice President for Technology of Witten Technologies, Washington, DC. Previously, he held positions with Eikonix Corporation, Bedford, MA, and Schlumberger-Doll Research, Ridgefield, CT. He was formerly the Topical Editor for propagation and scattering for the *Journal of the Optical Society of America*. Currently he is a member of the editorial boards of *Wave Motion*, *Electronic Imaging*, and *Inverse Problems* and is an Associate Editor of *Ultrasonic Imaging*. He has published more than 120 papers in refereed journals, holds six patents, and has presented scientific papers at a number of conferences and universities. He has also contributed articles to a number of books and is the author of the entry on acoustical holography in the *Encyclopedia of Applied Physics*, published by the American Institute of Physics.

Dr. Devaney is a Fellow of the Optical Society of America and a Member of the Acoustical Society of America.

A Volterra Series Approach to Nonlinear Traveltime Tomography

George A. Tsihrintzis and Anthony J. Devaney

Abstract—Nonlinear tomographic reconstruction algorithms are developed for inversion of traveltime measurements in scattering experiments in which data models are derived from an arbitrarily large (possibly infinite) number of terms in the perturbation solution to the ray or eikonal equations. The algorithms attain the form of a Volterra series of nonlinear operators, with the usual linear reconstruction algorithm of Traveltime Tomography as the leading term. A computer simulation study is included to illustrate the performance of the algorithms for the case of scattering objects with cylindrical symmetry.

Index Terms—Eikonal equation, inverse scattering, ray tracing, traveltime tomography, Volterra series.

I. INTRODUCTION

SIGNIFICANT research activity has taken place over the past 25 years on the problem of quantitative determination of the structure of an unknown object by computerized processing of measurements of the waves diffracted by the object in a set of scattering experiments [47], [13], [15], [43] [9], [48], [40]. The structure to be reconstructed is usually the spatial distribution of the (complex-valued) index of refraction inside the object. This activity falls, therefore, within the regime of inverse scattering [30], [9], [8], [11] and is applicable in a number of seemingly different scientific disciplines, such as crystal structure determination [26], medical ultrasound tomography [14], [45], [46], [43], [44], acoustic and electromagnetic underground surveying [28], optical and coherent x-ray microscopy [16], and elastic wave inverse scattering [21], [24]. It also falls within the regime of computerized tomography [1], in that it utilizes computerized processing of data collected from several probing directions to form an image of the distribution of the object's refraction index, and, in fact, x-ray computerized tomography is the special case of the general discipline of inverse scattering obtained as the wavelength of the probing radiation approaches zero.

The essence of tomographic imaging with diffracting waves is an attempt to invert the (nonlinear) mathematical mapping that relates the probing wave, the object's refraction index, and the measurable total wave. However, such a task is nontrivial to achieve and, to date, research has only produced theoretical

results and computer-intensive iterative numerical inversion algorithms. An alternative approach that has often met with success was based on approximations to the exact wave model that arose either from perturbation expansions or from application of the geometric (ray) theory of diffraction. The former led to the scientific regime known as (linearized) diffraction tomography [47], [13], [15], [48], [40], [41], while the latter led to the regime of traveltime tomography [36], [2], [29], [31], [3], [4]. Given the domain of validity of the underlying approximations (Born or Rytov), diffraction tomography is applicable when the wavelength of the probing radiation is comparable with the typical dimension of the object. On the other hand, the geometric theory of diffraction is valid in the limit of vanishing wavelength, and traveltime tomography is applicable when the wavelength of the probing radiation is several orders of magnitude smaller than the typical dimension of the object. Consequently, while diffraction tomography provides appropriate reconstruction algorithms when probing with low frequency waves, traveltime tomography is valid in the limit of high frequency probing.

A common assumption in the early developments of traveltime tomography was that the wave propagated along straight line rays through the object [42], [43]. Although this assumption is approximately valid in the limit of very weakly scattering objects (e.g., soft biological tissue), it breaks down in many practical cases. Therefore, ray tracing through the object becomes required. The ray traces can be determined by application of Fermat's principle and minimization of the wave path [1], [5], [9], but in imaging applications, the refraction index distribution inside the object is unknown. Therefore, exact inversion of traveltime measurements is a nonlinear problem. Iterative algorithms have been proposed that, in each iteration, invert the traveltime measurements on the basis of the current ray traces and use the refraction index distribution thus obtained to compute the ray traces for the next iteration [36], [6], [27], [7], [1], [3], [4]. The straight line ray propagation model is an inaccurate model, especially for strong scatterers, but the corresponding algorithms are similar to the algorithms for x-ray tomography [21], [24] and are fast to implement. On the other hand, the iterative algorithms based on successive ray tracing and reconstruction are valid over a much broader set of scatterers, at the cost, however, of significant computational load.

In the present paper, we derive nonlinear reconstruction algorithms for inversion of traveltime measurements that rely on perturbations of the straight line ray wave propagation model [22], [2], [29], [31]. The algorithms present a compromise in both validity and computational load, between the existing linear and iterative nonlinear algorithms and could, in fact, serve as either the final estimates or providers of good preliminary estimates

Manuscript received September 9, 1999; revised April 21, 2000. This work was supported in part by AFOSR Grant F49620-96-1-0039 and AFOSR Grant F49620-96-1-0028.

G. A. Tsihrintzis is with the Department of Informatics, University of Piraeus, Piraeus, Greece.

A. J. Devaney is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: devaney@ece.neu.edu).

Publisher Item Identifier S 0196-2892(00)05888-5.

Clearly, the term T_1 is a linear functional of the refraction index perturbation δn . On the other hand, the term T_2 is a homogeneous quadratic functional of δn , as can be seen from (2.17), and in general, T_j , $j = 1, 2, 3, \dots$ is a homogeneous functional of order j in δn .

D. Perturbation Series Solution to the Eikonal Equation

Now we return to the eikonal equation (2.1) and derive a perturbation series solution to it. Equation (2.1) is a first-order partial differential equation. Therefore, its general solution depends on one integration constant, the value of which is specified by the proper boundary conditions. Since the object is assumed to lie completely beyond the line $x = \tau u - s_0 v$, $\tau \in R^1$, s_0 , fixed, the proper boundary condition is

$$W(x = \tau u - s_0 v) = \langle v, x \rangle = -s_0 \quad (2.24)$$

simply stating that the incident wave has not yet interacted with the object and therefore, the eikonal W on the line $x = \tau u - s_0 v$ equals the eikonal W_0 in the object-free case. This boundary condition is compatible with the ray boundary conditions in (2.4) and (2.5). Indeed, from (2.2) and (2.24), we see that $n(dl/ds)|_{s=-s_0} = (\partial W/\partial l/\partial W/\partial s)|_{s=-s_0} = 0$, as in (2.5).

In the absence of the object ($f \equiv 0$), the eikonal is a function $W = W_0$, with $W_0(x) = \langle v, x \rangle$ at any point x . In the presence of the object f , we will express the eikonal as $W = W_0 + W_f$, i.e., as the sum of the object-free term W_0 and a correction term W_f that is solely due to the presence of the object and, therefore, contains all the information about the object function f . This decomposition is similar to the decomposition of the total (measurable) wave into the sum of the background and the scattered wave in diffraction tomography, the latter of which is solely due to the presence of the object and contains all information about it.

Under conditions, the solution W to the eikonal equation (2.1) admits a series representation of the form of an iterated perturbational expansion. To begin, f is replaced by ϵf in (2.1) and its solution $W(x; \epsilon)$ is obtained by substituting in (2.1) the formal series

$$W(x; \epsilon) = W_0(x) + \epsilon W_1(x) + \epsilon^2 W_2(x) + \dots \quad (2.25)$$

where W_0 is the eikonal in the absence of the object (i.e., for $\epsilon = 0$). The resulting equations that specify the individual terms in the series (2.25) are obtained by equating terms of the same power in ϵ in both sides of (2.1). The result is

$$\langle \nabla W_0, \nabla W_0 \rangle = 1 \quad (2.26)$$

$$\langle \nabla W_0, \nabla W_1 \rangle = f \quad (2.27)$$

$$\langle \nabla W_0, \nabla W_n \rangle = -\frac{1}{2} \sum_{k=1}^{n-1} \langle \nabla W_k, \nabla W_{n-k} \rangle; \quad n \geq 2. \quad (2.28)$$

Equations (2.26), (2.27), and (2.28) are each first-order partial differential equations. Therefore, their general solution will depend on one integration constant which will be specified from the proper boundary conditions. The solution to (2.26) is the complex phase of the incident field, normalized by $1/ik$ (i.e., $W_0(x) = \langle v, x \rangle = s$), which satisfies the boundary condition

(2.24). Therefore, the boundary condition to be satisfied by W_j , $j \geq 1$, requires that

$$W_j(\tau u - s_0 v) = 0; \quad \tau \in R^1, \quad j \geq 1. \quad (2.29)$$

The solution to (2.27) and (2.28) is given next.

Clearly, (2.27) is converted to

$$\langle v, \nabla W_1 \rangle = f \quad (2.30)$$

or, equivalently

$$\frac{\partial W_1}{\partial s} = f. \quad (2.31)$$

Thus

$$W_1(x = \tau u + s v) = \int_{-s_0}^s f(\tau u + s' v) ds' + c_1 \quad (2.32)$$

where from the boundary condition (2.29) $c_1 = 0$. Similarly,

$$W_j(x = \tau u + s v) = -\frac{1}{2} \int_{-s_0}^s \left[\sum_{k=1}^{j-1} \langle \nabla W_k, \nabla W_{j-k} \rangle \right] (\tau u + s' v) ds'; \quad j \geq 2. \quad (2.33)$$

E. Volterra Series Model for Traveltime Data from the Eikonal Equation

It can be shown [5] that $c_0 T(x_0 = \tau u + s_0 v) = \Re\{W(\tau u + s_0 v) - W(\tau u - s_0 v)\}$, where W is the eikonal of (2.1) and \Re denotes the real part. Therefore, similarly to (2.20), we have

$$T(x_0) = T_0(x_0) + \epsilon T_1(x_0) + \epsilon^2 T_2(x_0) + \dots \quad (2.34)$$

where

$$T_0(x_0) = \frac{2s_0}{c_0} \quad (2.35)$$

$$T_1(x_0) = \frac{1}{c_0} W_1(x_0) \quad (2.36)$$

$$T_2(x_0) = \frac{1}{c_0} W_2(x_0), \quad (2.37)$$

$$\vdots$$

F. Relation between Volterra Series Models for Traveltime Data

Equation (2.34) provides a Volterra series expansion to the traveltime time T measured at x_0 , similar to the power series of (2.20). Comparing the procedure that resulted in (2.20) and (2.34), we observe that while T_j in (2.20) is a homogeneous Volterra functional of order j in the refraction index perturbation δn , T_j in (2.34) is a homogeneous Volterra functional of order j in the object function f . However, for $f = \frac{1}{2}(n^2 - 1)$, with $n = 1 + \delta n$, we find $f = \delta n + ((\delta n)^2/2)$. Therefore, $\epsilon f = \epsilon \delta n + \mathcal{O}(\epsilon^2)$, which when combined with (2.22), (2.32), and (2.36), leads to the conclusion that the linear ray perturbation and the linear eikonal perturbation models result in identical traveltime predictions.

III. NONLINEAR TOMOGRAPHIC INVERSION OF TRAVELTIME DATA

A. The General Case

Define the Hilbert spaces \mathcal{F} and \mathcal{P} as follows. \mathcal{F} is $L^2(\mathcal{D}_a)$, where \mathcal{D}_a is a disk of radius a in R^2 , which includes the object

support \mathcal{V} , and \mathcal{P} is the space of functions defined on $R^1 \times [0, 2\pi)$, which are square-integrable with respect to the first variable. Define the nonlinear operators

$$T_j: \mathcal{F} \rightarrow \mathcal{P} \quad (3.1)$$

where

$$(T_j f)(\tau, \theta) = T_j(s_0; \theta, \tau) \quad (3.2)$$

and $T_j(s_0; \theta, \tau)$ is the j th term of the series (2.20) or (2.34) to the traveltime perturbation along the measurement line $s = s_0$ parameterized by the probing wave angle θ and the parameter τ . Clearly, T_j is a homogeneous Volterra operator of order j . More specifically, the operator $T_1: \mathcal{F} \rightarrow \mathcal{P}$ is linear and implementable via use of a fast Fourier transform-based algorithm. Indeed

$$(T_1 f)(\tau, \theta) = \frac{1}{c_0} \int_{-\infty}^{\infty} f(\tau u + s'v) ds'. \quad (3.3)$$

Therefore, the one-dimensional (1-D) Fourier transform $(\widetilde{T_1 f})(p, \theta) \equiv \int_{-\infty}^{\infty} e^{-ip\tau} T_1(s_0; \theta, \tau) d\tau$ is given as

$$\begin{aligned} (\widetilde{T_1 f})(p, \theta) &= \frac{1}{c_0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-ip\tau} f(\tau u + s'v) d\tau ds' \\ &= \tilde{f}(pu) \end{aligned} \quad (3.4)$$

where $\tilde{f}(K) = \int e^{-i(K,x)} f(x) dx$ is the 2-D Fourier transform of f . Equation (3.4) is a statement of the well-known *projection-slice theorem* of x-ray computerized tomography and relates the 1-D Fourier transform of a tomographic projection to a linear slice through the 2-D Fourier transform of the object function [21], [24].

Define now the linear operator

$$\mathcal{B}: \mathcal{P} \rightarrow \mathcal{F} \quad (3.5)$$

with

$$\begin{aligned} (\mathcal{B}d)(x = tu + sv) &= \frac{c_0}{(2\pi)^2} \int_0^\pi d\theta \int_{-\infty}^{\infty} e^{ipt} |p| \tilde{d}(p, \theta) dp, \\ &\quad -s_0 \leq s \leq s_0 \end{aligned} \quad (3.6)$$

where the tilde denotes a Fourier transform, i.e., $\tilde{d}(p, \theta) = \int_{-\infty}^{\infty} e^{-ipt} d(t, \theta) dt$. The operator $(\mathcal{B}|\mathcal{P})$ defines the *filtered backprojection algorithm* of x-ray computerized tomography and is thus the inverse of $(T_1|\mathcal{F})$ [21], [24]. Indeed, if $d(\tau, \theta) = \int_{-\infty}^{\infty} f(\tau u + s'v) ds'$, then the projection-slice theorem implies that $\tilde{d}(p, \theta) = \tilde{f}(pu)$, and algebraic manipulation of (3.6) shows that $(\mathcal{B}d)(x) = f(x)$ [21], [24].

Define the inhomogeneous Volterra operator of order N

$$\mathcal{T}^N = T_1 + T_2 + \cdots + T_N; \mathcal{F} \rightarrow \mathcal{P} \quad (3.7)$$

where

$$(\mathcal{T}^N f)(\tau, \theta) = (T_1 + T_2 + \cdots + T_N)f(s_0; \theta, \tau) \quad (3.8)$$

and $T_j(s_0; \theta, \tau)$, $j = 1, 2, \dots$, is the j th term of the series (2.20) or (2.34) to the traveltime along the measurement line $s = s_0$ parameterized by the probing angle θ and the parameter τ . The existence of a local inverse of the operator \mathcal{T}^N and its representation by a convergent series is guaranteed by the existence of the inverse $(\mathcal{B}|\mathcal{P})$ of the linear term $(T_1|\mathcal{F})$ [20], [35]. Specifically, there exists a region in the data space \mathcal{P} on which the operator series

$$B_1 + B_2 + \cdots + B_j + \cdots \quad (3.9)$$

is the inverse of \mathcal{T}^N , where B_j is a homogeneous Volterra operator of order j in the data. Conditions for convergence of inverse Volterra series of the form of (3.9) are stated in the relevant literature (e.g., [20], [25]). However, the application of the general conditions to the specific Volterra series examined here is not only nontrivial, but also nonpractical. Indeed, the conditions require knowledge of the object function, the unknown that the traveltime measurements seek to determine. A better and more practical approach would be to compute successive terms of the proposed series until convergence (or divergence) is observed. The general Volterra series theory guarantees that when the inverse series converges, its limit is the true inverse.

More specifically, the first few terms in the series (3.9) are

$$B_1 = \mathcal{B} \quad (3.10)$$

$$B_2 = -\mathcal{B}T_2\mathcal{B} \quad (3.11)$$

$$B_3 = \mathcal{B}T_2\mathcal{B}T_2\mathcal{B} - \mathcal{B}T_3\mathcal{B} \quad (3.12)$$

\vdots

It is observed that the terms B_1, B_2, B_3 , etc. are *universal*, in the sense that they do not depend on the order N of the forward model \mathcal{T}^N . Thus, the operator series in (3.9) can be truncated to desired order j and return inversion algorithms of order j in the data. For example, the usual filtered backprojection algorithm provides a universal inversion algorithm that is first-order (linear) in the data. A second observation is that if the data are sufficiently modeled by the first-order term (i.e., $T = T_1$ and $T_2 = T_3 = \cdots = 0$), then each higher-order term in series (3.9) returns an identically zero contribution to image reconstruction, as close examination of the expressions in (3.10)–(3.12) reveals. This fact is also verified in the simulations of the next section. A final observation is that the resulting algorithms are effectively of the form of *nonlinear* data filtering followed by backprojection. Indeed, the series (3.9) can be written as

$$\mathcal{B}(\mathcal{I} - T_2\mathcal{B} + T_2\mathcal{B}T_2\mathcal{B} - T_3\mathcal{B} + \cdots) \quad (3.13)$$

where $\mathcal{I}: \mathcal{P} \rightarrow \mathcal{P}$ is the identity operator. Effectively, the term in brackets in (3.13) is a higher-order nonlinear Volterra filter applied on the data and thus, the series (3.9) gives rise to an entire class of nonlinear filtered backprojection algorithms for inversion of traveltime data.

B. Cylindrically Symmetric Case

A particularly important imaging situation is that of objects with cylindrical symmetry or scatterers whose refraction index (object function) varies only with the radial distance from their center and thus is of the form

$$f(x) = f_r(r), \quad r = |x|. \quad (3.14)$$

This important situation arises in practical applications such as reconstruction of the structure of optical fibers or large molecules but also provides a paradigm for both analytical and computer simulation study of reconstruction algorithms. Both the filtered backprojection inversion algorithm and the nonlinear forward models attain a simpler form.

Indeed, let

$$\tilde{f}_r(|K|) = 2\pi \int_0^\infty r J_0(|K|r) f_r(r) dr. \quad (3.15)$$

Then, the Fourier transform of the object function f becomes

$$\hat{f}(K) = \hat{f}_r(|K|) \quad (3.16)$$

and is inverted as

$$f_r(r) = \frac{1}{2\pi} \int_0^\infty |K| J_0(|K|r) \hat{f}_r(|K|) d|K|. \quad (3.17)$$

The projection-slice theorem reads

$$(\widehat{T_1 f})(p) = \frac{1}{c_0} \hat{f}_r(|p|) \quad (3.18)$$

where the projections $T_1 f$ are independent of the view angle θ . The filtered backprojection algorithm is now put in the form

$$\hat{f}(x = tu) = \frac{c_0}{2\pi} \int_0^\infty J_0(p|t|) p \hat{d}(p) dp \quad (3.19)$$

where the integration with respect to view angle θ has been eliminated.

Similarly, (2.33) gives

$$(\mathcal{T}_2 f)(\tau, \theta) = -\frac{1}{2c_0} \int_{-s_0}^{s_0} \left[\left(\frac{\partial W_1}{\partial t} \right)^2 + \left(\frac{\partial W_1}{\partial s} \right)^2 \right] (\tau u + sv) ds \quad (3.20)$$

where

$$\frac{\partial W_1}{\partial s}(x) = f_r(|x|) \quad (3.21)$$

$$\frac{\partial W_1}{\partial t}(x = \tau u + sv) = \tau \int_{-s_0}^s \frac{f'_r(\sqrt{\tau^2 + s'^2})}{\sqrt{\tau^2 + s'^2}} ds'. \quad (3.22)$$

IV. COMPUTER SIMULATION

The proposed nonlinear tomographic reconstruction algorithms were implemented and studied in a limited computer simulation. The scatterer was cylindrically symmetric with a Gaussian object function

$$f(x) = f_r(|x|) \equiv A \exp\left(-\frac{|x|^2}{2\sigma^2}\right) \quad (4.1)$$

where we chose $\sigma = 0.5$ and considered the two cases $A = 0.002$ and $A = 0.02$. The case $A = 0.002$ represents a (perhaps unrealistically) weak scatterer for which traveltimes measurements are sufficiently modeled by only the first traveltimes term. On the other hand, the case $A = 0.02$ represents more common scatterers, such as those encountered in ultrasonic imaging of soft tissue [38], [37], [39]. Effectively, the object is contained within the disk $|x| \leq 1.5$, i.e., has a diameter of approximately 3. The measurement distance was set to $s_0 = 25.5$.

For this object, the eikonal terms W_n and the corresponding traveltimes terms T_n can be computed in closed form. Indeed, we find that (3.18) and (3.20) give the first two traveltimes terms as

$$T_1(\tau) = \sqrt{\frac{\pi}{2}} \frac{\sigma}{c_0} \exp\left(-\frac{\tau^2}{2\sigma^2}\right) \cdot \left[\operatorname{erf}\left(\frac{s_0}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(-\frac{s_0}{\sigma\sqrt{2}}\right) \right] \quad (4.2)$$

$$T_2(\tau) = -\frac{1}{2c_0} \int_{-s_0}^{s_0} \left[\left(\frac{\partial W_1}{\partial t} \right)^2 + \left(\frac{\partial W_1}{\partial s} \right)^2 \right] (\tau u + sv) ds \quad (4.3)$$

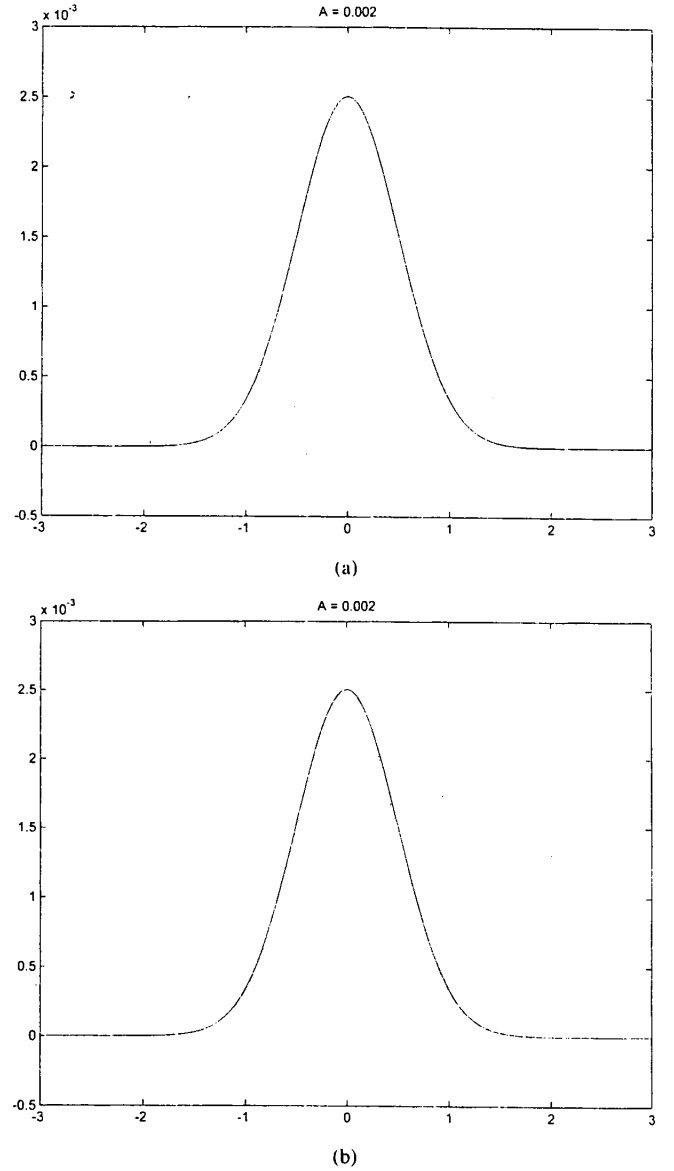


Fig. 2. (a) First (second) traveltimes term for the weak scatterer ($A = 0.002$) is plotted in solid (dashed) line. (b) The first-(second) order traveltimes approximation for the weak scatterer ($A = 0.002$) is plotted in solid (dashed) line.

where from (3.21) and (3.22), we have

$$\begin{aligned} \frac{\partial W_1}{\partial s}(x) &= A \exp\left(-\frac{|x|^2}{2\sigma^2}\right) \\ \frac{\partial W_1}{\partial t}(x = \tau u + sv) &= -\sqrt{\frac{\pi}{2}} \tau \exp\left(-\frac{\tau^2}{2\sigma^2}\right) \\ &\quad \cdot \left[\operatorname{erf}\left(\frac{s}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(-\frac{s}{\sigma\sqrt{2}}\right) \right]. \end{aligned} \quad (4.4)$$

In Fig. 2(a), we show plots of the traveltimes terms T_1 (solid line) and T_2 (dashed line), as computed from (4.2) and (4.3) for $c_0 = 1$ and for the weak scatterer with $A = 0.002$. Fig. 2(b) shows plots of the first-order (solid line) and second-order (dashed line) traveltimes approximations $T = T_1$ and $T = T_1 + T_2$, respectively. We observe that the second

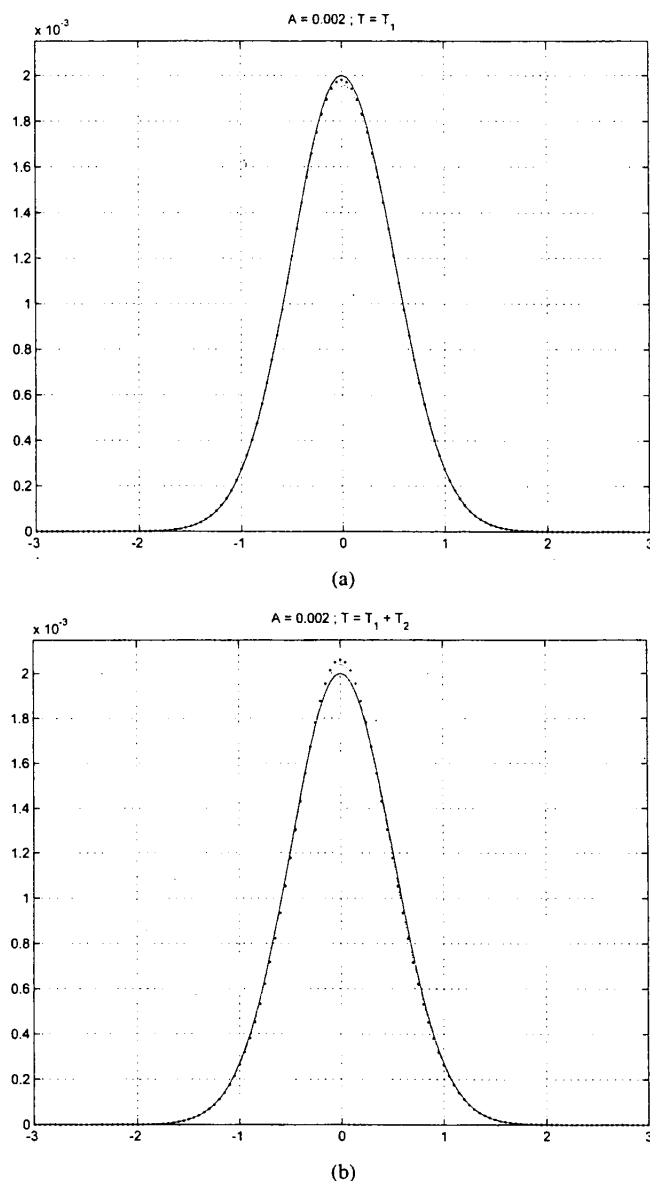


Fig. 3. (a) Original object function (solid line) and reconstructions from the first-order traveltime data returned by the filtered backprojection (dotted line) and the second-order algorithm (dashed line) in the computer simulation of the weak scatterer ($A = 0.002$). (b) Original object function (solid line) and reconstructions from the second-order traveltime data returned by the filtered backprojection (dotted line) and the second-order algorithm (dashed line) in the computer simulation of the weak scatterer ($A = 0.002$).

traveltime term is more than 40 times smaller in magnitude than the first traveltime term ($(\max |T_2| / \max |T_1|) \approx 0.0233$). Thus, the traveltime T is well (but not quite perfectly) approximated by the first term T_1 . Fig. 3 shows the object function reconstructions returned by the filtered backprojection and the second-order algorithm based on the traveltime series from first-[Fig. 3(a)] and second-order [Fig. 3(b)] traveltime approximations. Due to the weakness of the scatterer, the filtered backprojection algorithm returns an accurate reconstruction that is almost identical to the original object function even for data $T = T_1 + T_2$. The second-order reconstruction algorithm returns essentially the same reconstruction as the filtered backprojection algorithm for data $T = T_1$ and slightly improved for data $T = T_1 + T_2$.

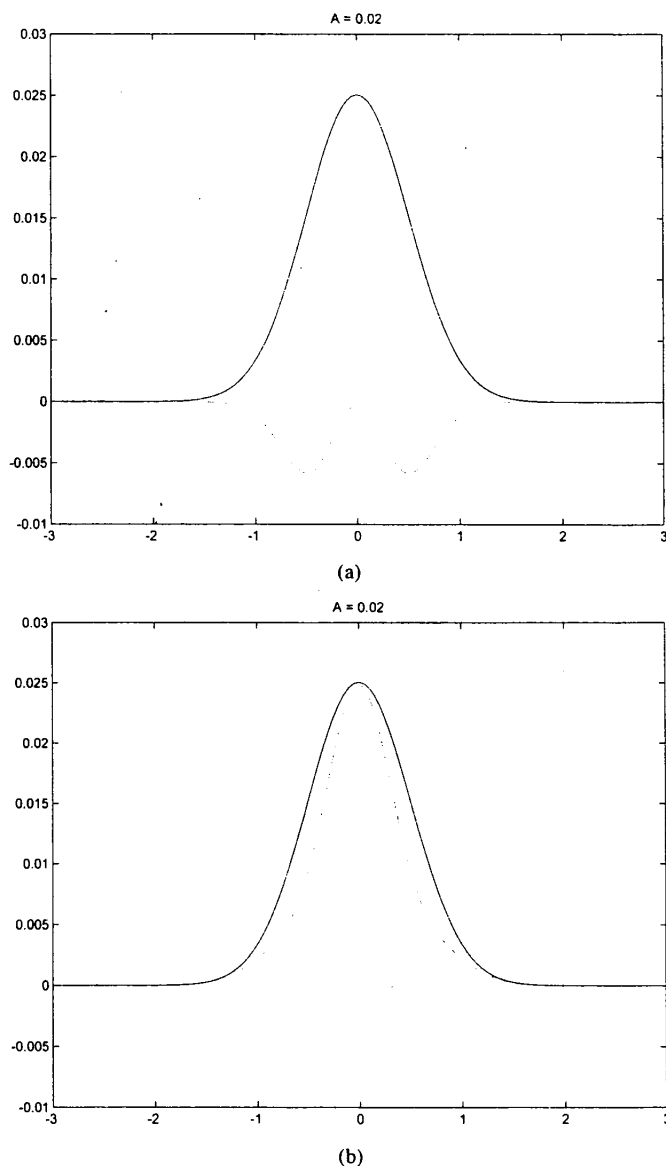


Fig. 4. (a) First (second) traveltime term for the stronger scatterer ($A = 0.02$) is plotted in solid (dashed) line. (b): The first-(second) order traveltime approximation for the stronger scatterer ($A = 0.02$) is plotted in solid (dashed) line.

In Fig. 4(a), we show plots of the traveltime terms T_1 (solid line) and T_2 (dashed line), as computed from (4.2) and (4.3) for $c_0 = 1$ and for the scatterer with $A = 0.02$. Fig. 4(b) shows plots of the first-order (solid line) and second-order (dashed line) traveltime approximations $T = T_1$ and $T = T_1 + T_2$, respectively. We observe that the second traveltime term is more than four times smaller in magnitude than the first traveltime term ($(\max |T_2| / \max |T_1|) \approx 0.2330$). Additionally, the third traveltime term T_3 (not shown) is about 16 times smaller in magnitude than the first traveltime term T_1 . Thus, the traveltime T is now well (but not perfectly) approximated by the sum of the first two terms T_1 and T_2 . Fig. 5 shows the object function reconstructions returned by the filtered backprojection and the second-order algorithm based on the traveltime series, from first- [Fig. 5(a)] and second-order [Fig. 5(b)] traveltime approximations. Clearly, the second-order algorithm returns the same reconstruction as the filtered backprojection algorithm in the

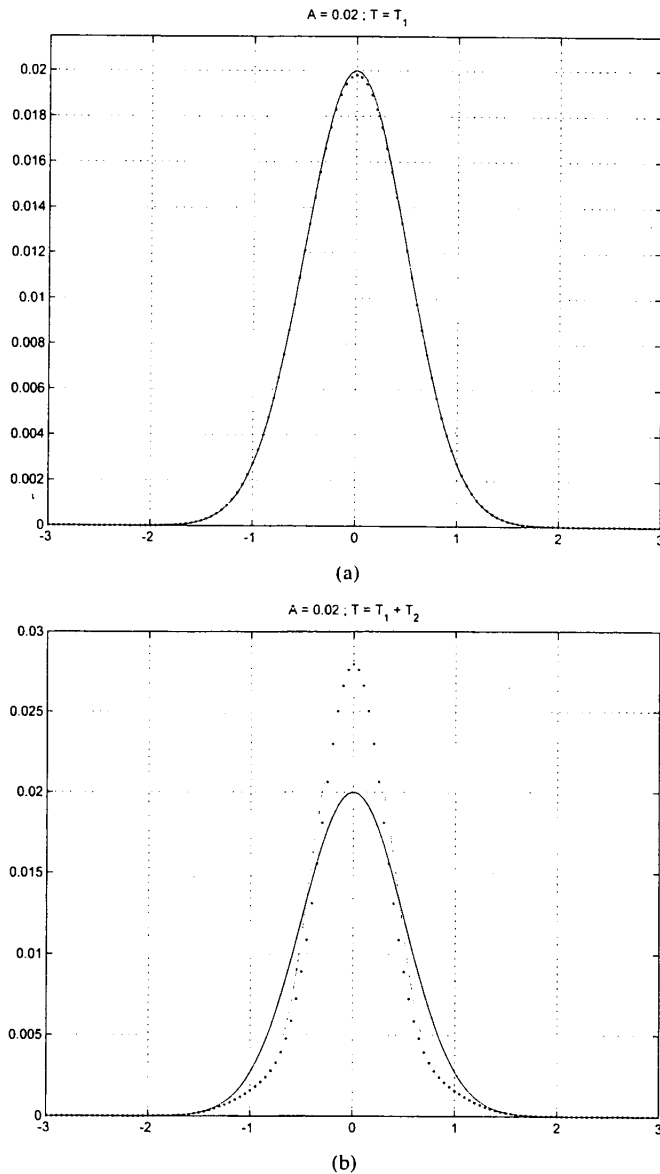


Fig. 5. (a) Original object function (solid line) and reconstructions from the first-order traveltime data returned by the filtered backprojection (dotted line) and the second-order algorithm (dashed line) in the computer simulation of the stronger scatterer ($A = 0.02$). (b) Original object function (solid line) and reconstructions from the second-order traveltime data returned by the filtered backprojection (dotted line) and the second-order algorithm (dashed line) in the computer simulation of the stronger scatterer ($A = 0.02$).

case of data consisting of only the first traveltime term. However, the filtered backprojection algorithm underperforms the second-order algorithm by a significant margin when the second traveltime term is not negligible. We also observe in Fig. 5(b) that the inverted profiles still differ from the true profile even when a second-order inversion algorithm is used. This implies that the third and perhaps higher terms in the inverse series (3.9) are significant.

Further testing of the proposed nonlinear inversion algorithms is currently under way. These tests involve processing traveltime measurements collected using a pulsed laser and scatterers whose object functions are well-known. These tests will be completed soon and the results will be announced at that point.

V. SUMMARY, CONCLUSIONS, AND FUTURE RESEARCH

In this paper, linearized traveltime tomography was extended into a higher-order (nonlinear) regime via use of an arbitrarily large number of terms from the ray or eikonal series as the data-generating model in scattering experiments. Nonlinear tomographic reconstruction algorithms were developed for inversion of wavefield data and image formation, which attained the form of a series of Volterra operators in which the leading (linear) term was the usual filtered backprojection algorithm of linearized traveltime tomography. A limited computer simulation study was included to illustrate the performance of the algorithms in the case of scattering objects with cylindrical symmetry. From this study, it was found that a second-order algorithm returned the same reconstruction as the usual filtered backprojection algorithm when applied to first-order traveltime data, but clearly outperformed the filtered backprojection algorithm when applied to data consisting of the second-order traveltime approximation.

This paper establishes a unified paradigm for development of linear and nonlinear reconstruction algorithms for traveltime tomography and thus further investigation in its direction is due. Future relevant research may follow the avenues of higher-order traveltime tomography with alternative data measurement configurations (e.g., spherical incident waves) and higher-order traveltime tomography of stochastic objects. These and other research avenues are currently being explored, and the findings will be reported elsewhere.

APPENDIX ELEMENTS OF THE VOLTERRA THEORY OF NONLINEAR SYSTEMS

A homogeneous Volterra operator (VO) of order $k \geq 1$ is a nonlinear system represented by a functional relation of the form

$$y(\cdot) = \int \cdots \int d\tau_1 \cdots d\tau_k h_k(\cdot; \tau_1, \cdots, \tau_k) x(\tau_1) \cdots x(\tau_k) \quad (\text{A.1})$$

between its input $x(\cdot)$ and the corresponding output $y(\cdot)$. The $(k+1)$ -dimensional function $h_k(\cdot; \tau_1, \cdots, \tau_k)$ is the *kernel* of the homogeneous k th-order VO, and for uniqueness, it is assumed symmetric (i.e., its value does not change under a permutation of its arguments). Clearly, a homogeneous VO of order $k = 1$ is a linear system and its kernel $h_1(\cdot; \tau_1)$ is also its impulse response. A homogeneous VO of order k as in (A.1) is a homogeneous polynomial system of order k in the sense that the system response to input $cx(\cdot)$ is $c^k y(\cdot)$ for arbitrary constant c .

A general (inhomogeneous) VO of order N is a nonlinear system consisting of the summation of homogeneous VO's of the form of (A.1) of order less than or equal to N , i.e., a nonlinear system represented by a functional relation of the form

$$y(\cdot) = h_0 + \sum_{k=1}^N \int \cdots \int d\tau_1 \cdots d\tau_k h_k(\cdot; \tau_1, \cdots, \tau_k) x(\tau_1) \cdots x(\tau_k) \quad (\text{A.2})$$

where h_0 is a constant. Clearly, an inhomogeneous VO of order N is a polynomial system of order N in the sense that its response to input $cx(\cdot)$ is a polynomial in c of order N for arbitrary constant c . (A.2) can be generalized further by letting the order N of the inhomogeneous VO become infinite, assuming convergence of the resulting functional series

$$y(\cdot) = h_0 + \sum_{k=1}^{\infty} \int \cdots \int d\tau_1 \cdots d\tau_k h_k(\cdot; \tau_1, \cdots, \tau_k) x(\tau_1) \cdots x(\tau_k). \quad (\text{A.3})$$

In practice, both the number of terms in (A.3), as well as the support of the Volterra kernels must be truncated. Issues of approximation of nonlinear systems with infinite and truncated Volterra systems, as well as issues of existence and uniqueness of inverses of Volterra systems, have been extensively addressed in the literature (e.g., [10], [20], [25], [18], [35], [33], [32], [17], [34]), to which the interested reader is pointed.

REFERENCES

- [1] A. H. Andersen and A. C. Kak, "Digital ray tracking in two-dimensional refractive fields," *J. Acoust. Soc. Amer.*, vol. 72, pp. 1593–1606, 1982.
- [2] R. H. T. Bates and G. C. McKinnon, "Toward improving images in ultrasonic transmission tomography," *Australasian Phys. Sci. Med.*, vol. 2–3, pp. 134–140, 1979.
- [3] J. G. Berryman, "Fermat's principle and nonlinear traveltime tomography," *Phys. Rev. Lett.*, vol. 62, pp. 2953–2956, 1989.
- [4] —, "Stable iterative reconstruction algorithm for nonlinear traveltime tomography," *Inv. Probl.*, vol. 6, pp. 21–42, 1990.
- [5] M. Born and E. Wolf, *Principles of Optics*, 6th ed. New York: Pergamon, 1983.
- [6] S. Cha and C. M. Vest, "Interferometry and reconstruction of strongly refracting asymmetric-refractive-index fields," *Opt. Lett.*, vol. 4, pp. 311–313, 1979.
- [7] —, "Tomographic reconstruction of strongly refracting fields and its application to interferometric measurement of boundary layers," *Appl. Opt.*, vol. 20, pp. 2787–2794, 1981.
- [8] K. Chadon, D. Colton, L. Paivarinta, and W. Rundell, *An Introduction to Inverse Scattering and Inverse Spectral Problems*. Philadelphia: SIAM, 1997.
- [9] W. C. Chew, *Waves and Fields in Inhomogeneous Media*. New York: Van Nostrand Reinhold, 1990.
- [10] G. S. Christensen, "On the convergence of Volterra series," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 736–737, 1968.
- [11] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering*, 2nd ed. Berlin, Germany: Springer-Verlag, 1998.
- [12] A. J. Devaney, "A computer simulation study of diffraction tomography," *IEEE Trans. Biomed. Eng.*, vol. BME-30, p. 377, 1982.
- [13] —, "A filtered backpropagation algorithm for diffraction tomography," *Ultrason. Imag.*, vol. 4, p. 336, 1982.
- [14] —, "Geophysical diffraction tomography," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-22, p. 3, Jan. 1984.
- [15] —, "Reconstructive tomography with diffracting wavefields," *Inv. Probl.*, vol. 2, p. 161, 1986.
- [16] —, "Elastic wave inverse scattering," in *Elastic Waves and Ultrasonic Nondestructive Evaluation*, S. K. Datta, J. D. Achenbach, and Y. S. Rajapakse, Eds. New York: Elsevier Science, 1990.
- [17] R. J. P. De Figueiredo, "A generalized Fock space framework for nonlinear system and signal analysis," *IEEE Trans. Circuits Syst.*, vol. CAS-30, pp. 637–647, 1983.
- [18] R. J. P. De Figueiredo and T. A. W. Dwyer, "A best approximation framework and implementation for simulation of large-scale nonlinear systems," *IEEE Trans. Circuits Syst.*, vol. CAS-27, pp. 1005–1014, 1980.
- [19] J. F. Greenleaf, "Computerized tomography with ultrasound," *Proc. IEEE*, vol. 71, p. 330, 1983.
- [20] A. Halme and J. Orava, "Generalized polynomial operators for nonlinear system analysis," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 226–228, 1972.
- [21] G. T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*. New York: Academic, 1980.
- [22] M. J. Jacobson, W. L. Siegman, N. L. Weinberg, and J. G. Clark, "Perturbation method for determining acoustic rays in a two-dimensional sound-speed medium," *J. Acoust. Soc. Amer.*, vol. 57, pp. 843–855, 1975.
- [23] A. C. Kak, "Computerized tomography with X-ray emission and ultrasound sources," *Proc. IEEE*, vol. 67, p. 1245, 1979.
- [24] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. New York: IEEE Press, 1988.
- [25] C. Lesiak and A. J. Krener, "The existence and uniqueness of Volterra series for nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 1090–1095, 1978.
- [26] H. Lipson and W. Cochran, *The Determination of Crystal Structures*. Ithaca, NY: Cornell Univ. Press, 1966.
- [27] R. J. Lytle and K. A. Dines, "Iterative ray tracing between boreholes for underground image reconstruction," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-18, pp. 234–240, 1980.
- [28] M. H. Maleki, A. J. Devaney, and A. Schatzberg, "Tomographic reconstruction from optical scattered intensities," in *J. Opt. Soc. Amer. A*, 1992, vol. A, pp. 1356–1363.
- [29] G. C. McKinnon and R. H. T. Bates, "A limitation on ultrasonic transmission tomography," *Ultrason. Imaging*, vol. 2, pp. 48–54, 1980.
- [30] R. G. Newton, *Scattering Theory of Waves and Particles*. Berlin, Germany: Springer Verlag, 1982.
- [31] S. J. Norton and M. L. Linzer, "Correction for ray refraction in velocity and attenuation tomography: A perturbation approach," *Ultrason. Imaging*, vol. 4, pp. 201–233, 1982.
- [32] I. W. Sandberg, "The mathematical foundations of associated expansions for mildly nonlinear systems," *IEEE Trans. Circuits Syst.*, vol. CAS-30, pp. 441–455, 1983.
- [33] —, "Volterra-like expansions for solutions of nonlinear integral and nonlinear differential equations," *IEEE Trans. Circuits Syst.*, vol. CAS-30, pp. 68–77, 1983.
- [34] —, "Uniform approximation with doubly finite volterra series," *IEEE Trans. Signal Processing*, vol. SP, pp. 1438–1442, 1992.
- [35] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York: Wiley, 1980.
- [36] H. Schomberg, "An improved approach to reconstructive ultrasound tomography," *J. Phys. D: Appl. Phys.*, vol. 11, pp. L181–L185, 1978.
- [37] N. Sponheim and I. Johansen, "Experimental results in ultrasonic tomography using a filtered backpropagation algorithm," *Ultrason. Imag.*, vol. 13, p. 56, 1991.
- [38] N. Sponheim, I. Johansen, and A. J. Devaney, "Initial testing of a clinical ultrasound mammograph," in *Acoustical Imaging*, H. Lee and A. J. Devaney, Eds. New York: Plenum Press, 1990, vol. 18.
- [39] J. J. Stamnes, L.-J. Gelius, I. Johansen, and N. Sponheim, "Diffraction tomography applications in seismics and medicine," in *Proc. NATO Advanced Workshop on Inverse Problems in Scattering and Imaging*, North Falmouth, MA, 1991.
- [40] G. A. Tsihrintzis and A. J. Devaney, "Maximum likelihood techniques in ultrasonic diffraction tomography," in *Medical Imaging Techniques and Applications*, C. T. Leondes, Ed. Newark, NJ: Gordon and Breach, 1998, vol. 6, pp. 43–126.
- [41] G. A. Tsihrintzis, A. J. Devaney, and E. Heyman, "Estimation of object location from wideband scattering data," *IEEE Trans. Image Processing*, vol. 8, pp. 996–1001, 1999.
- [42] C. M. Vest, "Interferometry of strongly refracting axisymmetric phase objects," *Appl. Opt.*, vol. 14, pp. 1601–1128, 1975.
- [43] Y. M. Wang and W. C. Chew, "An iterative solution of the two-dimensional electromagnetic inverse scattering problem," *Int. J. Imag. Sci. Syst.*, vol. 1, pp. 100–108, 1989.
- [44] A. Witten and W. C. King, "Acoustical imaging of subsurface features," *J. Environ. Eng.*, vol. 116, p. 166, 1990.
- [45] A. Witten and E. Long, "Shallow applications of geophysical diffraction tomography," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-24, p. 654, 1986.
- [46] A. Witten, J. Tuggle, and R. C. Waag, "A practical approach to ultrasonic imaging using diffraction tomography," *J. Acoust. Soc. Amer.*, vol. 83, p. 1645, 1988.
- [47] E. Wolf, "Three dimensional structure determination of semi transparent objects from holographic data," *Opt. Commun.*, vol. 1, p. 153, 1969.
- [48] —, "Principles and development of diffraction tomography," in *Trends in Optics*, A. Consortini, Ed. San Diego, CA: Academic, 1996.

George Tsihrintzis (S'88–M'93) was born in Athens, Greece. He received the electrical engineer degree from the National Technical University of Athens (Hons.), Athens, Greece, in 1987, and the M.S. and Ph.D. degrees in electrical engineering from Northeastern University, Boston, MA, in 1988 and 1992, respectively.

He is currently an Associate Professor with the Department of Informatics, The University of Piraeus, Piraeus, Greece, and a Visiting Professor with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA. Previously, he was a Research Associate with the Signal and Image Processing Institute, University of Southern California, Los Angeles, and an Assistant Professor of Electrical Engineering, University of Virginia, Charlottesville. His expertise combines signal and image processing and electromagnetic wave theory, and his research interests include pattern recognition, detection and estimation theory, statistical signal processing, and inverse problems in scattering with applications in imaging, radar, and communications. He has contributed over 100 publications in these areas and is the author of the article on *Radar Applications* in the *Wiley Encyclopedia of Electrical and Electronics Engineering* (1999).

Anthony J. Devaney (M'87) received the B.S.E.E. in electrical engineering from Northeastern University, Boston, MA, in 1964, the M.S. degree in engineering and applied science from Yale University, New Haven, CT, in 1965, and the Ph.D. degree from the Institute of Optics, University of Rochester, Rochester, NY, in 1971.

He is currently Professor of Electrical and Computer Engineering with Northeastern University, and is also President of A. J. Devaney Associates, Boston, MA and Vice President of Technology of Witten Technologies, Washington, DC. Previously, he held positions with Eikonix Corporation, Bedford, MA, and Schlumberger–Doll Research, Ridgefield, CT. He has published more than 120 papers in refereed journals, holds six patents, and has presented scientific papers at a number of conferences and universities.

Dr. Devaney is a Fellow of the Optical Society of America and a member of the Acoustical Society of America. He was formerly the topical Editor for propagation and scattering for the *Journal of the Optical Society of America*. Currently, he is a member of the editorial boards of the journals *Wave Motion*, *Electronic Imaging*, and *Inverse Problems*, and is an Associate Editor of *Ultrasonic Imaging*.

Time Domain Processing of Frequency Domain GPR Signatures for Buried Land Mine Detection¹

Audrey J. Dumanian^{a,b}, Carey M. Rappaport^a, and Ann W. Morgenthaler^a

^aCenter for Subsurface Sensing and Imaging Systems
Northeastern University
Boston, MA 02115
Tel: (617) 373-2043
Fax: (617) 373-8627
Email: rappaport@cer.neu.edu

^bMIT Lincoln Laboratory
244 Wood Street
Lexington, MA 02420
Email: audrey@ll.mit.edu

Abstract

This paper investigates the feasibility of detecting plastic antipersonnel land mines buried in lossy, dispersive, rough soils using a stepped-frequency ultra wideband (WB) ground-penetrating radar (GPR). Realistic land mine scenarios were modeled using a two-dimensional (2D) finite difference frequency domain (FDFD) technique. Assuming normal incidence plane wave excitation, the scattered fields were generated over a large frequency bandwidth (.5 to 5 GHz) for a variety of mine-like shapes, different soil types, and multiple receiver locations. The simulation results showed that for a ground penetration sensor located just above the soil surface, the strong reflection signals received from the rough ground surface obscured the buried target's frequency response signal.

The simulated GPR WB frequency response data at each receiver location was transformed to the time domain using the fast fourier transform. Time domain processing permits high resolution measurement of target features that are invariant to the ground roughness and also that are dependent on the soil characteristics as well as the burial depth and size of the mine. Specifically, two or more characteristic timing peaks are observed in the simulation results suggesting that the ultra-wideband spectral radar response may yield particular advantages not exploited by currently employed detection systems. It is also shown that by using time-gating to remove the strong ground reflection signals, the target signals are selectively enhanced (as expected), but more surprisingly, the target frequency response signature is almost completely recovered.

Keywords: mine detection, ground-penetrating radar, ultra-wideband radar, resonance, dielectric target imaging

¹This work is sponsored by the Department of the Air Force under Air Force contract #F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Air Force.

1. Introduction

The detection and removal of buried land mines is a serious problem that affects both civilians and soldiers throughout the world. Metal detectors are the current sensors of choice for detecting mines, but they are relatively ineffective in locating plastic land mines that contain little to no metal. In this case, an improved sensor capability and/or advanced signal processing technology is necessary to effectively detect all types of mines.

This paper investigates the feasibility of using an ultra wideband stepped-frequency ground penetrating radar sensor to detect nonmetallic antipersonnel land mines buried in lossy dispersive soil. GPR is one of the oldest sensor technologies for subsurface sensing. The geophysical community has used it successfully for decades as a cost effective and noninvasive measure to probe the ground in determining the underlying strata. GPR has proven to be a useful sensor when (1) the soil is not exceedingly lossy, (2) the target or layer of interest has a high-dielectric constant contrast with the surrounding or adjacent ground, (3) the target is not excessively deep, and (4) the search area is not extremely large. The limiting factor in using GPR for mine detection within the surface a few feet is not so much attenuation, but clutter. In radar terminology, the GPR problem is clutter-limited not noise limited. The sources of clutter can include vegetation and other non-metallic objects such as rocks, tree branches, etc. The rejection of clutter is an essential part of a successful GPR system. In addition, for GPR applications the target of interest is characterized not only by the inherent properties of the buried object and soil but also by the strong effects of the air-ground interface. This is due to the fact that the rough soil surface may also scatter much of the transmitted sensing signal and be a significant source of detection clutter.

In order for a GPR to be successful sensor in mine detection, it must be able to identify the buried target, i.e. determine the mine's shape, size, and material characteristics. This is a necessary requirement in order to separate the reflected mine signals from the clutter signals (rocks, branches, debris, etc.). For most subsurface sensing systems, the GPR relies on the magnitude of the target backscatter as its only classification measure. Detection difficulties arise when you have a low dielectric constant contrast target (small permittivity difference between the mine and soil) that would produce a weak target response, for example, a TNT mine buried in sand. In addition due to the small size of mine, the mine's response would be even weaker at the lower frequencies. Yet it is extremely difficult to separate a buried mine from the ground surface clutter even at low frequencies. At higher frequencies the propagation loss and clutter from both the ground roughness and the soil inhomogeneity will become severe and therefore also increase the detection difficulty. Another frequency dependent response factor to consider is the conductivity of the soil. If the soil has high loss then accordingly the strength of the reflected target signal is attenuated as the frequency increases. Therefore a WB GPR experiences several frequency dependent effects.

The signal processing transformation of ultra-wideband GPR-based frequency domain reflection signals to the time domain will enable highly resolved target features to appear that are invariant to the ground roughness. The power spectral density of the reflected signal yields a time domain signature dependent on the soil characteristics as well as the burial depth and size of the mine. Specifically, two or more characteristic time peaks are uniquely observed suggesting that the ultra-wideband spectral

radar response may yield particular discrimination advantages not exploited by currently employed detection systems.

It is shown that by using time-gating to remove the strong ground reflection signals, the target signals are selectively enhanced (as would be expected), but more surprisingly, the target frequency response signature is almost completely recovered. This study determines the feasibility of detecting mine-like targets based on time domain processed GPR frequency signatures.

2. Frequency Domain Scattering

For this investigation, a 2D FDFD electromagnetic modeling simulation code was used to generate the near zone scattered electric fields resulting from a normally incident TM uniform plane wave on a random rough dielectric half-space containing a buried dielectric mine-like target. The simulated scattered electric field results generated by the FDFD code were validated with the scattered field results computed from another numerical electromagnetic simulation technique, a semi-analytic mode matching (SAMM) algorithm¹⁻². Excellent scattering agreement was achieved between the two codes for dielectric targets buried beneath randomly rough soils³⁻⁴. FDFD is an ideal technique for computing the near-field scattered fields since it permits an accurate representation at each frequency for the random rough surface and various target shapes. The only shortcoming was the extensive computational requirement in both memory and runtime.

Several realistic land mine simulation scenarios were modeled. A variety of modeling parameters were considered such as the mine shape, the surface roughness, and the soil type. The datasets were generated over a densely sampled wide frequency bandwidth (.5 to 5 GHz, 100 MHz increments). Each FDFD data set for the entire frequency bandwidth took approximately 12 hours to run on a Silicon Graphics Octane workstation (R12000; 300 MHz). Since most anti-personnel mines can be described by a simple geometry, three buried targets were considered in this investigation --- a circular-shaped TNT mine (10 cm diameter), a rectangular-shaped TNT mine (10 cm by 4 cm), and an elliptical-shaped TNT mine (major axis: 10 cm and minor axis: 6 cm). All the targets were modeled having the electrical characteristics of TNT ($\epsilon_m = 2.9(1+i0.001)\epsilon_o$), which was held constant over the frequency bandwidth.

Two different ground soil types were considered in this analysis. One is representative of the material characteristics of a low loss, dry sand ($\epsilon_g = 2.55(1+i0.01)\epsilon_o$) that remained constant over the frequency band. Note the permittivity of the sand is very similar to the permittivity of the TNT mine. The second soil model was a dispersive Bosnian soil⁵⁻⁷ ($\epsilon_g = 8.82(1+i0.13)\epsilon_o$ at 3 GHz), which was measured at the test site Alicia. The Bosnian soil had a higher permittivity than the mine and also a higher loss than the dry sand case. Therefore in these two cases, the TNT target buried in sand is considered a low dielectric constant contrast target and in the Bosnian soil the mine is a high dielectric constant contrast target. The simple targets were buried in these soils at a fixed depth of 10 cm from the air-ground surface. The air-ground interface was modeled both as smooth interface and a randomly rough interface. The rough soils had an average roughness equal to 1 cm and a correlation length equal to 4 cm.

Figure 1 schematically shows two FDFD input scattering model geometries representative of an infinite dielectric cylinder buried in the ground beneath a smooth and rough air-ground interface. In both cases a TNT target (dielectric constant = 2.9, loss tangent = 0.001) is embedded in a half space of dry sand (dielectric constant = 2.55, loss tangent = 0.01) or dispersive Bosnian soil (dielectric constant = 8.9, loss tangent = 0.13 at 3 GHz). All the rough surface scattering results to follow will exhibit the same roughness interface as displayed in the figure.

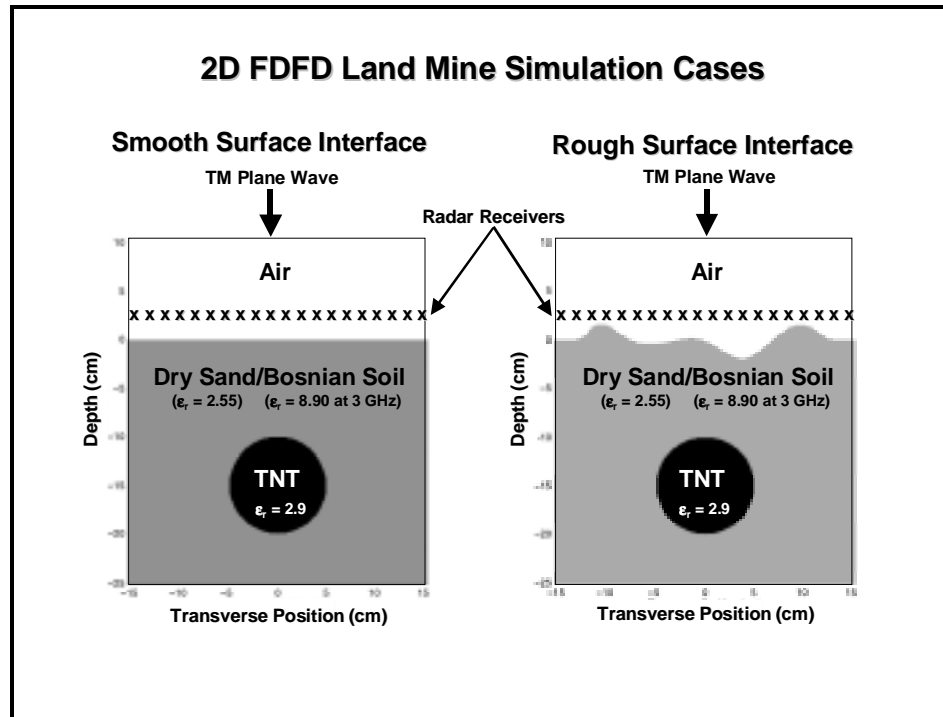


Figure 1: Representative 2D FDFD input geometry for an infinite dielectric cylindrical target buried 10 cm deep in the ground beneath both a smooth and rough air-ground interface.

The scattered TM electric field results (magnitude) in Figure 2 were generated by the 2D FDFD code using Figure 1's input model geometries and assuming normal incidence plane wave excitation. Displayed in the figure are the frequency-based spatial (as a function of depth and transverse position) scattered electric fields. Specifically, Figure 2 displays the scattered fields for the circular TNT target buried in both the sand and the Bosnian soil at individual frequencies (.5, 1, 2, 3, and 4 GHz) and beneath both the smooth and rough air-ground surface. All of the scattered field results had subtracted from them the scattering due to the smooth air ground interface alone, i.e. the ground reflection. Note that the scales in the figure vary to show the dominant scatterers in each plot in the figure.

The major differences in the scattering results for the circular mine buried under a smooth interface in the two different soils are as follows: (1) the TNT mine buried in the Bosnian soil exhibited a higher scattering response than the dry sand since it was a higher contrast target, (2) and at high frequencies the same target (buried in Bosnian soil) behaved like a cavity and thus the scattering was localized inside the target. When a rough surface interface was included for both ground types, the rough

surface scattering dominated the target scattering response as the frequency increased. Also notice in the Bosnian soil rough surface cases attenuation occurred at the higher frequencies due to the loss in the model.

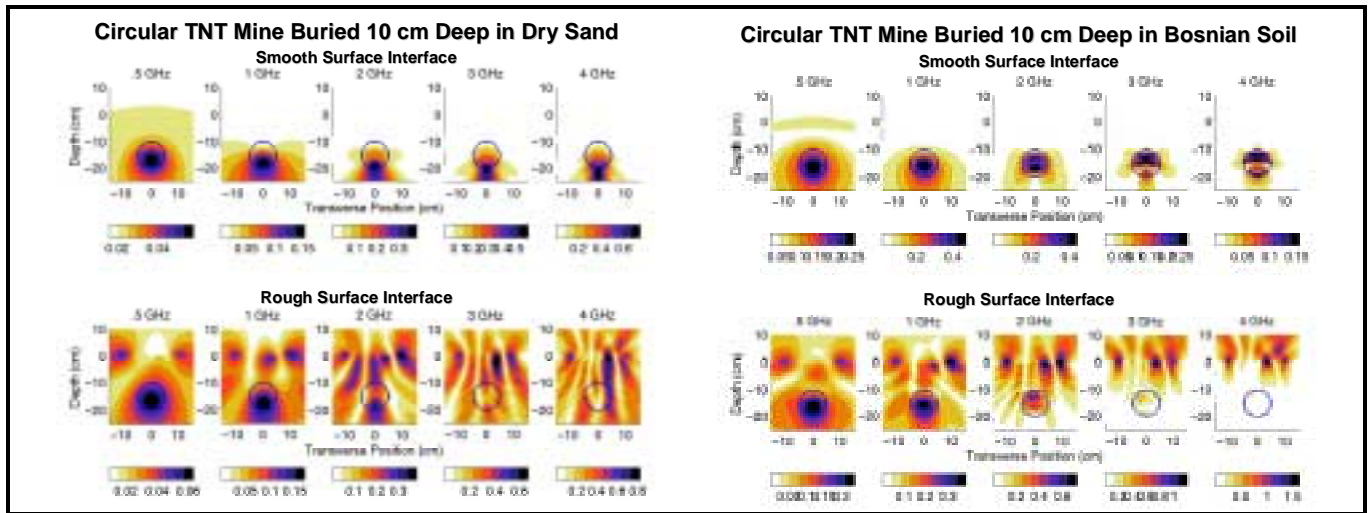


Figure 2: Spatial (Depth versus Position) E-Field scattering (magnitude) at individual frequencies (.5, 1, 2, 3, and 4 GHz) for a cylindrical mine (as denoted by the circles) buried 10 cm deep beneath a smooth and rough air-ground interface in two different soils (dry sand and Bosnian soil).

While the spatial scattering results (as displayed in Figure 2) helps one to understand the frequency dependent electromagnetic propagation and scattering effects associated with the different soil properties combined with the electrical characteristics of the target and the air-ground interface effects, the most important response to examine would be at the radar receiver/receivers locations. For this case, the analysis further assumes the radar receivers, as seen in Figure 1, to be placed across in the transverse position above the mine, and 2.5 cm above the ground surface. As would be expected, the most ideal radar receiver location is to be as close to the ground surface as possible for strong signal reception.

Figure 3 displays scattered fields (dBsm) for the three targets (circular, elliptical, and rectangular) as a function of frequency and transverse position of the receivers. The three objects were modeled as being buried 10 cm deep in the two soils (dry sand and Bosnian) beneath both the smooth and the rough air-ground interface. Also modeled was the case for the rough surface alone and no target is present. For the dry sand and a smooth air-ground interface cases, the received scattered fields for the three targets exhibited a constant resonating pattern across the frequency bandwidth. But in the case of the Bosnian soil, the received signal attenuated or faded as the frequency increased due to the presence of loss in the soil. The scattered field pattern of the rectangular object seem to be more distinctive than the circular and elliptical objects. As expected due to the large frequency bandwidth, the objects buried under a rough surface regardless of the soil type were indistinguishable from the rough surface case when no target was present. Also for both soil types the characteristic resonating pattern as seen in the smooth surface case was hidden by the rough ground surface reflections.

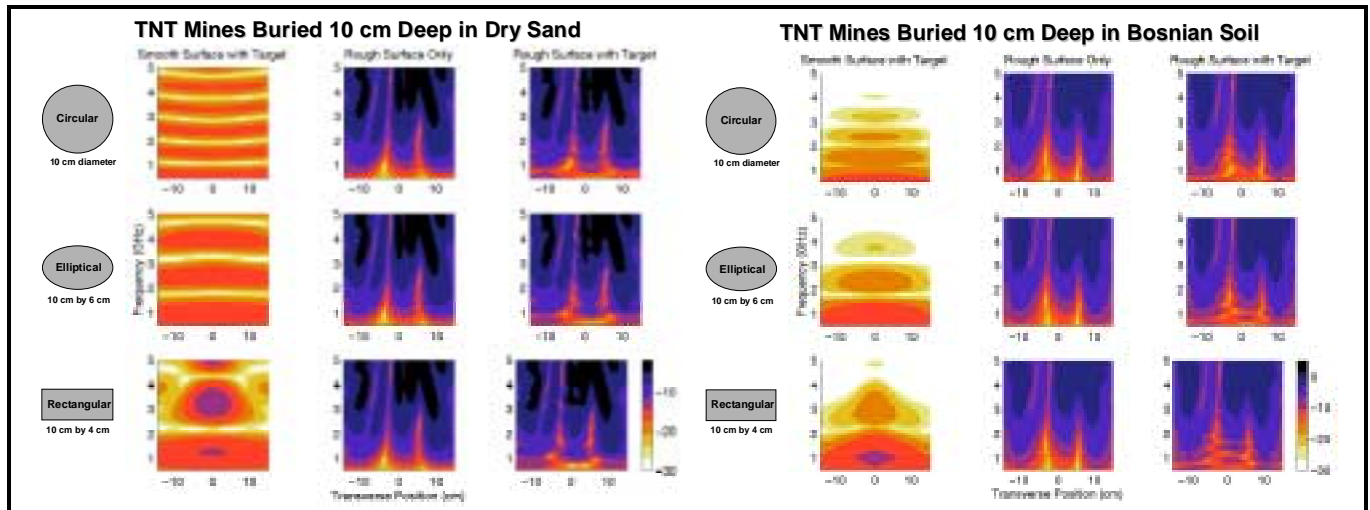


Figure 3: E-Field scattering (dBsm) versus frequency (GHz) and transverse position (cm) when radar receivers are placed 2.5 cm above a smooth and rough air-ground interface with a target buried 10 cm deep in soils (dry sand and Bosnian soil), and a rough air-ground interface with no buried targets present.

3. Time Domain Scattering Derived from GPR Frequency Signatures

This investigation examines the spectral response derived from the GPR frequency signatures as a viable discrimination technique in identifying a buried land mine from the clutter. A variety of spectral estimation techniques could be utilized to transform the GPR wideband frequency response to the time domain. These techniques have their associated benefits and limitations as widely known in the signal processing community. For example, before a waveform (one sweep over a given frequency range or bandwidth) is fourier transformed a weighting function is typically applied. This weighting function is used to mitigate the sidelobes that are produced by using a truncated fourier transform. A non-weighted transform will have a sidelobe level 13 dB below the peak response, but will have maximum resolution at 6 dB below the peak. In most cases, a Hamming weight is used to reduce the sidelobes to better than 40 dB below the peak. Unfortunately, the Hamming weight produces a loss of resolution by a factor of about 1.8. Therefore applying the Hamming weight is a trade-off since it helps to reduce the sidelobe levels but with the associated cost of decreasing the resolution.

After the wideband frequency data is fourier transformed to the time domain, highly resolved target features appear which directly correspond to the target's shape, size, and material content, the target's burial depth in the soil, and the soil's electrical properties. In particular, two or more characteristic timing peaks occur in the processed time domain signatures. For the case of a circular-shaped mine, as displayed in Figure 4, the peaks directly correspond to the scattering response from the top and bottom of the target. Due to the fact that the target is penetrable, these timing peaks (T_0 and T_1) correspond to the two-way travel times. T_0 corresponds to the round trip time for a transmitted wave to hit the top of mine and return back to the receiver. And accordingly, T_1 is the round trip time for the transmitted wave to hit the bottom of the mine and then return back to the receiver. Note due to the diffraction effects, the timing paths do not ideally correspond to a straight line as displayed in the figure. Other peaks can occur due to the geometry of the object, which might also include scattering effects from

sharp corners. If the dielectric constant of the land mine and the soil are known a priori then these timing peaks can be used to uniquely determine the target size and depth. Time domain processing shows features that are invariant to the ground roughness.

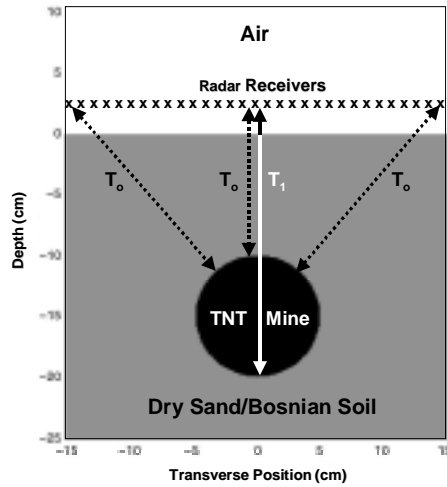


Figure 4: Schematic drawing of two way travel times, T_0 and T_1 , for a buried cylinder when radar receivers are transversely placed 2.5 cm above a smooth air-ground interface.

The simulated frequency response results for the buried mine-like targets in Figure 3 were transformed to the time domain using a traditional signal processing technique, the fast fourier transform (FFT). The WB frequency response for each receiver location was fourier transformed to the time domain producing a synthetic aperture-like response. Specifically, each column (or WB frequency response results for each receiver position, labeled transverse position) in the data sets displayed in Figure 3 was individually fourier transformed to the time domain. The time domain results as a function of transverse position (at fixed height 2.5 cm above the surface) or equivalently angle are displayed in Figure 5. As seen in the figure, the timing peaks as a function of position, as denoted by T_0 and T_1 , are clearly resolved for the three different targets shapes buried in both soils (dry sand and Bosnian soil) and beneath a smooth air-ground interface with the ground reflections removed. Since the dielectric constant of the Bosnian soil was higher than that of the sand the velocity of wave through the soil is slower and therefore the timing peaks are delayed in time. The difference between the timing peaks ($T_0 - T_1$), which directly corresponds to the size of the object, is always constant regardless of the soil type. Thus the quantity $T_0 - T_1$ is independent of the soil.

The timing peaks in Figure 5 are also distinguishable for the cases when the targets are buried beneath a rough surface for both soil types. The time domain results for the rough surface interface alone (no targets present) are also displayed as a visual comparison. When the buried target is modeled beneath a rough surface, the timing peaks associated with the target are sometimes obscured in the sidelobes of the signal at different transverse positions. But a different signal processing technique combined with coherent integration in position could also help this artifact. Due to the fourier transform of the data to the time domain, there is now a difference between the case when the target is present beneath a rough surface interface and the case when it is not present, notably, the timing peaks beneath the rough surface are observable.

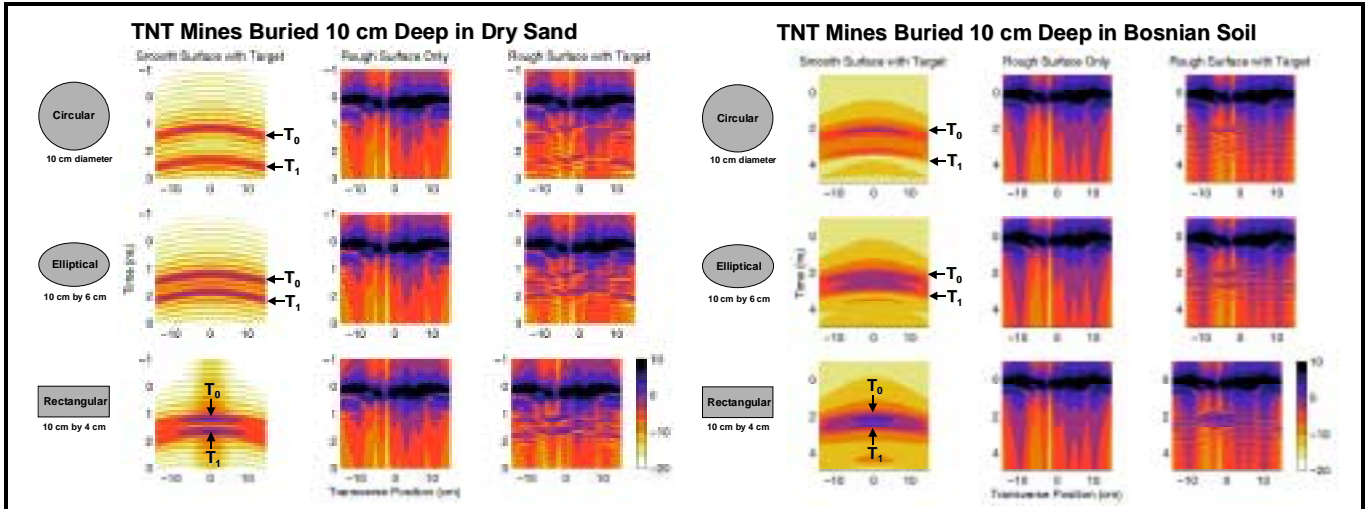


Figure 5: Processed time domain E-Field scattering (dBsm) versus time (ns) and transverse position (cm) for the three targets buried in soils beneath both a smooth air-ground interface and a rough air-ground interface, and a rough interface only (no target present)

The processed time domain scattered field results in Figure 5 were nominally time gated using a gaussian tapered high pass filter to remove some of the ground surface clutter effects as seen in Figure 6.

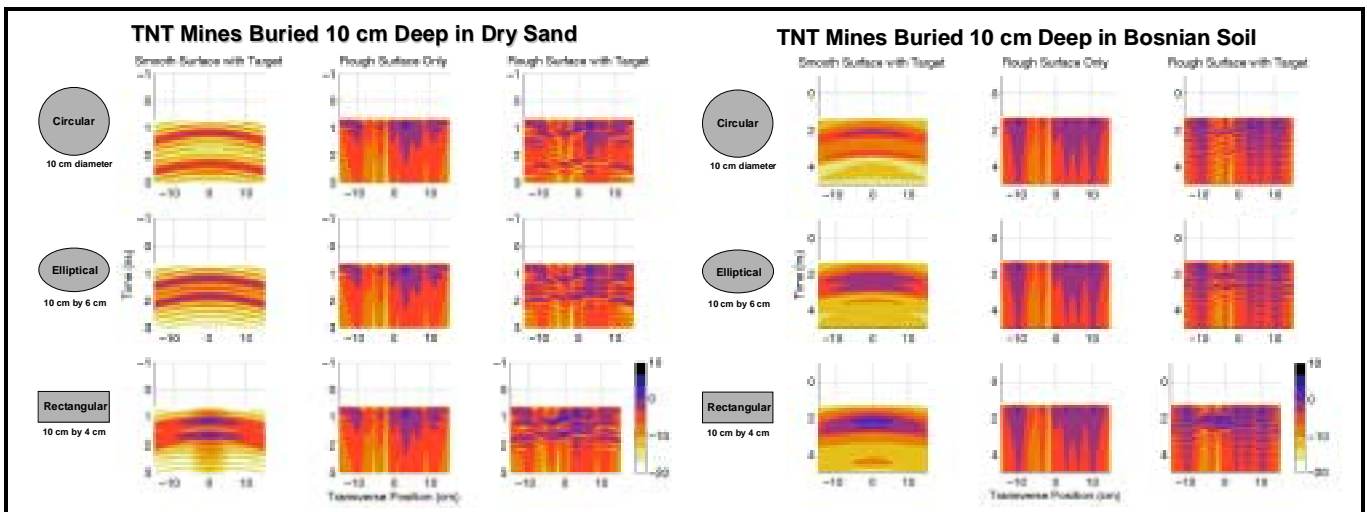


Figure 6: Removal of strong ground reflections: time domain E-Field scattering (dBsm) with gaussian tapered high pass filtering for the three targets buried in sand and Bosnian soil beneath both a smooth air-ground interface and a rough air-ground interface, and a rough interface only (no target present)

The time-gated scattered fields from Figure 6 were then inverse fourier transformed to return back to the frequency domain. It is shown in Figure 7 that by using time-gating to remove the strong ground reflection signals, the target signals are selectively enhanced (as would be expected), but more surprisingly, the target frequency response signature is almost completely recovered (as seen in the right most column for each ground type in the figure).

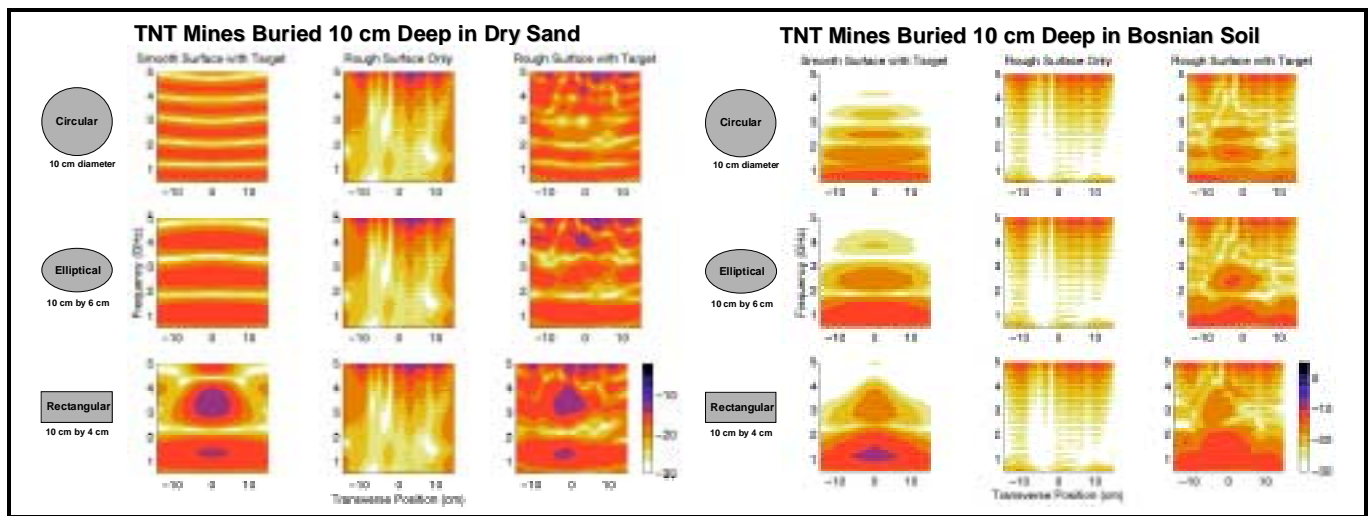


Figure 7: Processed frequency domain E-field scattering (dBsm). The processed time-gated data in Figure 6 was inverse fourier transformed for the three targets buried in sand and Bosnian soil beneath both a smooth air-ground interface and a rough air-ground interface, and a rough interface only (no target present).

4. Conclusions

The FDFD electromagnetic modeling technique proved to be a valuable computational tool in understanding the wideband propagation and scattering effects of buried dielectric mine-like objects in lossy dispersive rough soils. The reflected target frequency response is severely dependent on the nature of the soil both permittivity and loss, the target dielectric constant contrast with the soil, and the roughness of the soil. It is vital to understand these combined effects before a useful discrimination technique can be identified and developed. The wideband stepped-frequency GPR-based simulation results showed that the clutter from the rough ground reflections masked the target's frequency response signatures.

The signal processing of the ultra WB GPR simulated data to the time domain helped to resolve unique target features that are invariant to the ground roughness. In particular two or more timing peaks are observed which can specifically determine the size and depth of a target if the electrical characteristics of the soil and target are known a priori. These time-domain based features could uniquely characterize the target from the clutter. The target frequency response signature was almost completely recovered after nominally time-gating the ground reflection signals and then inverse fourier transforming the signatures to the frequency domain.

To be able to solve the challenging inverse problem of identifying small buried plastic land mines, one must be able to separate the ground surface reflections from the target signal response. If using a frequency resonance-based discrimination algorithm, the ground clutter clearly necessitates removal as seen in the simulation results presented herein.

Future target detection and identification algorithms will be explored in the delayed time domain environment after signal processing the frequency response signatures. This will include the detection of a land mine-like target from multiple objects (rocks, branches, debris, etc.) buried in lossy rough soils.

5. References

1. A. W. Morgenthaler and C. M. Rappaport, "Semi-analytic mode matching for detecting nonmetallic mines buried in realistic soils," *Proc. Of SPIE Aerosense, Detection and Remediation Technologies for Mines and Minelike Targets V*, **4038**:2, pp. 1149-59, 2000.
2. A. W. Morgenthaler and C. M. Rappaport, "The semi-analytic mode matching algorithm for accurate, efficient computation of scattered fields from 2D and 3D dielectric objects buried in randomly rough lossy backgrounds," to be submitted to *IEEE Trans. on Ant. and Propagat.*
3. A. W. Morgenthaler and C. M. Rappaport, "Scattering from Lossy Dielectric Objects Buried Beneath Randomly Rough Ground: Validating the Semi-Analytic Mode Matching Algorithm with 2-D FDFD," *IEEE Trans. on Geoscience and Remote Sensing*, **39**:11, pp. 2421-2428, Nov 2001.
4. A. W. Morgenthaler and C. M. Rappaport, "Scattering from Dielectric Objects Buried Beneath Random Rough Ground: Validating the Semi-Analytic Mode Matching Algorithm with Two-Dimensional FDFD," *Proc. Of IEEE Intl Geosci. And Remote Sens. Symp.*, **4**, pp. 1634-1636, 2000.
5. B. Yang and C. M. Rappaport, "Response of Realistic Soil for GPR Applications with Two Dimensional FDTD," *IEEE Trans. on Geoscience and Remote Sensing*, **39**:6, pp. 1198-1205, June 2001.
6. C. Rappaport, S. Wu, and S. Winton, "FDTD Wave Propagation Modeling in Dispersive Soil Using a Single Pole Conductivity Model," *IEEE Transactions on Magnetism*, **35**:5, pp. 1542-1545, May 1999.
7. J. Curtis, "Dielectric properties of soils: various sites in Bosnia," US Army Corp. of Eng., Waterways Experim. Station Data Rep., 1996.ed cavities," *IEEE Trans. Microwave Theory and Tech.*, **43**:8, pp. 1848-55, 1995.

FDTD Computation of Wave Propagation in Dispersive Soil with a Perfectly Matched ABC Lattice Termination

Carey M. Rappaport and Scott C. Winton
Center for Electromagnetics Research
Northeastern University
Boston, MA 02115

In order to better predict the propagation and scattering of radar signals in inhomogeneous soil, it is important to use efficient computational techniques, such as the Finite Difference Time Domain method, suitably adjusted for lossy, dispersive media. In particular, the recently reported Perfectly Matched Layer absorbing boundary condition (Berenger *J. Comp. Phys.*, 10/1994) must be modified to most effectively prevent reflections from the computational lattice boundary.

The FDTD method is ideally suited for ultra-wideband ground penetrating radar simulation in applications of identifying land mines, buried waste, and excavation obstacles. FDTD efficiently models the irregular surface boundary and provides the needed flexibility in predicting radar scattering from both metal and plastic targets, buried in soil with rock inclusions and topped with vegetation. Such realistic scenarios are impractical for moment method calculations.

Dispersive FDTD algorithms exist which include the effects of frequency-dependent conductivity and dielectric constant (Weedon & Rappaport *IEEE Trans. Ant. Prop.*, to appear 1997; Gandhi, *IEEE Trans. Micro. Thry. Tech.*, 4/93). These methods are based on converting the frequency domain dispersive constitutive relation into differential or difference equations in the time domain. Because of the computational expense of storing previous time steps, it is essential to keep the order of any electromagnetic parameter approximation as low as possible.

Another important difference in using FDTD (or FDFD or finite element analysis) with a lossy, dispersive medium is the ABC used to terminate the computational grid. While the PML is an excellent ABC when applied at the periphery of a free space scattering domain, unless its parameters are modified, it will not be as effective when terminating a region modeling soil. The soil-tuned PML must be inherently matched to the soil at the ABC boundary. In the frequency domain, this is accomplished by matching the impedances of the two regions:

$$\eta_{soil} = \sqrt{\frac{\mu}{\epsilon_0(\epsilon_r - j\sigma/\omega\epsilon_0)}} = \eta_{PML} = \sqrt{\frac{\mu_{PML}}{\epsilon_{PML}}}$$

To maintain the frequency independence of this match the permeability and permittivity of the PML must have the form:

$$\mu_{PML} = \mu(1 - j\frac{\sigma_{PML}}{\omega\epsilon_0})$$

and

$$\sigma_{PML} = \epsilon_0(\epsilon_r - j\frac{\sigma}{\omega\epsilon_0})(1 - j\frac{\sigma_{PML}}{\omega\epsilon_0})$$

Converting this last expression into the time domain is problematic because of the double frequency dependence—which leads to a second time derivative in Ampere's Law. However, computed FDTD results for several soil cases show excellent suppression of unwanted edge boundary reflections.

Higher Order (Nonlinear) Diffraction Tomography: Inversion of the Rytov Series

George A. Tsihrintzis, *Member, IEEE*, and Anthony J. Devaney, *Member, IEEE*

Abstract—Nonlinear tomographic reconstruction algorithms are developed for inversion of data measured in scattering experiments in which the complex phase of the wavefields is modeled by an arbitrarily large (possibly infinite) number of terms in the Rytov series. The algorithms attain the form of a Volterra series of nonlinear operators, with the usual filtered backpropagation algorithm of Diffraction Tomography as the leading linear term. A computer simulation study is included to illustrate the performance of the algorithms for the case of scattering objects with cylindrical symmetry.

Index Terms—Born series, diffraction tomography, inverse scattering, Rytov series, Volterra series.

I. INTRODUCTION

A. Early Development of Diffraction Tomography

DIFFRACTION Tomography (DT) emerged over the past twenty years as a linearized approach to the problem of quantitative determination of the structure of an unknown object from measurement of the waves diffracted by the object in a suite of scattering experiments [102], [17], [22], [103]. The structure to be reconstructed by DT is usually the spatial distribution of the complex-valued index of refraction inside the object. Therefore, DT lies within the field of Inverse Scattering [56], [6], [9] and is applicable in a number of seemingly different scientific disciplines, such as crystal structure determination [48], medical ultrasound tomography [33], acoustic and electromagnetic underground surveying [19], [98], [99], [97], [52], [100], optical and coherent X-ray microscopy [51], and elastic wave inverse scattering [25]. DT differs from other Inverse Scattering theories [104], [35], [2], [5], [4], [37], [92], [57], [41], [42], [12], [13] in that it utilizes linearizing approximations, namely, the Born or the Rytov approximation, to the nonlinear mathematical mapping that relates the probing wave, the object refraction index, and the measurable total wave.

The foundation of modern linearized DT lies in the generalized projection-slice theorem of (A.10), which was recognized by Wolf [102] in 1969. Wolf [102] employed the Born approximation together with truncation of the evanescent wave modes

in the Weyl expansion and the need was recognized for a tomographic procedure employing different directions of illumination for a reasonable reconstruction of an object's complex-valued index of refraction. Wolf's formulation was extended in 1974 by Iwata and Nagata [36] to determine the structure of a less restrictive class of scatterers satisfying the Rytov rather than the Born approximation. In 1979, Mueller *et al.* [55] employed the same concepts of the Born and Rytov approximations and presented Fourier interpolation-based algorithms for the inverse problem of ultrasound tomography, while in 1982, Devaney [17] derived an elegant inversion algorithm, named "*the filtered backpropagation algorithm of DT*," for the inversion of full view, scattered field data under the Born or Rytov approximations. When scattering experiments are done at a wavelength λ , the filtered backpropagation algorithm returns an estimate of the unknown index of refraction distribution whose frequency content is the same as of the true distribution over a circular disk in Fourier space of radius $\frac{2\pi}{\lambda}\sqrt{2}$ and zero elsewhere [23].

The filtered backpropagation algorithm has been recognized as the one providing highest quality in the reconstructed images [59] and modifications to it have been presented by Devaney [19] in 1984 and Deming and Devaney [11] in 1996 to adjust it to the configurations employed in geophysical tomographic surveys. Tsihrintzis and Devaney addressed the reconstruction problem of linearized DT from noisy scattered field data and showed that the optimum (Wiener) estimation filter attains again the form of a filtered backpropagation algorithm [85], [86]. Finally, iterative algorithms for inversion of an angularly limited set of noise-free linearized scattering data were presented by Ladas and Devaney in [44], [45].

B. Recent Developments in Diffraction Tomography

Linearized DT has reached today the stage of being implemented in prototype commercial tomographic scanners for ultrasonic [79], [84], underground [62], [97], [100], and optical [51] imaging systems. Particularly successful have been geophysical DT algorithms when applied to a range of underground imaging problems such as oil field prospecting and reservoir monitoring [101], locating underground tunnels between North and South Korea [97], [100], and searching for dinosaur bones in the New Mexico desert [96], [100]. The success of the linearized DT algorithms depends critically, however, on two assumptions: 1) linearity and 2) availability of multiple experiments [14]. In many cases, the linearity assumption fails, while different constraints (economic, safety, operating, geometric, or physical) limit the number of scattering experiments that can be performed and/or provide low signal-to-noise ratio data. Even though algebraic reconstruction techniques reduce the effect of

Manuscript received April 20, 1999; revised April 16, 2000. This work was supported by ARO under Grant DAAG55-97-1-0013, by AFOSR under Grant F49620-96-1-0039, and by the Center for Subsurface Sensing and Imaging Systems (CenSSIS) of Northeastern University.

G. A. Tsihrintzis is with the Department of Informatics, University of Piraeus, Piraeus 18534, Greece (e-mail: geoatsi@unipi.gr).

A. J. Devaney is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (tonydev2@aol.com).

Communicated by A. O. Hero III, Guest Editor.

Publisher Item Identifier S 0018-9448(00)05458-4.

availability of only a small number of scattering experiments, the effects of nonlinearity are much harder to combat and remain an issue of current research. To overcome these limitations a more modest inverse scattering problem was addressed by Tsihrintzis and Devaney, originally within the framework of linearized [27], [82], [84], [87] and later exact [83], [84], [90], [91] scattering theory.¹ The goal of this more modest inverse problem, motivated in part by earlier work on a related problem in X-ray computed tomography [66], was to estimate the location of a single *known* scatterer having unknown central location from noisy scattered field data. It was found, that for monochromatic planewave probing, the optimum (in the maximum-likelihood sense) location estimate could be obtained via a filtered backpropagation algorithm, in which partial images formed by filtering and backpropagating scattered field data for different probing directions were coherently summed. The algorithm yields an *image* of the log-likelihood function of the object's location and can be used for target detection and classification as well as target location estimation. The detection/estimation/classification procedure is optimum (in the maximum-likelihood sense) for a single target and any given number of scattering experiments and returns good estimates even from a single experiment as long as the wavelength of the probing radiation is comparable with the typical dimensions of the target [27]. Its independent implementation in practice has revealed robustness to both high noise levels and the presence of unmodeled objects.

C. Present Contribution

One practically important imaging situation arises when the object consists of a number of distinct scatterers. As pointed out by Azimi and Kak [1] and Slaney, Kak, and Larsen [73], even though each scatterer individually may be weak enough for validity of the Born or Rytov approximate model, multiple scattering interactions among several scatterers degrade the performance of linearized DT reconstruction algorithms. The situation can be partially ameliorated if the reconstruction algorithms are based on higher order (nonlinear) scattering models and, indeed, formal series solutions to the inverse scattering problem have been presented in the literature [63]–[65], [28], [49], [50], [93], [94], [74]–[76]. In [28], [49], [50] more specifically, perturbative expansions of the scattering object's Fourier transform were utilized to develop DT reconstruction algorithms of arbitrary order, which contained linear reconstruction algorithms as special cases and effectively attained the form of nonlinear data filtering followed by a linear operation. The algorithms in [28], [49], and [50] attempt to estimate samples of the Fourier transform of the object function and use them to obtain an estimate of the object function itself. Recent attempts ([61], [3] and references therein) to invert a second-order scattering model have resulted in algorithms of the form of iterative numerical solutions of systems of quadratic equations and revealed significantly higher fidelity than their linear counterparts. In this paper, we address the reconstruction problem of DT within the

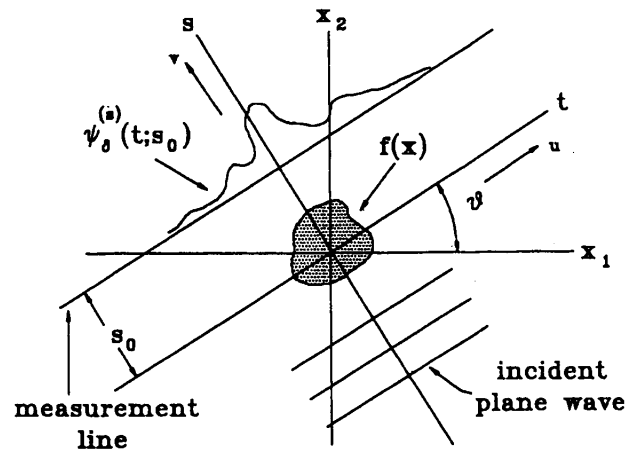


Fig. 1. The classical scan configuration of Diffraction Tomography

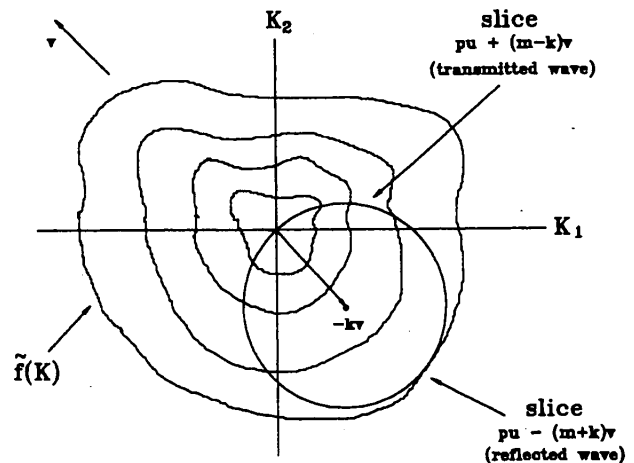


Fig. 2. The generalized projection-slice theorem of Diffraction Tomography

framework of higher order (nonlinear) scattering approximations. We derive reconstruction algorithms for the so-called, "classical scan configuration of DT" of Fig. 1 (and the associated Fig. 2), that can be implemented via use of a fast Fourier transform (FFT) algorithm, and illustrate them in a computer simulation study. Thus the paper presents a successive approximation procedure that recursively extracts increased information about a scatterer, where "increased information" is interpreted in the sense of reduction of the bias of the solution. In this sense, the present paper is also related to our previous work on information extraction from imperfectly modeled scattered wavefield data [27], [83], [82], [85], [84], [86], [90], [91] because it attempts to construct a wave-theoretic model for the error term that originally appeared as random noise in that work and suppress its biasing effect on the inversion algorithm output via the use of rigorous methodologies provided by the theory of Volterra system inversion.

More specifically, the paper is organized as follows. Section II defines nonlinear data-generating models for scattering experiments through nonlinear mathematical mappings from the object structure to the complex phase of the observed total wavefields that arise from an arbitrarily large number of terms in the

¹The algorithm was originally developed for homogeneous background media, but was recently extended to arbitrary one-dimensional (e.g., planar layered) media [91]. In the same manner, it can be extended to arbitrary two-dimensional media (e.g., a waveguide of arbitrary cross section).

Rytov series. Section III is devoted to study of the inversion of the nonlinear data-generating models, a discussion of the implementation procedure of the proposed algorithms, and a limited computer simulation, while Section IV summarizes the paper, draws conclusions, and points to future related research. Three appendices are attached to the paper. Appendix A summarizes the Born series representation of the mapping from object structure to scattered wavefield data. Appendix B makes use of the Weyl expansion to derive the generalized projection-slice theorem and define the propagation transform of linearized DT. Appendix C summarizes the key elements of the Volterra theory of nonlinear systems.

II. NONLINEAR DATA-GENERATING MODELS FOR DIFFRACTION TOMOGRAPHY

A. Configuration and Scattering Equations

Consider the data collection configuration illustrated in Fig. 1, where a known monochromatic plane wave $\psi_0(x) = e^{ik\langle v, x \rangle}$ of wavenumber k propagates in the direction of the unit vector v and is incident on a scattering object of support \mathcal{V} . The total wave is measured over a line located entirely outside the object and perpendicular to the unit vector v . A point on the measurement line is denoted with the coordinate vector $x_0 = t_0 u + s_0 v$, $t_0 \in \mathbb{R}^1$, where s_0 (fixed) is the distance of the measurement line from the origin. The object is assumed to be embedded in a *known* nonattenuating, homogeneous background medium of wave velocity c_0 and wavenumber $k = \frac{\omega}{c_0}$ and characterized by its distribution of complex-valued index of refraction $n(x) = c_0/c(x)$, where $c(x)$ is the wave velocity distribution inside the scatterer.

The interaction of the incident wave with the object results in the formation of a wavefield $\psi(x)$ that satisfies the time-independent inhomogeneous Helmholtz equation

$$(\nabla^2 + k^2)\psi(x) = -2k^2 f(x)\psi(x) \quad (2.1)$$

where $f(x) = \frac{1}{2}[n^2(x) - 1]$ is the *object function*, the quantity whose determination is the ultimate goal of inverse scattering theory and DT. The solution of the Helmholtz equation (2.1) can be decomposed into the sum of the incident field and a scattered field in the form $\psi(x) = \psi_0(x) + \psi^{(s)}(x)$, where the scattered field component is related to the object function and total field via the integral equation

$$\psi^{(s)}(x) = -2k^2 \int_{\mathcal{V}} dx' f(x') \psi(x') G(x - x') \quad (2.2)$$

with $G(x - x') = -\frac{i}{4} H_0^{(1)}(k|x - x'|)$ the outgoing wave Green function to the Helmholtz equation [53].

Equation (2.2) provides a link between source (radiation) and scattering problems in wave theory and shows that the scattered field *outside* the scattering region \mathcal{V} is uniquely determined from the total field *inside* \mathcal{V} . Subsequently, approximate scattering models, such as the Born or Rytov models, seek to replace the “induced source” $f\psi$ in (2.2) with an approximate expression that leads to a tractable inverse problem. The first Born and Rytov approximations lead to linear models that are inverted via the algorithms of linearized DT outlined in Section I. The objec-

tive of this paper is to improve these linear models by including higher order terms in the Rytov series and to establish a unified paradigm of both linear and nonlinear inversion algorithms for DT.

B. The Rytov Series

The Rytov series was originally developed within the context of wave scattering in random and turbulent media [81], [7], [72] and, as an approximation to exact scattering, its domain of validity is broader than the domain of validity of the Born series. It arises from a perturbational expansion of the complex phase (i.e., the complex logarithm) of the field rather than the field itself. To begin, f is replaced by ϵf in (2.1) and its solution $\psi(x; \epsilon)$ is set to

$$\psi(x; \epsilon) = e^{ikW(x; \epsilon)} \quad (2.3)$$

where $W(x; \epsilon)$ is the “wavenumber-normalized complex phase” of the total field $\psi(x; \epsilon)$. Substitution of (2.3) in (2.1) shows that the phase W satisfies the nonlinear Riccati equation

$$\nabla W \cdot \nabla W + \frac{1}{ik} \nabla^2 W - 1 = 2\epsilon f. \quad (2.4)$$

So far, no approximation has been made. Equation (2.4) is exact and mathematically equivalent to (2.1).

The Rytov series solution to (2.4) is obtained by substituting in (2.4) the formal series

$$W(x; \epsilon) = W_0(x) + \epsilon W_1(x) + \epsilon^2 W_2(x) + \dots \quad (2.5)$$

with $\psi_0 = e^{ikW_0}$. The resulting equations that specify the individual terms in the series (2.5) are

$$\nabla W_0 \cdot \nabla W_0 + \frac{1}{ik} \nabla^2 W_0 = 1 \quad (2.6)$$

$$2\nabla W_0 \cdot \nabla W_1 + \frac{1}{ik} \nabla^2 W_1 = 2f \quad (2.7)$$

$$2(\nabla W_0 \cdot \nabla W_2 + \nabla W_1 \cdot \nabla W_1) + \frac{1}{ik} \nabla^2 W_2 = 0$$

$$\vdots$$

$$2(\nabla W_0 \cdot \nabla W_n + \nabla W_1 \cdot \nabla W_{n-1} + \dots) + \frac{1}{ik} \nabla^2 W_n = 0 \quad (2.8)$$

From (2.8), it is clear that, unlike the Born series in which the n th-order Born term is computed directly from the Born term of order $n - 1$ only, computation of the n th-order Rytov term W_n makes use of all the Rytov terms $W_{n-1}, W_{n-2}, \dots, W_0$. The solution to (2.6)–(2.8) is given next. The solution to (2.6) is the complex phase of the incident field, i.e., $W_0(x) = \langle v, x \rangle$. To solve (2.7), the Rytov transformation [81], [7], [72] is applied

$$W_1 = e^{-ikW_0} F. \quad (2.9)$$

The quantity F is seen to satisfy the equation

$$(\nabla^2 + k^2)F = -\frac{2kf}{i} e^{ikW_0} \quad (2.10)$$

which is solved using Green function techniques to obtain²

$$F(x) = -\frac{2k}{i} \int_{\mathcal{V}} dx' G(x-x') f(x') e^{ikW_0(x')}. \quad (2.11)$$

Substitution in (2.9) of the expression in (2.11) leads to

$$W_1(x) = -\frac{2ke^{-ikW_0(x)}}{i} \int_{\mathcal{V}} dx' G(x-x') f(x') e^{ikW_0(x')}. \quad (2.12)$$

To obtain the term W_n in series (2.5), we rewrite (2.8) as

$$2\nabla W_0 \cdot \nabla W_n + \frac{1}{ik} \nabla^2 W_n = -\sum_{j=1}^{n-1} \nabla W_j \cdot \nabla W_{n-j} \quad (2.13)$$

and apply the Rytov transformation to it. The result is the expression in the right-hand side of (2.12) with $2f$ replaced by $-\sum_{j=1}^{n-1} \nabla W_j \cdot \nabla W_{n-j}$, i.e.,

$$W_n(x) = \frac{ke^{-ikW_0(x)}}{i} \int_{\mathcal{V}} dx' G(x-x') \cdot \left[\sum_{j=1}^{n-1} \nabla W_j \cdot \nabla W_{n-j} \right] (x') e^{ikW_0(x')}. \quad (2.14)$$

The Born and the Rytov terms are related [72], [67]. A Rytov term of order n can be expressed in terms of the Born terms of order $0, 1, \dots, n$. Indeed, substitution of the Born and the Rytov series in (2.3) gives

$$\sum_{n=0}^{\infty} \epsilon^n \psi_n = \exp \left(ik \sum_{n=0}^{\infty} \epsilon^n W_n \right) \quad (2.15)$$

which, after expanding around $\epsilon = 0$ and equating terms of equal power in ϵ , gives

$$W_0 = \frac{1}{ik} \log \psi_0 \quad (2.16)$$

$$W_1 = \frac{1}{ik} \frac{\psi_1}{\psi_0} \quad (2.17)$$

$$W_2 = \frac{1}{ik} \left[\frac{\psi_2}{\psi_0} - \frac{1}{2} \left(\frac{\psi_1}{\psi_0} \right)^2 \right] \quad (2.18)$$

$$W_3 = \frac{1}{ik} \left[\frac{\psi_3}{\psi_0} - \frac{\psi_2}{\psi_0} \frac{\psi_1}{\psi_0} + \frac{\psi_1^3}{3!} \right] \quad (2.19)$$

\vdots

Similar expressions can be written to express a Born term of order n in terms of the Rytov terms of order $0, 1, \dots, n$. Equations (2.16)-(2.19) reveal a pointwise (memoryless), with respect to the x variable, transformation from Born to Rytov terms. Given the fact that Born term computation can be approximately done by algorithms that are based on the fast Fourier transform, (2.16)-(2.19) indicate that an efficient

procedure for computing a Rytov term of order n would be to compute first the Born terms of order $0, 1, \dots, n$ and from these to find the n th-order Rytov term via a pointwise (memoryless), with respect to the x variable, transformation.

The first-order Rytov approximation, simply referred to as the Rytov approximation, consists of approximating the field phase W with the sum of only the zeroth- and first-order Rytov terms as $W(x) \approx W_0(x) + W_1(x)$, where the term W_1 is to be computed from (2.12) using the fact that $W_0(x) = \langle v, x \rangle$. In terms of the total field at point x

$$\begin{aligned} \psi(x) &\approx e^{ikW_0(x)} e^{ikW_1(x)} \\ &= \exp\{ik\langle v, x \rangle\} \exp \left\{ -2ke^{-ik\langle v, x \rangle} \int_{\mathcal{V}} dx' G(x-x') \cdot f(x') e^{ik\langle v, x' \rangle} \right\}. \end{aligned} \quad (2.20)$$

Clearly, the Rytov approximation estimates the total field as the product of the incident field $\psi_0 = e^{ikW_0}$ and a multiplicative factor e^{ikW_1} that depends on the object function f . This is in contrast with the Born approximation which estimates the total field by correcting the incident field by an additive term that depends on the object function f . As a result, the Rytov approximation implies a *nonlinear* mapping from the object function to the total and scattered fields. It is, however, a linearizing approximation for the direct scattering problem in the sense that the mapping from the object function to the *complex phase* of the total field has been linearized.

The Rytov approximation is less restrictive than the Born approximation. Indeed, the Rytov approximation is valid under the mild condition [40]

$$|\nabla W_1|^2 \ll f_m k^2, \quad \text{where } f_m = \max_x |f(x)| \quad (2.21)$$

in which no restriction on the size of the scattering region \mathcal{V} has been imposed. The main disadvantages, however, of the Rytov approximation are that 1) it works with the phase of the fields rather than the fields themselves, which creates phase unwrapping problems in inverse scattering applications [38], and 2) that it is a near-field approximation, since as observation points x_0 are considered further away from the scattering region \mathcal{V} , it deteriorates fast and in the far field it becomes identical to the Born approximation [10], [24]. This latter fact has led [24] to a modified form of the Rytov approximation, which combines aspects of both the Born and the Rytov models, named the "hybrid model."

Both the Born and the Rytov approximations have been extensively used for predictions in direct [81], [7], [72] and inverse [15], [18] scattering theory. Their domains of validity, as well as their relative advantages and disadvantages, have been extensively explored in the literature [40], [39], [58], [95], [77], [78], [47] and also compared with experimental results [80], [51]. A concise treatment of these and other approximate models for inverse scattering applications can be found in [80] and the references therein. In this paper, use is made of the Rytov series, on the basis of the fact that experimentation has shown for it better modeling properties than the Born series for the type of scattering problems encountered in DT [80], [51].

²Clearly, F is simply equal to the wavenumber-normalized Born approximation $\frac{\psi_1}{ik}$ to the scattered field, as seen from the top of (A.2) and the fact that $\psi_0 = e^{ikW_0}$. This leads to the simple expression $\psi_1 = ik\psi_0 W_1$ relating the first terms in the Born and the Rytov series. These two approximations have quite different domains of validity, however.

C. Nonlinear Data-Generating Models for Two-Dimensional DT

Consider again the classical scan configuration of Fig. 1, but assume that the complex phase of the total field along the line $x_0 = t_0 u + s_0 v$, $t_0 \in R^1$, s_0 : fixed, is well approximated by the first N terms of the Rytov series, where the possibility of $N \rightarrow \infty$ is not excluded

$$W(x_0) \approx W_0(x_0) + W_1(x_0) + \cdots + W_N(x_0). \quad (2.22)$$

Define the Hilbert spaces \mathcal{F} , \mathcal{F}_L , and \mathcal{P} as follows: \mathcal{F} is $L^2(\mathcal{D}_a)$, where \mathcal{D}_a is a disk of radius a in R^2 that includes the scattering region \mathcal{V} . \mathcal{F}_L is the subspace of \mathcal{F} consisting of the functions in \mathcal{F} lowpass filtered down to a disk of radius $k\sqrt{2}$ in Fourier space, i.e.,³

$\mathcal{F}_L = (f_L | \mathcal{D}_a): \exists (f | \mathcal{D}_a) \in \mathcal{F}$ such that

$$\tilde{f}_L(K) = \begin{cases} \tilde{f}(K), & \text{for } |K| \leq k\sqrt{2} \\ 0, & \text{otherwise.} \end{cases}$$

Finally, \mathcal{P} is the space of functions defined on $R^1 \times [0, 2\pi]$ whose frequency content with respect to the first variable is zero outside the $[-k, k]$ interval.

Define the nonlinear operators

$$\mathcal{W}_n : \mathcal{F}_L \rightarrow \mathcal{P} \quad (2.23)$$

where

$$(\mathcal{W}_n f_L)(t_0, \theta) = W_n^L(t_0 u + s_0 v, \theta) \quad (2.24)$$

and $W_n^L(t_0 u + s_0 v, \theta)$ is the n th term of the Rytov series to the wavenumber-normalized complex phase of the total field along the measurement line $x_0 = t_0 u + s_0 v$, lowpass-filtered with respect to t_0 down to the frequency interval $[-k, k]$. From (2.16) to (2.19) and Appendix A, it is seen that \mathcal{W}_n is a homogeneous Volterra operator (see Appendix C) of order n .

Define now the linear operator

$$\mathcal{B} : \mathcal{P} \rightarrow \mathcal{F}_L \quad (2.25)$$

with

$$(\mathcal{B}d)(x = tu + sv) = \frac{1}{2} \frac{1}{(2\pi)^2} \int_0^{2\pi} d\theta \int_{-k}^k dp |p| e^{ipt} e^{i(m-k)(s-s_0)} \tilde{d}(p, \theta). \quad (2.26)$$

The operator $(\mathcal{B}|\mathcal{P})$ is the inverse of $(\mathcal{W}_1|\mathcal{F}_L)$, as stated in (A.13).

³Since the functions f in \mathcal{F} are compactly supported, their Fourier transform is an entire function. Therefore, for given f_L , the corresponding f is unique and can, in principle, be obtained by analytically continuing the Fourier transform of f_L . The lowpass filtering is essentially a regularization procedure that makes the inverse of the propagation transform of linearized DT continuous.

D. The Cylindrically Symmetric Case

A particularly important imaging situation is that of scattering objects with cylindrical symmetry, that is, scatterers whose properties vary only with the radial distance from the scatterer center and, thus their object function is of the form

$$f(x) = f_r(r), \quad r = |x|. \quad (2.27)$$

This important situation arises in practical applications such as reconstruction of the structure of optical fibers or large molecules or buried pipes, but also provides a paradigm for both analytical and computer simulation study of DT reconstruction algorithms. Both the filtered backpropagation algorithm (A.13) and the nonlinear models (2.24) attain a simpler form that makes use of one-dimensional Fourier–Bessel transforms.

Indeed, let

$$\tilde{f}_r(|K|) = 2\pi \int_0^\infty dr r J_0(|K|r) f_r(r). \quad (2.28)$$

Then, the Fourier transform of the object function f becomes

$$\tilde{f}(K) = \tilde{f}_r(|K|) \quad (2.29)$$

and the generalized projection-slice theorem (A.10) (see Fig. 2) reads⁴

$$\begin{aligned} \widetilde{P_\theta f}(p; s_0) &= \begin{cases} \frac{k}{m} e^{i(m-k)s_0} \tilde{f}_r \left[\sqrt{2k(k-m)} \right], & \text{if } |p| \leq k \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (2.30)$$

where the generalized projections $P_\theta f$ are independent of the view angle θ . The filtered backpropagation algorithm is now put in the form

$$\hat{f}(x) = \frac{1}{2\pi} \int_0^k dp p e^{i(k-m)s_0} J_0 \left[r \sqrt{2k(k-m)} \right] \widetilde{P_\theta f}(p; s_0), \quad r = |x| \quad (2.31)$$

where the integration with respect to view angle θ has been eliminated. Finally, the nonlinear models (2.24) can be computed in a similar manner. For example, the first Born term is given by

$$\tilde{\psi}_1(pu + s_0 v) = \begin{cases} \frac{ik^2}{m} e^{im s_0} \tilde{f}_r \left[\sqrt{2k(k-m)} \right], & \text{if } |p| \leq k \\ \frac{ik^2}{m} e^{im s_0} \tilde{f}_r(|p|\sqrt{2}), & \text{if } |p| > k \end{cases} \quad (2.32)$$

and allows the computation of the first term W in the Rytov series. Similarly derivable, yet more complicated, expressions

⁴More generally,

$$\tilde{W}_1(pu + s_0 v) = \begin{cases} \frac{k}{m} e^{i(m-k)s_0} \tilde{f}_r \left[\sqrt{2k(k-m)} \right], & \text{if } |p| \leq k \\ \frac{k}{m} e^{i(m-k)s_0} \tilde{f}_r(|p|\sqrt{2}), & \text{if } |p| > k \end{cases}$$

which includes evanescent wavemodes in the wavefield data.

for terms in the Rytov series of order higher than the first can also be derived.

III. INVERSION OF NONLINEAR DATA-GENERATING MODELS FOR DT

A. Theory

Define the inhomogeneous Volterra operator of order N (where the possibility of $N \rightarrow \infty$ is not excluded)

$$\mathcal{W}^N = \mathcal{W}_1 + \mathcal{W}_2 + \cdots + \mathcal{W}_N: \mathcal{F}_L \rightarrow \mathcal{P} \quad (3.1)$$

where

$$(\mathcal{W}^N f_L)(t_0, \theta) = \frac{e^{-iks_0}}{ik} (W_1^L + W_2^L + \cdots + W_N^L)(t_0 u + s_0 v, \theta) \quad (3.2)$$

and $W_n^L(t_0 u + s_0 v, \theta)$, $n = 1, 2, \dots$, is the n th term of the Rytov series to the scattered field along the measurement line $x_0 = t_0 u + s_0 v$, lowpass-filtered with respect to t_0 down to the frequency interval $[-k, k]$. The existence of a local inverse of the operator \mathcal{W}^N and its representation by a convergent series is guaranteed by the existence of the inverse $(\mathcal{B}|\mathcal{P})$ of the linear term $(\mathcal{W}_1|\mathcal{F}_L)$ [71], [46].⁵ Specifically, there exists a region in the data space \mathcal{P} , on which the operator

$$\mathcal{B}_1 + \mathcal{B}_2 + \cdots + \mathcal{B}_n + \cdots \quad (3.3)$$

is the inverse of \mathcal{W}^N , where \mathcal{B}_n is a homogeneous Volterra operator of order n in the data.

More specifically,

$$\mathcal{B}_1 = \mathcal{B} \quad (3.4)$$

$$\mathcal{B}_2 = -\mathcal{B}\mathcal{W}_2\mathcal{B} \quad (3.5)$$

$$\mathcal{B}_3 = \mathcal{B}\mathcal{W}_2\mathcal{B}\mathcal{W}_2\mathcal{B} - \mathcal{B}\mathcal{W}_3\mathcal{B} \quad (3.6)$$

\vdots

It is observed that the terms $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3$, etc., are *universal* in the sense that they do not depend on the order N of the forward model \mathcal{W}^N . Thus the operator series in (3.3) can be truncated to desired order n and return DT inversion algorithms of order n in the data. For example, the usual filtered backpropagation algorithm provides a universal DT inversion algorithm of first order (linear) in the data. A second observation is that if the data are sufficiently modeled by the first Rytov term, i.e., $\delta W = W_1$ and $W_2 = W_3 = \cdots = 0$, then each higher order term in series (3.3) returns an identically zero contribution to image reconstruction, as close examination of the expressions in (3.4)-(3.6) reveals. This fact is also verified in the simulations of this section. A final observation is that the resulting algorithms are effectively of the form of *nonlinear* data filtering followed by backpropagation. Indeed, the series (3.3) can be written as

$$\mathcal{B}(\mathcal{I} - \mathcal{W}_2\mathcal{B} + \mathcal{W}_2\mathcal{B}\mathcal{W}_2\mathcal{B} - \mathcal{W}_3\mathcal{B} - \cdots) \quad (3.7)$$

⁵In [46], a detailed statement of a general theorem on inversion of Volterra systems is given, which includes a statement of an estimate of the range of existence and convergence of the inverse series.

where $\mathcal{I}: \mathcal{P} \rightarrow \mathcal{P}$ is the identity operator. Effectively, the term in brackets in (3.7) is a higher order nonlinear Volterra filter applied on the data and, thus the series (3.3) gives rise to an entire class of *nonlinear filtered backpropagation* algorithms for inversion of DT data.

B. Implementation of the Inversion Algorithms

The implementation of the proposed nonlinear diffraction tomographic reconstruction algorithms consists of several steps and requires the implementation of corresponding algorithmic modules. More specifically, distinct modules to be used are as follows.

- 1) An estimator of the complex phase (complex log) of the measured total wavefield and a phase unwrapper. The estimation of the complex phase of the total wavefield from the measurements may require the use of a proper phase unwrapping algorithm when noise is present in the data. The problem is not serious in the fully three-dimensional imaging case and, in fact, becomes increasingly less serious in higher dimensions. The best practical approach probably consists of two precautionary measures: i) to exploit the fact that the wave signal is bandlimited with zero Fourier content outside the frequency interval $[-k, k]$ and lowpass-filter the measurements accordingly to reduce the noise level and ii) to increase the sampling rate beyond the corresponding Nyquist rate. In previous (independent) experimentation with the theories of linearized DT, it was possible to image two-dimensional (2-D) objects, where the phase wrapping problem is worse, from scattered wavefield measurements in a prototype ultrasonic scanner, a prototype optical microscope, and several underground imaging devices (acoustic and ground penetrating radar), as discussed in Section I and references therein. If a better phase unwrapping approach is needed in some applications, the one we propose is based on modeling the phase with a low degree approximating polynomial, which is possible from the very requirement of phase continuity, and estimating the polynomial coefficients from the measurements.
- 2) An implementer of the two- or three-dimensional filtered backpropagation or equivalent linear inversion algorithm, which can be based on use of fast Fourier transform algorithms and, today, is considered a routine procedure. Even though the filtered backpropagation algorithm is particularly robust to high noise levels (see, for example, [27, the reconstruction in Fig. 13]), it is possible that regularization of the linear algorithm may be required in certain instances (e.g., wave probing only from a very limited number of directions). To counteract the effects of noise, a number of related publications exist which include [85], [11], [60]. In [60], more specifically, an infinite class of regularized inversion algorithms are proposed and studied, which propagate noise differently.
- 3) Born term calculators, that is, algorithms to compute Born terms of a given two- or three-dimensional input function using an FFT-based approach as discussed in [88], [89]. No regularization of these algorithms is required because

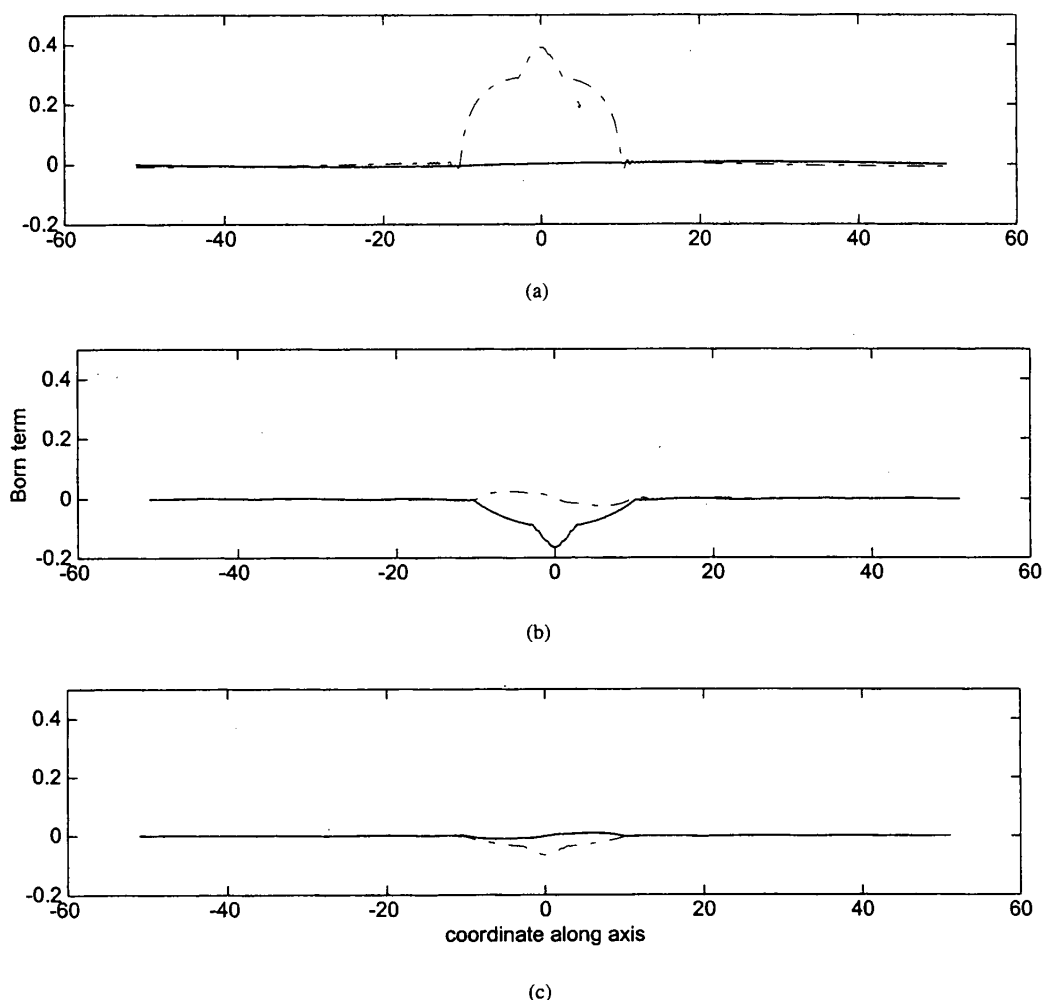


Fig. 3. Scattered fields in the computer simulation. (a) The first Born term ψ_1 . (b) The second Born term ψ_2 . (c) The third Born term ψ_3 . The real (imaginary) part of the fields is plotted in solid (dashed) line.

the corresponding operators are bounded and therefore continuous.

- 4) Implementers of pointwise Born to Rytov transformations, as in (2.16)–(2.19).
- 5) Combiners (adders) of the output of the above modules.

Once the above modules become available, the nonlinear diffraction tomographic reconstruction algorithms can be efficiently implemented in a scheme that successively adds higher order corrections to reduce the bias in the solution returned by a lower order inversion algorithm with repeated utilization of filtered backpropagation, Born term computation, Born to Rytov transformation, and addition as in (3.4)–(3.6).

C. Computer Simulation

The proposed nonlinear diffraction tomographic reconstruction algorithms were implemented and studied in a limited computer simulation, following the steps outlined in the previous section. For simplicity, the scattering object consisted of a circular core and three concentric circular coatings and constitutes a realistic model for cylindrical objects such as optical fibers,

large molecules, or buried pipes. Specifically, the object function was set to

$$f(x) = \begin{cases} 0.0025, & \text{if } |x| \leq 0.75 \\ 0.002, & \text{if } 0.75 \leq |x| \leq 1.8 \\ 0.0015, & \text{if } 1.8 \leq |x| \leq 2.6 \\ 0.001, & \text{if } 2.6 \leq |x| \leq 10 \\ 0, & \text{else.} \end{cases} \quad (3.8)$$

The wavenumber was equal to $k = 6\pi$, corresponding to a wavelength $\lambda = \frac{1}{3}$ and the measurement distance was set to $s_0 = 0$. This is a fairly big object, sixty wavelengths in diameter, for which the Born series converges slowly. The sampling rate was set to 0.04, which corresponds to approximately eight samples per wavelength and thus equals about four times the Nyquist rate. Thus the sample density is high enough to provide good numerical approximations to the continuous space signals and algorithms considered in here. Unfortunately, the exact scattered wave solution for the given scatterer is not known in analytic form. However, if the Born or Rytov series converge, their limit will be the exact solution. Fig. 3 shows plots of the

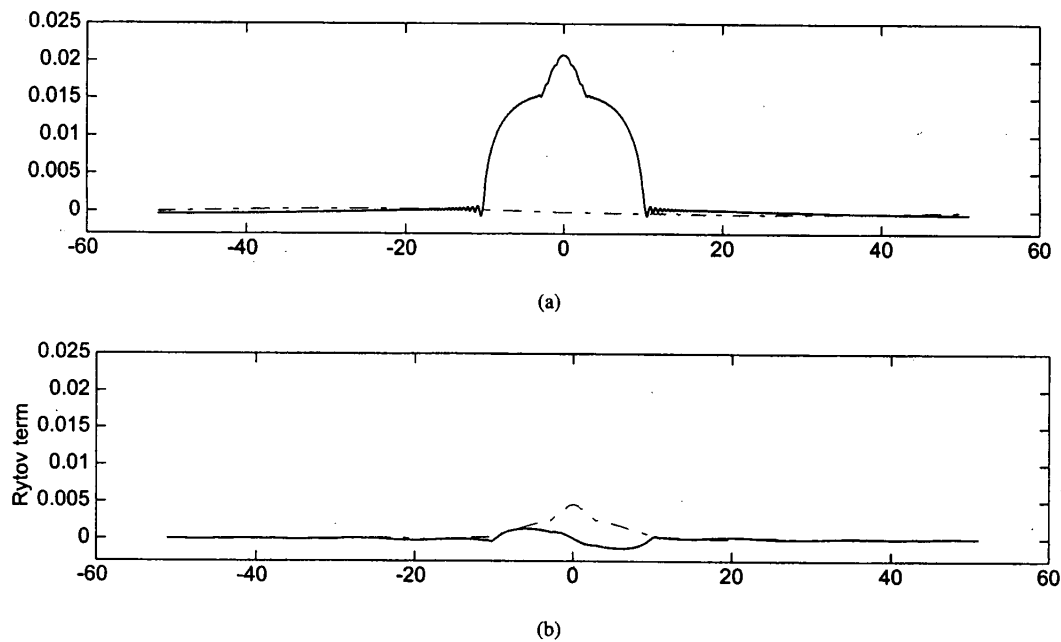


Fig. 4. Complex phase of total fields in the computer simulation. (a) The first Rytov term W_1 . (b) The second Rytov term W_2 . (c) The third Rytov term W_3 . The real (imaginary) part of the complex phase is plotted in solid (dashed) line.

first three Born terms, while plots of the first three Rytov terms are shown in Fig. 4. Clearly, a geometric similarity exists between Born and Rytov terms which is due to the transformations (2.16)–(2.19) and the relative simplicity of the object. We note, however, that the Rytov terms need to be nonlinearly distorted (via exponentiation) to yield the total wave and this may, in turn, distort the simple geometric correspondence.

The third Born term is approximately only six times smaller in magnitude than the first Born term ($\frac{\max |\psi_1|}{\max |\psi_3|} \approx 6$). By comparison, the third Rytov term is approximately thirty times smaller in magnitude than the first Rytov term ($\frac{\max |W_1|}{\max |W_3|} \approx 30$). Thus the Rytov series provides more accurate modeling of the complex phase of the total field than the Born series provides for the scattered field, a fact that is consistent with earlier investigations into the domains of validity of the Born and Rytov approximations [40], [39], [58], [95], [77], [78], [80], [47], [51].

Fig. 5 shows the object function reconstructions returned by the filtered backpropagation and a second-order algorithm based on the Born series [89], from first- (Fig. 5(a)), second- (Fig. 5(b)), and third-order (Fig. 5(c)) Born approximations. Clearly, the inversion procedure based on the Born series diverges.⁶ Fig. 6 shows the object function reconstructions returned by the filtered backpropagation and the second-order algorithm based on the Rytov series, from first- (Fig. 6(a)), second- (Fig. 6(b)), and third-order (Fig. 6(c)) Rytov approximations. Clearly, the second-order algorithm returns the same reconstruction as the filtered backpropagation algorithm in the case of data consisting of only the first Rytov term, as theoretically expected. In the other cases, however, the filtered backpropagation algorithm un-

derperforms the second-order algorithm by a significant margin, especially in the area close to the core of the object.

IV. SUMMARY, CONCLUSIONS, AND FUTURE RESEARCH

In this paper, linearized Diffraction Tomography was extended into a higher order (nonlinear) regime via the use of an arbitrarily large (possibly infinite) number of terms from the Rytov series as the data-generating model in scattering experiments. Nonlinear tomographic reconstruction algorithms were developed for inversion of wavefield data and image formation, which attained the form of a series of Volterra operators in which the leading (linear) term was the usual filtered backpropagation (or equivalent) algorithm of linearized DT. A limited computer simulation study was included to illustrate the performance of the algorithms in the case of imaging objects with cylindrical symmetry. From this study, it was found that a second-order algorithm returned essentially the same reconstruction as the usual filtered backpropagation algorithm when applied to first-order Rytov data, but clearly outperformed the filtered backpropagation algorithm when applied to data consisting of the second- or third-order Rytov approximation.

The paper establishes a unified paradigm for development of linear and nonlinear reconstruction algorithms for DT and thus further investigation in its direction is due. Future relevant research may follow the avenues of higher order DT with alternative data measurement configurations and higher order Diffraction Tomography of stochastic objects (including Bayesian inversion algorithms) and their time-domain counterparts. It seems that none of these research avenues are intractable in principle, but may require modification of the procedures presented in the present paper. Other related research directions to be followed in the future include study of the discretization of the proposed continuous space models and

⁶We observe, however, that the filtered backpropagation algorithm returns identical reconstructions when applied to either first-order Born (Fig. 5(a)) or first-order Rytov (Fig. 6(a)) data.

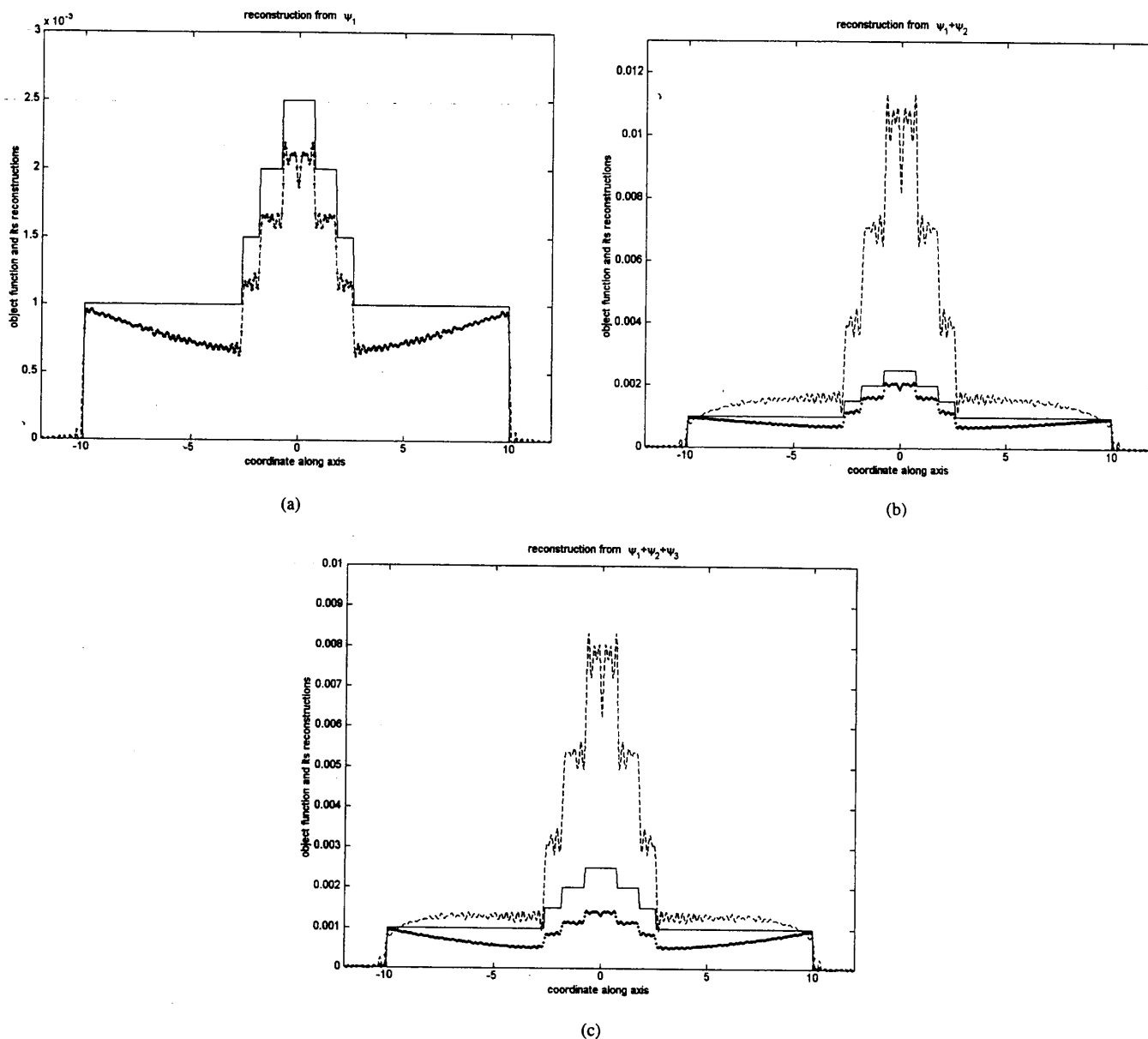


Fig. 5. Original object function (solid line) and reconstructions returned from the filtered backpropagation (dotted line) and the second-order algorithm based on the Born series (dashed line) in the computer simulation. (a) Reconstruction from the first-order Born approximation. (b) Reconstruction from the second-order Born approximation. (c) Reconstruction from the third-order Born approximation.

algorithms and study of the discretization error and the efficient implementation of the discrete space algorithms. Finally, the application of these algorithms on fully three-dimensional real data is a high priority and, in fact, preliminary results of this process for data collected in a testbed that we are developing have been announced at recent conferences [26], [31]. These and other research avenues are currently being explored and the findings will be reported elsewhere.

APPENDIX A THE BORN SERIES

Equation (2.2) implies a nonlinear mapping from the object function f to the scattered field $\psi^{(s)}$, which under certain weak scattering conditions can be expanded into a convergent Liouville–Neumann expansion, known as the Born series. The Born

series is thus a perturbational expansion for the field scattered by an object f , obtained by introducing a smallness parameter ϵ and replacing the object function f in (2.1) and (2.2) by ϵf . The total field $\psi(x; \epsilon)$ inside the scattering region V is assumed an analytic function of the parameter ϵ around the point $\epsilon = 0$, so that the Taylor series expansion around $\epsilon = 0$

$$\psi(x; \epsilon) = \psi_0(x) + \epsilon\psi_1(x) + \epsilon^2\psi_2(x) + \dots \quad (\text{A.1})$$

converges for sufficiently small ϵ . The terms $\epsilon\psi_1(x)$, $\epsilon^2\psi_2(x)$, \dots are the first-, second-, \dots order corrections made to the incident field $\psi_0(x)$ to yield the total field $\psi(x; \epsilon)$. Clearly, this expansion is compatible with the fact that the scattered field

$$\psi^{(s)}(x; \epsilon) = \epsilon\psi_1(x) + \epsilon^2\psi_2(x) + \dots$$

is solely due to the presence of the scatterer with object function ϵf and in the absence of it (i.e., for $\epsilon = 0$) the total field $\psi(x; \epsilon = 0)$ reduces simply to the known incident field $\psi_0(x)$.

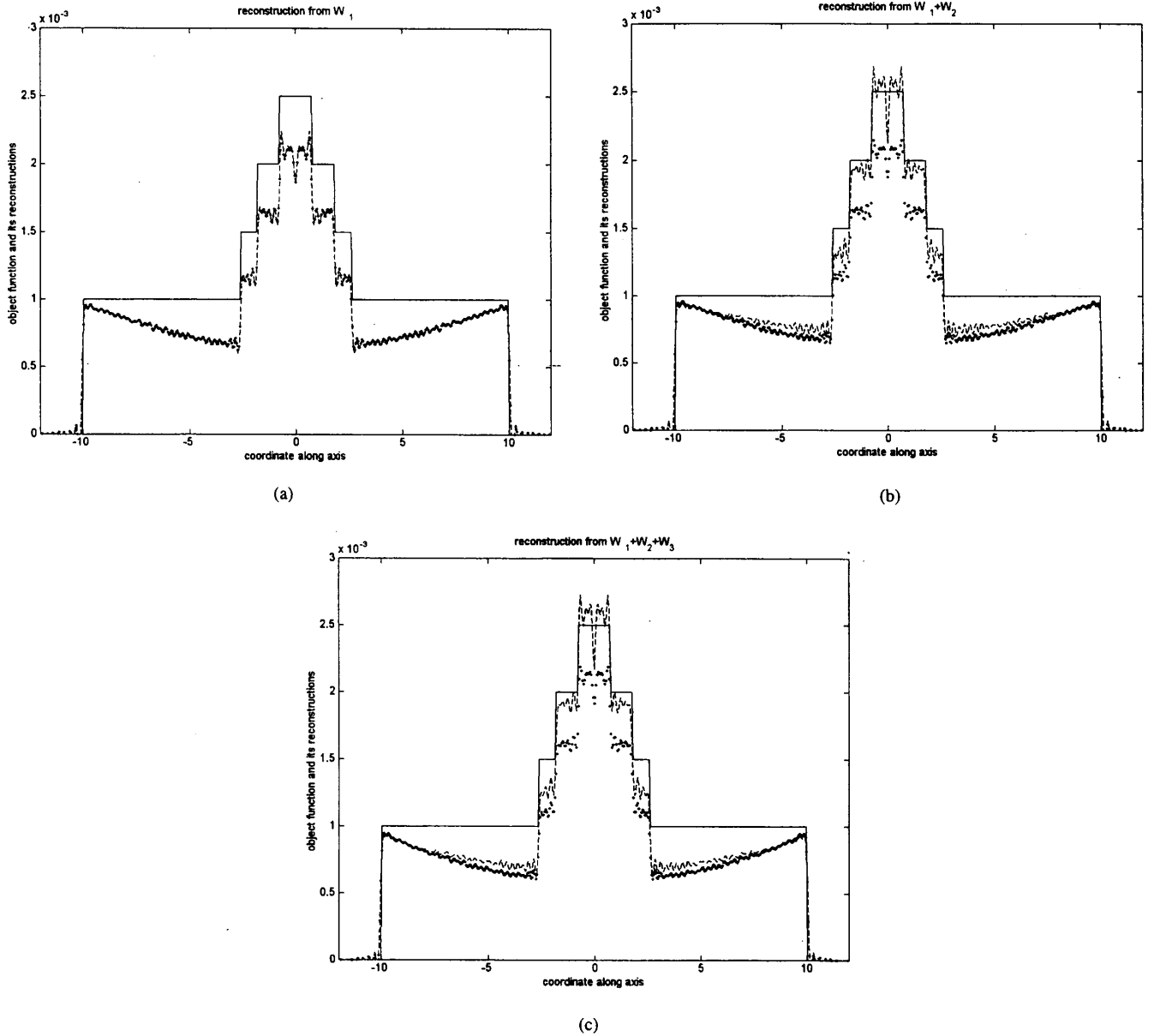


Fig. 6. Original object function (solid line) and reconstructions returned from the filtered backpropagation (dotted line) and the second-order algorithm based on the Rytov series (dashed line) in the computer simulation. (a) Reconstruction from the first-order Rytov approximation. (b) Reconstruction from the second-order Rytov approximation. (c) Reconstruction from the third-order Rytov approximation.

Inserting (A.1) into (2.2) and equating the coefficients of equal powers of ϵ on the two sides of the resulting expression, one obtains

$$\begin{aligned}
 \psi_1(x_0) &= -2k^2 \int_{\mathcal{V}} dx' G(x_0 - x') f(x') \psi_0(x') \\
 \psi_2(x_0) &= -2k^2 \int_{\mathcal{V}} dx' G(x_0 - x') f(x') \psi_1(x') \\
 &\vdots \\
 \psi_n(x_0) &= -2k^2 \int_{\mathcal{V}} dx' G(x_0 - x') f(x') \psi_{n-1}(x') \\
 &\vdots
 \end{aligned} \tag{A.2}$$

Clearly, the Born term of order n can be computed from the Born term of order $n-1$, $n = 1, 2, 3, \dots$, by replacing the induced source $f\psi$ with the product $f\psi_{n-1}$. After computation of the Born terms up to some desired order N , the total field ψ

resulting from the interaction of the incident wave ψ_0 with the scatterer f is computed as $\psi(x_0) = \psi_0(x_0) + \psi^{(s)}(x_0)$ with the scattered field approximated as

$$\psi^{(s)}(x_0) \approx \psi_1(x_0) + \psi_2(x_0) + \dots + \psi_N(x_0). \tag{A.3}$$

Of course, the assumption is made that the expansion (A.1) is valid for $\epsilon = 1$.

The first-order Born approximation, simply referred to as the Born approximation [56], has been extensively employed in quantum-mechanical scattering theory [56], [32]. From (A.2), the Born approximation estimates the scattered field $\psi^{(s)}$ by substituting the incident field ψ_0 for the total field ψ in the integrand in the right-hand side of (2.2) and, therefore, requires that i) $|\psi^{(s)}| \ll |\psi_0|$ inside the scatterer support \mathcal{V} and ii) the volume of scatterer support \mathcal{V} itself be small [40].

It is seen that the Born approximation is quite restrictive, since it requires both the size and the strength of the scattering

object to be “small.” Additionally, it is a low-frequency approximation, in that it requires the frequency of the incident wave to be low. The Born approximation has been extensively used for predictions in direct [81], [7], [72] and inverse [18] scattering theory. Its domain of validity has been extensively explored in the literature [40], [39], [58], [95], [77], [78], [47] and compared with experimental results [80], [51].

APPENDIX B MATHEMATICAL FORMULATION OF LINEARIZED DIFFRACTION TOMOGRAPHY

The first Born approximation for the field scattered from the object when illuminated with a plane wave⁷ $\psi_\theta(x) = e^{ik\langle v, x \rangle} = e^{iks}$ propagating in the direction of the unit vector v is computed from the first of (A.2) as

$$\psi_\theta^B(x_0) = -2k^2 \int_{\mathcal{V}} dx' G(x_0 - x') e^{ik\langle v, x' \rangle} f(x'). \quad (\text{A.4})$$

At this point, the Weyl expansion [53] of the Green function G is introduced

$$\begin{aligned} G(t'u + s'v) &= -\frac{i}{4\pi} \int_{-\infty}^{\infty} dp e^{ipt'} \frac{e^{im|s'|}}{m} \\ &\approx -\frac{i}{4\pi} \int_{-k}^k dp e^{ipt'} \frac{e^{im|s'|}}{m} \\ &= G^L(t'u + s'v), \end{aligned} \quad (\text{A.5})$$

where

$$m = \begin{cases} \sqrt{k^2 - p^2}, & \text{if } |p| \leq k \\ i\sqrt{p^2 - k^2}, & \text{if } |p| > k. \end{cases}$$

The expansion in the right-hand side of (A.4) contains both homogeneous (corresponding to $|p| \leq k$) and evanescent (corresponding to $|p| > k$) planewave terms and can be truncated to its homogeneous spectrum only, as indicated by the approximate equality sign in (App. 5), if the corresponding observation point x_0 is taken at least several wavelengths away from the scattering region \mathcal{V} .

Substitution of (A.5) into (A.4) gives

$$\begin{aligned} \psi_\theta^B(t_0 u + s_0 v) &\approx \frac{2ik^2}{4\pi} \int_{\mathcal{V}} dt' ds' \int_{-k}^k dp e^{ipt_0 - t'} \\ &\quad \cdot \frac{e^{im(s_0 - s')}}{m} f(t'u + s'v) e^{iks'} \\ &= \frac{ik^2}{2\pi} \int_{-k}^k dp e^{ipt_0} \frac{e^{ims_0}}{m} \int_{\mathcal{V}} dt' ds' \\ &\quad \cdot e^{-ipt'} e^{-i(m-k)s'} f(t'u + s'v) \\ &= \frac{ik^2}{2\pi} \int_{-k}^k dp e^{ipt_0} \frac{e^{ims_0}}{m} \tilde{f}[pu + (m-k)v] \end{aligned} \quad (\text{A.6})$$

⁷A subscript θ has been introduced to index the incident, total, and scattered waves corresponding to the different tomographic experiments using the classical scan configuration of DT.

in which \tilde{f} denotes the two-dimensional Fourier transform of f . Equation (A.6) defines the “propagation transform” $P_\theta f$ of linearized DT [17]

$$P_\theta f(t_0; s_0) = \begin{cases} \frac{e^{-iks_0}}{ik} \psi_\theta^{(s)}(x_0) & \text{within the Born approximation} \\ \delta W_\theta(x_0) & \text{within the Rytov approximation} \end{cases} \quad (\text{A.7})$$

which relates to the object function f through the convolutional relation [21], [23]

$$P_\theta f(t_0; s_0) = \int_{\mathcal{V}} \int_{\mathcal{V}} dt' ds' \Gamma_\theta(t_0 - t', s_0 - s') f(t'u + s'v) \quad (\text{A.8})$$

for $t_0 \in R^1$, s_0 : fixed, $\theta \in \Theta \subset [0, 2\pi)$, and kernel Γ_θ expressed in spectral form as

$$\Gamma_\theta(t', s') = \frac{1}{2\pi} \int_{-k}^k dp \frac{k}{m} e^{[ipt' + (m-k)s']}. \quad (\text{A.9})$$

Equations (A.8) and (A.9) result in a simple relation when considered in the Fourier domain. This relation, following upon substitution of (A.9) into (A.8), reads

$$\widetilde{P_\theta f}(p; s_0) = \begin{cases} \frac{k}{m} e^{i(m-k)s_0} \tilde{f}[pu + (m-k)v], & \text{if } |p| \leq k \\ 0, & \text{otherwise} \end{cases} \quad (\text{A.10})$$

and is known as the *generalized projection-slice theorem* [102], [21], [23]. Fig. 2 is an illustration of the generalized projection-slice theorem, which in words, implies that the one-dimensional Fourier transform $\widetilde{P_\theta f}$ of the propagation transform is proportional to a semicircular slice through the two-dimensional Fourier transform \tilde{f} of the object function itself.⁸

The above formulation of the inverse problem of DT is compatible with the usual formulation of the inverse problem of conventional Computerized Tomography (CT) and reduces to it in the limiting case of probing radiation of very short wavelength. Indeed,

$$k \rightarrow \infty, \quad \left(\frac{k}{m}\right) \rightarrow 1, \quad (m-k) \rightarrow 0, \quad \text{as } \lambda \rightarrow 0$$

and the kernel $\Gamma_\theta(t', s')$ reduces to

$$\Gamma_\theta(t', s') \approx \frac{1}{2\pi} \int_{-\infty}^{\infty} dp e^{ipt'} = \delta(t'), \quad \text{as } \lambda \rightarrow 0. \quad (\text{A.11})$$

Thus the propagation transform becomes the usual Radon transform

$$\begin{aligned} P_\theta f(t_0) &= \int_{\mathcal{V}} \int_{\mathcal{V}} dt' ds' \delta(t_0 - t') f(t'u + s'v) \\ &= \int_{-\infty}^{\infty} ds' f(t_0 u + s'v). \end{aligned} \quad (\text{A.12})$$

The generalized projection-slice theorem (A.10) has formed the basis for most treatments to the inverse problem of DT, namely, Fourier interpolation [36], [55], [54], [59] and filtered backpropagation [17], [59], [23], while (A.8) formed the basis for the development of iterative algorithms for DT [20], [22],

⁸The remainder of the circular slice is associated with the reflected field, i.e., the scattered field along the measurement line $s = -s_0$.

[44], [43]. Of particular importance to this paper is the filtered backpropagation algorithm [17], [22], which provides a reconstruction \hat{f} that corresponds to a lowpass-filtered version of the unknown object function f . In particular, the filtered backpropagation algorithm estimates \hat{f} as [17]

$$\hat{f}(x) = \frac{1}{2} \frac{1}{(2\pi)^2} \int_0^{2\pi} d\theta \cdot \int_{-k}^k dp |p| e^{ipt} e^{i(m-k)(s-s_0)} \widetilde{P_\theta} f(p; s_0) \quad (\text{A.13})$$

where $x = tu + sv$. It is readily shown that the estimate \hat{f} has the same spatial frequency content as the unknown f over a disk D in Fourier space of radius $k\sqrt{2}$ and zero frequency content outside that disk. Since its first appearance [17], the filtered backpropagation algorithm has been modified and extended to other data acquisition geometries [19] and tested on computer simulated [16], [59] and experimental data [79], [84].

APPENDIX C

ELEMENTS OF THE VOLTERRA THEORY OF NONLINEAR SYSTEMS

A homogeneous Volterra operator (VO) of order $k \geq 1$ is a nonlinear system represented by a functional relation of the form

$$y(\cdot) = \int \cdots \int d\tau_1 \cdots d\tau_k h_k(\cdot; \tau_1, \cdots, \tau_k) x(\tau_1) \cdots x(\tau_k) \quad (\text{A.14})$$

between its input $x(\cdot)$ and the corresponding output $y(\cdot)$. The $(k+1)$ -dimensional function $h_k(\cdot; \tau_1, \cdots, \tau_k)$ is the *kernel* of the homogeneous k th-order VO and, for uniqueness, is assumed symmetric, i.e., its value does not change under a permutation of its arguments. Clearly, a homogeneous VO of order $k = 1$ is a linear system and its kernel $h_1(\cdot; \tau_1)$ is also its impulse response. A homogeneous VO of order k as in (A.14) is a homogeneous polynomial system of order k in the sense that the system response to input $cx(\cdot)$ is $c^k y(\cdot)$ for arbitrary constant c .

A general (inhomogeneous) VO of order N is a nonlinear system consisting of the summation of homogeneous VO's of the form of (A.14) of order less than or equal to N , i.e., a nonlinear system represented by a functional relation of the form

$$y(\cdot) = h_0 + \sum_{k=1}^N \int \cdots \int d\tau_1 \cdots d\tau_k h_k(\cdot; \tau_1, \cdots, \tau_k) x(\tau_1) \cdots x(\tau_k) \quad (\text{A.15})$$

where h_0 is a constant. Clearly, an inhomogeneous VO of order N is a polynomial system of order N in the sense that its response to input $cx(\cdot)$ is a polynomial in c of order N for arbitrary constant c . Equation (A.15) can be generalized further by letting the order N of the inhomogeneous VO become infinite, assuming convergence of the resulting functional series

$$y(\cdot) = h_0 + \sum_{k=1}^{\infty} \int \cdots \int d\tau_1 \cdots d\tau_k h_k(\cdot; \tau_1, \cdots, \tau_k) x(\tau_1) \cdots x(\tau_k). \quad (\text{A.16})$$

In practice, both the number of terms in (A.16) as well as the support of the Volterra kernels have to be truncated. Issues of approximation of nonlinear systems with infinite and truncated Volterra systems have been extensively addressed in the literature [8], [34], [46], [30], [69], [68], [29], [70]. Issues of inverting a given Volterra system are discussed in great length in [71] and in a concise and rigorous manner in [46].

REFERENCES

- [1] M. Azimi and A. C. Kak, "Distortion in diffraction tomography caused by multiple scattering," *IEEE Trans. Med. Imag.*, vol. MI-2, pp. 176-195, 1983.
- [2] S. Barkeshli and R. G. Lautzenheizer, "An iterative method for inverse scattering problems based on an exact gradient search," *Radio Sci.*, vol. 29, pp. 1119-1130, 1994.
- [3] A. Brancaccio and R. Pierri, "Information content of Born scattered fields: Results in the circular cylindrical case," *J. Opt. Soc. Amer. A*, vol. 15, pp. 1909-1917, 1998.
- [4] S. Caorsi, C. L. Gragnani, S. Medicina, M. Pastorino, and A. Pinto, "A Gibbs random fields-based active electromagnetic method for noninvasive diagnostics in biomedical applications," *Radio Sci.*, vol. 30, pp. 291-301, 1995.
- [5] S. Caorsi, G. L. Gragnani, and M. Pastorino, "Two-dimensional microwave imaging by a numerical inverse scattering solution," *IEEE Trans. Microwave Theory Tech.*, vol. 38, pp. 981-989, 1990.
- [6] K. Chadon, D. Colton, L. Paivarinta, and W. Rundell, *An Introduction to Inverse Scattering and Inverse Spectral Problems*. Philadelphia, PA: SIAM, 1997.
- [7] L. A. Chernov, *Wave Propagation in a Random Medium*. New York: Dover, 1967.
- [8] G. S. Christensen, "On the convergence of Volterra series," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 736-737, 1968.
- [9] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering*, 2 ed. Berlin, Germany: Springer-Verlag, 1998.
- [10] D. Colton and P. Monk, "The inverse scattering problem for acoustic waves in an inhomogeneous medium," in *Inverse Problems in Partial Differential Equations*, D. Colton, R. Ewing, and W. Rundell, Eds. Philadelphia, PA: SIAM, 1990.
- [11] R. W. Deming and A. J. Devaney, "Diffraction tomography for monostatic ground penetrating radar imaging," *Inverse Probl.*, vol. 13, pp. 29-45, 1997.
- [12] P. M. Van der Berg and R. E. Kleinman, "A total variation enhanced modified gradient algorithm for profile reconstruction," *Inverse Probl.*, vol. 11, pp. L5-L10, 1995.
- [13] —, "A contrast source inversion method," *Inverse Probl.*, vol. 13, pp. 1607-1620, 1998.
- [14] A. J. Devaney, "Nonuniqueness in the inverse scattering problem," *J. Math. Phys.*, vol. 19, p. 1526, 1978.
- [15] —, "Inverse scattering within the Rytov approximation," *Opt. Lett.*, vol. 6, p. 374, 1981.
- [16] —, "A computer simulation study of diffraction tomography," *IEEE Trans. Biomed. Eng.*, vol. BME-30, p. 377, 1982.
- [17] —, "A filtered backpropagation algorithm for diffraction tomography," *Ultrason. Imag.*, vol. 4, p. 336, 1982.
- [18] —, "Inversion formula for inverse scattering within the Born approximation," *Opt. Lett.*, vol. 7, p. 111, 1982.
- [19] —, "Geophysical diffraction tomography," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-22, p. 3, 1984.
- [20] —, "An iterative ART algorithm for geophysical diffraction tomography," in *IGARSS'85 Dig.*, K. R. Carver, Ed. New York: IEEE, 1985.
- [21] —, "Mathematical topics in diffraction tomography," Final Project, Rep. NSF, Grant DMS-8460595, 1985.
- [22] —, "Reconstructive tomography with diffracting wavefields," *Inverse Probl.*, vol. 2, p. 161, 1986.
- [23] —, "The limited-view problem in diffraction tomography," *Inverse Probl.*, vol. 5, p. 501, 1989.
- [24] —, "Approximate scattering models in inverse scattering: Past, present, and future," in *Directions in Electromagnetic Modeling*, L. B. Felsen and H. Bertoni, Eds. New York: Plenum, 1990.
- [25] —, "Elastic wave inverse scattering," in *Elastic Waves and Ultrasonic Nondestructive Evaluation*, S. K. Datta, J. D. Achenbach, and Y. S. Rajapakse, Eds. New York: Elsevier, 1990.

- [26] —, "A new three-dimensional optical microscope," in *Annu. Meet. Optical Society of America*, Santa Clara, CA, Sept. 1999.
- [27] A. J. Devaney and G. A. Tsihrintzis, "Maximum likelihood estimation of object location in diffraction tomography," *IEEE Trans. Signal Processing*, vol. 39, no. 3, p. 672, 1991.
- [28] A. J. Devaney and E. Wolf, "A new perturbation expansion for inverse scattering from three-dimensional finite-range potentials," *Phys. Lett.*, vol. 89A, no. 6, pp. 269–272, 1982.
- [29] R. J. P. De Figueiredo, "A generalized Fock space framework for nonlinear system and signal analysis," *IEEE Trans. Circuits Syst.*, vol. CAS-30, pp. 637–647, 1983.
- [30] R. J. P. De Figueiredo and T. A. W. Dwyer, "A best approximation framework and implementation for simulation of large-scale nonlinear systems," *IEEE Trans. Circuits Syst.*, vol. CAS-27, pp. 1005–1014, 1980.
- [31] Y. Glina, G. A. Tsihrintzis, C. M. Warner, D. O. Hogenboom, and C. A. DiMarzio, "On the use of the optical quadrature method in tomographic microscopy," in *BIOS'99*, vol. 3605, San Jose, CA, 1999.
- [32] K. Gottfried, *Quantum Mechanics, Volume I. Fundamentals*. Reading, MA: Addison-Wesley, 1989.
- [33] J. F. Greenleaf, "Computerized tomography with ultrasound," *Proc. IEEE*, vol. 71, p. 330, 1983.
- [34] A. Halme and J. Orava, "Generalized polynomial operators for nonlinear system analysis," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 226–228, 1972.
- [35] J. M. Harris, "Diffraction tomography with arrays of discrete sources and receivers," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-25, no. 4, p. 448, 1987.
- [36] K. Iwata and R. Nagata, "Calculation of refractive index distribution from interferograms using the Born and Rytov's approximation," *Japan J. Appl. Phys.*, vol. 14, p. 379, 1974.
- [37] N. Joachimowicz, C. Pichot, and J.-P. Hugonin, "Inverse scattering: An iterative numerical method for electromagnetic imaging," *IEEE Trans. Antennas Propagat.*, vol. 39, pp. 1742–1752, 1991.
- [38] M. Kaveh, M. Soumekh, and J. F. Greenleaf, "Signal processing for diffraction tomography," *IEEE Trans. Sonics Ultrason.*, vol. SU-31, p. 230, 1984.
- [39] M. Kaveh, M. Soumekh, and R. K. Mueller, "A comparison of Born and Rytov approximation," in *Acoustical Imaging, Vol. 11*, H. Lee and G. Wade, Eds. New York: Plenum, 1981, p. 325.
- [40] J. B. Keller, "Accuracy and validity of the Born and Rytov approximation," *J. Opt. Soc. Amer.*, vol. 59, p. 1003, 1969.
- [41] R. E. Kleinman and P. M. Van der Berg, "A modified gradient method for two-dimensional problems in tomography," *J. Comput. Appl. Math.*, vol. 42, pp. 17–35, 1992.
- [42] —, "An extended range modified gradient technique for profile inversion," *Radio Sci.*, vol. 28, pp. 877–884, 1993.
- [43] K. T. Ladas, "Iterative methods in diffraction tomography," Ph.D. dissertation, Northeastern Univ., Boston, MA, 1991.
- [44] K. T. Ladas and A. J. Devaney, "Generalized ART algorithm for diffraction tomography," *Inverse Probl.*, vol. 7, p. 109, 1991.
- [45] —, "Iterative methods in geophysical diffraction tomography," *Inverse Probl.*, vol. 8, pp. 119–132, 1992.
- [46] C. Lesiak and A. J. Krener, "The existence and uniqueness of Volterra series for nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 1090–1095, 1978.
- [47] F. C. Lin and M. A. Fiddy, "On the issue of the Born–Rytov controversy: I. Comparing analytical and approximate expressions for the one-dimensional deterministic case," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1102–1110, 1992.
- [48] H. Lipson and W. Cochran, *The Determination of Crystal Structures*. Ithaca, NY: Cornell Univ. Press, 1966.
- [49] Z.-Q. Lu, "JKM perturbation theory, relaxation perturbation theory, and their applications to inverse scattering: Theory and reconstruction algorithms," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. UFFC-33, pp. 722–730, 1986.
- [50] Z.-Q. Lu and Y.-Y. Zhang, "Acoustical tomography based on the second-order transform perturbation approximation," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 43, pp. 296–302, 1996.
- [51] M. H. Maleki, A. J. Devaney, and A. Schatzberg, "Tomographic reconstruction from optical scattered intensities," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1356–1363, 1992.
- [52] J. E. Molyneux and A. J. Witten, "Diffraction tomographic imaging in a monostatic measurement geometry," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 507–511, 1993.
- [53] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics, Parts I, II*. New York: McGraw-Hill, 1953.
- [54] R. K. Mueller, M. Kaveh, and R. D. Inverson, "A new approach to acoustic tomography using diffraction techniques," in *Acoustical Imaging*, A. F. Metherell, Ed. New York: Plenum, 1980.
- [55] R. K. Mueller, M. Kaveh, and G. Wade, "Reconstructive tomography and applications to ultrasonics," *Proc. IEEE*, vol. 67, p. 567, 1979.
- [56] R. G. Newton, *Scattering Theory of Waves and Particles*. Berlin, Germany: Springer-Verlag, 1982.
- [57] M. K. Nguyen and A. Mohammad-Djafari, "Bayesian approach with maximum entropy priors in image reconstruction from microwave scattered fields data," *IEEE Trans. Med. Imag.*, vol. 13, pp. 254–261, 1994.
- [58] M. L. Oristaglio, "Accuracy of the Born and Rytov approximations for the reflection and refraction at a plane interface," *J. Opt. Soc. Amer.*, vol. 2, p. 1987, 1985.
- [59] S. X. Pan and A. C. Kak, "A computational study of reconstruction algorithms for diffraction tomography," *IEEE Trans. Acous., Speech, Signal Processing*, vol. ASSP-31, p. 1262, 1982.
- [60] X. Pan, "Unified reconstruction theory for diffraction tomography, with consideration of noise control," *J. Opt. Soc. Amer. A*, vol. 15, pp. 2312–2326, 1998.
- [61] R. Pierri and A. Brancaccio, "Imaging of a rotationally symmetric dielectric cylinder by a quadratic approach," *J. Opt. Soc. Amer. A*, vol. 14, pp. 2777–2785, 1997.
- [62] R. B. Pratt and M. H. Worthington, "The application of diffraction tomography to cross-hole seismic data," *Geophys.*, vol. 53, p. 1284, 1988.
- [63] R. T. Prosser, "Formal solutions of inverse scattering problems," *J. Math. Phys.*, vol. 10, pp. 1819–1822, 1969.
- [64] —, "Formal solutions of inverse scattering problems, II," *J. Math. Phys.*, vol. 17, pp. 1775–1779, 1976.
- [65] —, "Formal solutions of inverse scattering problems, III," *J. Math. Phys.*, vol. 21, pp. 2648–2653, 1980.
- [66] D. J. Rossi and A. S. Willsky, "Reconstruction from projections based on detection and estimation of objects, Parts I and II: Performance analysis and robustness analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, p. 886, 1984.
- [67] M. I. Sancer and A. D. Varvatsis, "A comparison of the Born and Rytov methods," *Proc. IEEE*, vol. 58, pp. 140–141, 1970.
- [68] I. W. Sandberg, "The mathematical foundations of associated expansions for mildly nonlinear systems," *IEEE Trans. Circuits Syst.*, vol. CAS-30, pp. 441–455, 1983.
- [69] —, "Volterra-like expansions for solutions of nonlinear integral and nonlinear differential equations," *IEEE Trans. Circuits Syst.*, vol. CAS-30, pp. 68–77, 1983.
- [70] —, "Uniform approximation with doubly finite volterra series," *IEEE Trans. Signal Processing*, vol. 40, pp. 1438–1442, 1992.
- [71] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York, NY: Wiley, 1980.
- [72] R. A. Schmeltzer, "Means, variances, and covariances, for laser beam propagation through a random medium," *Quart. Appl. Math.*, vol. 24, pp. 339–354, 1967.
- [73] M. Slaney, A. C. Kak, and L. E. Larsen, "Limitations of imaging with first-order diffraction tomography," *IEEE Trans. Microw. Theory Tech.*, vol. MTT-32, pp. 860–874, 1984.
- [74] R. Snieder, "The role of the Born approximation in nonlinear inversion," *Inverse Probl.*, vol. 6, pp. 247–266, 1990.
- [75] —, "An extension of Backus–Gilbert theory to nonlinear inverse problems," *Inverse Probl.*, vol. 7, pp. 409–433, 1991.
- [76] —, "The role of nonlinearity in inverse problems," *Inverse Probl.*, vol. 14, pp. 387–404, 1998.
- [77] M. Soumekh, "An improvement to the Rytov approximation in diffraction tomography," *IEEE Trans. Ultrason., Ferroelect., and Freq. Contr.*, vol. UFFC-33, no. 4, p. 394, 1986.
- [78] M. Soumekh and M. Kaveh, "A theoretical study of model approximation errors in diffraction tomography," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. UFFC-33, no. 1, p. 10, 1986.
- [79] N. Sponheim, L.-J. Gelius, I. Johansen, and J. J. Stamnes, "Quantitative results in ultrasonic tomography of large objects using line sources and curved detector arrays," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 38, p. 370, 1991.
- [80] N. Sponheim, I. Johansen, and A. J. Devaney, "Initial testing of a clinical ultrasound mammograph," in *Acoustical Imaging, Vol. 18*, H. Lee and G. Wade, Eds. New York: Plenum, 1990.
- [81] V. I. Tatarskiy, *Wave Propagation in a Turbulent Medium*. New York: McGraw-Hill, 1961.
- [82] G. A. Tsihrintzis and A. J. Devaney, "Estimation of object location from diffraction tomographic intensity data," *IEEE Trans. Signal Processing*, vol. 39, p. 2136, Sept. 1991.

- [83] —, "Maximum likelihood estimation of object location in diffraction tomography, Part II: Strongly scattering objects," *IEEE Trans. Signal Processing*, vol. 39, p. 1466, June 1991.
- [84] —, "Application of a maximum likelihood estimator in an experimental study of ultrasonic diffraction tomography," *IEEE Trans. Med. Imag.*, vol. 12, pp. 545–554, 1993.
- [85] —, "Stochastic diffraction tomography: Theory and computer simulation," *Signal Processing*, vol. 39, pp. 49–64, 1993.
- [86] —, "Stochastic geophysical diffraction tomography," *Int. J. Imag. Syst. Technol.*, vol. 5, pp. 239–242, 1994.
- [87] —, "Maximum likelihood techniques in ultrasonic diffraction tomography," in *Medical Imaging Techniques and Applications*, C. T. Leondes, Ed. Newark, NJ: Gordon and Breach, 1998, vol. 6, pp. 43–126.
- [88] —, "Higher-order (nonlinear) diffraction tomography," in *Proc. CISS'99*, Baltimore, MD, 1999.
- [89] —, "Higher-order (nonlinear) diffraction tomography: Reconstruction algorithms and computer simulation," *IEEE Trans. Image Processing*, vol. 9, Sept. 2000, to be published.
- [90] G. A. Tsihrintzis, A. J. Devaney, and E. Heyman, "Estimation of object location from wideband scattering data," *IEEE Trans. Image Processing*, vol. 8, pp. 996–1001, 1999.
- [91] G. A. Tsihrintzis, P. M. Johansen, and A. J. Devaney, "Buried object detection and location estimation from electromagnetic field measurements," *IEEE Trans. Antennas Propagat.*, vol. 47, pp. 1742–1744, 1999.
- [92] Y. M. Wang and W. C. Chew, "An iterative solution of the two-dimensional electromagnetic inverse scattering problem," *Int. J. Imag. Sci. Syst.*, vol. 1, pp. 100–108, 1989.
- [93] A. B. Weglein, F. A. Gasparotto, P. M. Carvalho, and R. H. Stolt, "An inverse-scattering series method for attenuating multiples in seismic reflection data," *Geophys.*, vol. 62, pp. 1975–1989, 1997.
- [94] A. B. Weglein and K. H. Matson, "Inverse scattering internal multiple attenuation: An analytic example and subevent interpretation," in *SPIE Proceedings, Volume 3453: Mathematical Methods in Geophysical Imaging*: SPIE, 1998, pp. 1008–1017.
- [95] V. H. Weston, "On the convergence of the Rytov approximation for the reduced wave equation," *J. Math. Phys.*, vol. 26, p. 1979, 1985.
- [96] A. Witten, D. Gillette, W. C. King, and J. Sypniewski, "Geophysical diffraction tomography at a dinosaur site," *Geophys.*, vol. 57, pp. 187–195, 1992.
- [97] A. Witten and W. C. King, "Acoustical imaging of subsurface features," *J. Envir. Eng.*, vol. 116, p. 166, 1990.
- [98] A. Witten and E. Long, "Shallow applications of geophysical diffraction tomography," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-24, p. 654, 1986.
- [99] A. Witten, J. Tuggle, and R. C. Waag, "A practical approach to ultrasonic imaging using diffraction tomography," *J. Acoust. Soc. Amer.*, vol. 83, p. 1645, 1988.
- [100] A. J. Witten, J. E. Molyneux, and J. E. Nyquist, "Ground penetrating radar tomography: Algorithms and case studies," *IEEE Trans. Geosci. Remote Sensing*, vol. GRS-32, pp. 461–467, 1994.
- [101] E. J. Witterholt, J. L. Kretzsehmar, and K. L. Jay, "The application of crosshole electromagnetic wave measurements to mapping of a steam flood," in *Proc. Petroleum Soc. CIM*, 1982.
- [102] E. Wolf, "Three dimensional structure determination of semi transparent objects from holographic data," *Opt. Commun.*, vol. 1, p. 153, 1969.
- [103] —, "Principles and development of diffraction tomography," in *Trends in Optics*, A. Consortini, Ed. San Diego, CA: Academic, 1996.
- [104] R. J. Tombell and R. D. Murch, "The reconstruction of dielectric objects from scattered field data using the distorted-wave Born approximation," *J. Electromagn. Waves Applic.*, vol. 7, pp. 687–702, 1993.

Hypothesis Testing Detection of Mines Buried Under Rough Ground *Surfaces using 2-D FDTD Modeling

He (Sophia) Zhan^{*a}, Carey Rappaport^{**a}, Magda El-Shenawee^b, Eric Miller^a

^a Center for Subsurface Sensing and Imaging Systems
Department of Electrical and Computer Engineering
Northeastern University, 235 Forsyth Building, Boston, MA 02115

^b Department of Electrical Engineering, 3217 Bell Engineering Center,
University of Arkansas, Fayetteville, Arkansas 72701

ABSTRACT

In ground penetrating radar (GPR) antipersonnel mine sensing, in which the target is small, shallow and often of low dielectric contrast, detection is challenging. One of the difficulties is that it is hard to distinguish the target signal from the omnipresent random rough ground reflection clutter. In this work, a Monte Carlo computational simulation using 2-dimensional (2-D) transverse magnetic (TM) finite difference time domain (FDTD) with multiple rough surfaces is implemented to investigate single TNT target buries in dispersive soil.

Based on the effects of the random rough surface on an impulse GPR signal and the knowledge of wave propagation differences in different media – air, soil, and TNT – a special background average process using physics based signal processing (PBSP) is performed to remove the ground clutter signal. This procedure first involves shifting and scaling multiple time signals from target-free random rough ground to establish the nominal (average) ground reflection pulse shape. Next, this nominal pulse shape is correlated in time with each trial signal, then shifted and scaled to match the ground surface clutter of that trial signal. Subtracting this shifted scaled clutter signal from the trial signal ideally leaves the target signal (with some additional multiple scattering between the target and ground surface). The PBSP algorithm reapplied in cases for which surface scattering occurs at multiple points. The statistical results of PBSP surface clutter removal indicate that the detection performance degrades with increasing surface roughness and decreasing burial depth. Hypothesis testing on the processed results proved to be successful in a detection and estimation point of view. This paper presents the detection performances in terms of Receiver Operating Characteristics (ROC) for various ground surface roughness and target burial depth cases. Also demonstrated is the performance improvements expected from multiple views: indicating that a multi-bistatic configuration appears to be superior to multistatic transmitter/receiver geometry with minimum combinations.

Keywords: hypothesis testing, FDTD, GPR, ROC, performance, bistatic

1. INTRODUCTION

Detecting buried dielectric targets – such as antipersonnel mines – with GPR is important since million of mines are still be hidden in the ground, a cost and labor effective technique is needed. However, it is also challenging, because the dielectric constant of nonmetallic mine targets are similar to those of the surrounding soil, and because their sizes are comparable to the thickness of soil layer above them. In addition, the soil dielectric constant may not be well characterized, and the ground surface will usually be rough, often with roughness of the order of the target burial depth. The focuses of mine detection addressed in this paper are ground clutter removal and hypothesis testing which qualifies

*hzhzan@ece.neu.edu; phone (617)373-4874; 467 Egan Research Building, 320 Huntington Avenue, Boston, MA 02115

**rappaport@neu.edu; phone (617)373-2043; fax: (617)-373-8627; 302 Stearns Center, 320 Huntington Avenue, Boston, MA 02115

the detection performance based on 2D TM FDTD model. Although 3D FDTD is more accurate and visible, it is computationally complex and expensive while 2D FDTD can provide much information.

The computational geometry shown in Fig. (1) is typical bistatic geometry based on the Geo-Centers, Inc. EFGPR system¹. The measured Geo-Centers TEMR GPR antenna element radiated signal is used as the excitation, shown in Fig. (2). The TEMR is a wide-band antenna, throughout the 200MHz to 2.0GHz range.

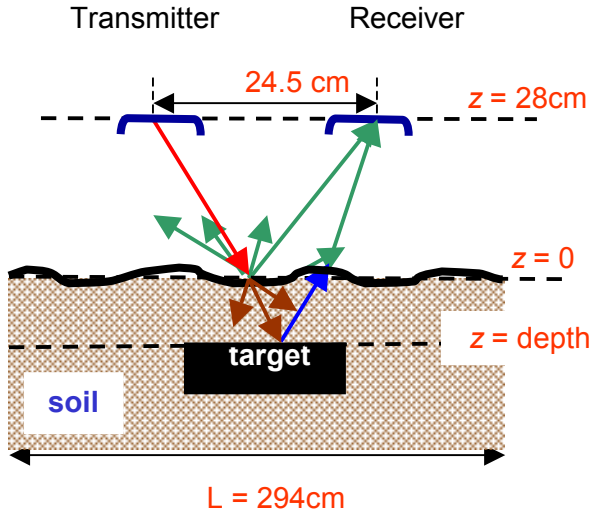


Figure 1. Rough surface computational geometry

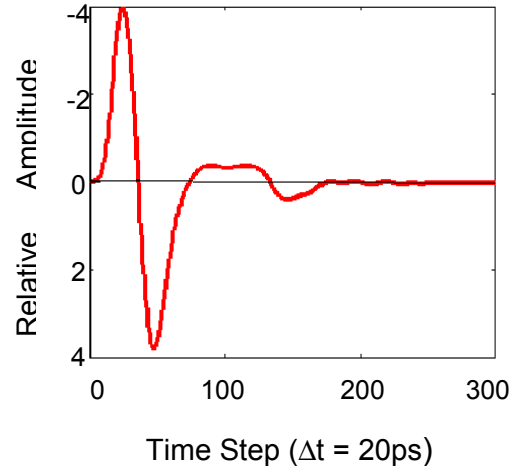


Figure 2. incident measured waveform

The impulsive electromagnetic wave propagates through the dispersiver soil and is partially reflected. The received components in time are: the transmitter-receiver coupling signal with much stronger amplitude than the target signal; the ground reflection clutter signal, with distorted shape due to surface roughness; and the target return that is buried in the clutter.

The soil model is Puerto Rican clay loam² with 10% moisture and 1.4g/cc density, with average dielectric constant $\epsilon' = 6.2$. Since the soil is frequency-dependent medium, it is modeled by a single pole conductivity model based on the observation that the frequency variation of the real dielectric constant does not significantly affect either the real propagation constant or the decay rate^{3,4,5}. The probability density function for the random height of the ground surface has a Gaussian distribution with zero mean and standard deviation σ_h . Also, the surface profiles spectrum is assumed Gaussian^{6,7}. Since we interest in the clutter effects on target discrimination, the single scale Gaussian assumption is reasonable⁸. The frequency-independent target is the dry TNT with dielectric constant 2.9, and almost 0 conductivity⁹, and its size is about 10cm by 5cm.

In the FDTD¹⁰ code, the time step used was $\Delta t = 20ps$, which is matched to the 2D excitation time interval, and spatial difference is calculated to be $\Delta = 1.22cm$. The boundary condition used is the perfect matched layer (PML) absorbing boundary condition¹¹, with which the theoretical reflection factor of a plane wave striking a vacuum layer interface can be made insignificant at any frequency and at any incidence angle. In this computation, 16 PML half-layers are used, and it gives good result for incident angle less than 75 degrees¹².

The simulations are implemented for a variety of rough surfaces characterized by standard deviation for height σ_h and correlation length l_c , also for various depths of buried target, and without target as well. In each simulation, 500 random rough surface realizations are generated.

2. SIGNAL AND CLUTTER SEPARATION MODELING AND PROCESSING

The goal of the process is to maximally suppress the clutter and retain the target signature. Since the surface is rough, the statistical rough ground reflection signal is obtained by the cross-correlation model, and eliminated from the received signal. Furthermore, because the shift of target signal is related to the shift of the ground clutter signal, the target signal is enhanced by the alignment procedure.

When the correlation length l_c is large, the received pulse is a shifted and scaled version of transmitted signal. The time shifting and amplitude scaling are due to air/soil partition of the wave propagation path. By identifying these parameters and measuring their statistics separately, much added information about the ground clutter becomes available. The time shifting is found by the correlation function. Cross-correlation function between reference signal ' f ', which is from an ideal soil half-space with a flat boundary, and any realization signal ' i ' is^{7,13}:

$$C_{fi}(m) = \sum_{n=1}^{N-|m|} S_f(n) S_i(n+m) / \sqrt{\sum_{n=1}^N |S_f(n)|^2} \sqrt{\sum_{n=1}^N |S_i(n)|^2} \quad \text{for } m \geq 0 \quad (1)$$

$$C_{fi}(m) = C_{if}(-m) \quad \text{for } m \leq 0 \quad (2)$$

The relative scaling is obtained from:

$$A_i = \sqrt{\sum_{n=1}^N |S_i(n)|^2} / \sqrt{\sum_{n=1}^N |S_f(n)|^2} \quad (3)$$

where $i = 1, 2, \dots, M$ is the rough surface realization index, M is the size of Monte Carlo sample, and N is the total number of time steps. Another way to obtain the scaling factor is to use the maximum correlation normalization instead. Since the energy normalization is superior for rougher surfaces that generate greater pulse distortion, the relative scaling is calculated using Eq. (3). The flow chart of the PBSP is shown in Fig.(3).

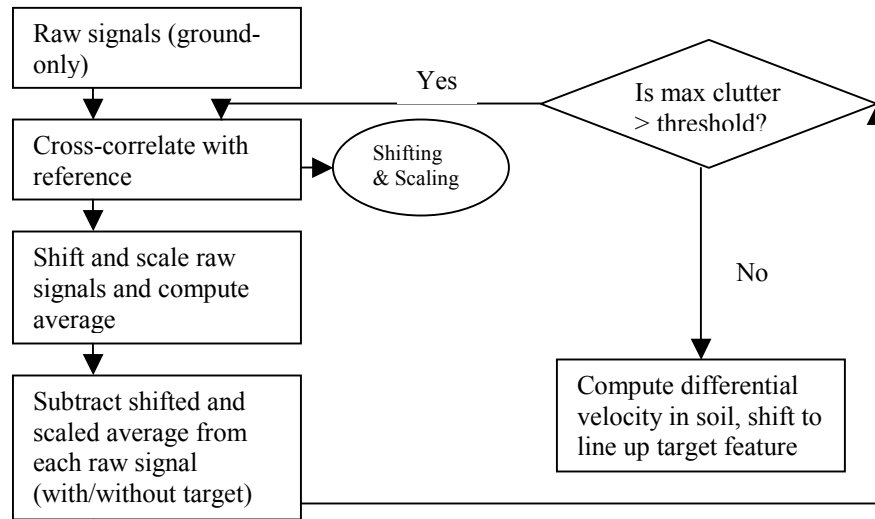


Figure 3. Flowchart of single/multiple PBSP to suppress the ground clutter(s)

The nominal background signal is found by shifting received ground-only signal by $\tau_{gnd,i}$, scaling by A_i , and averaging. The ensemble average gives the best estimate for the local clutter that corrupts the target signal. To reduce ground clutter, the nominal background signal is then shifted back by $\tau_{gnd,i}$, scaled by A_i , and subtracted from each received signal S_i . With the assumption or the knowledge of the depth of the buried target, the path of the scatter signal can be estimated. The wave propagation velocity in soil compared with that of in air is reduced by a factor $1/\sqrt{\epsilon'}$, the shifting factor for the target signal $\tau_{mine,i}$ due to the change of soil amount can be calculated according to the geometry path. Then, the background-subtracted signals are once again realigned to the expected target position^{7,13}. If the surface scattered wave were primarily due to a single specular reflection – as would be the case if l_c were large – then this procedure would suppress most of the surface clutter. However, it is possible that the surface scattering occurs at multiple points. In that case, the cross-correlation/shifts ensemble average/subtraction procedure is repeated.

One hundred computed scattered signals with/without in-ground-target are shown in Fig. (4.a1) and Fig. (4.b1) along with the ± 1 standard deviation confidence interval, with the direct transmitter-receiver coupling signals removed. Those two plots are almost identical. Then, the ground clutter removal procedure is applied, with results shown in Fig. (4.a2) and Fig. (4.b2). Comparing the target-free mean plot to the target-in-ground mean plot, although there are clutter variations involved, there is visible difference in time interval [200 300] – where the target signal would be expected to occur. Therefore, it is possible to distinguish target-in-ground plot from target-free plot. However, it is difficult to localize the target to a certain depth by only looking at Fig. (4.b2), since the clutter amplitudes are comparable to those of the target signal. The realigned signals of the complete PBSP procedure are plotted in Fig. (4.a3) and Fig. (4.b3). The target signals are now well aligned, and the target position can be located by investigating the mean signal in the target-in-ground plot. Furthermore, as expected, the standard deviation is much smaller for the target region of the signal than for the clutter region. The double clutter removal results are shown in Fig. (4.a4) and Fig. (4.b4). Since the clutter due to the second ground reflection has been suppressed, the amplitude of the clutter signal is reduced. Meanwhile, the target signature remains as prominent as expected.

By using the PBSP, most of the ground clutter can be suppressed. However, the processing results are roughness dependent and related to the depth of a buried target. In Fig. (5), the mean signal and STD of several rough surfaces and various depth cases are present.

3. SIMPLE BINARY HYPOTHESIS TESTING

We simplify the mine detection problem to be: whether there is mine at certain depth with the knowledge of the surface roughness. Each of the two answers corresponds to a hypothesis:

H_1 corresponds to the presence of a target (i.e., mine is present)
 H_0 corresponds to the absence of a target (i.e., no mine)

Applying the likelihood ratio test, which is derived to be¹⁴:

$$-\frac{1}{2}(\mathbf{r}^T - \mathbf{m}_0^T)\mathbf{Q}_0(\mathbf{r} - \mathbf{m}_0) - \frac{1}{2}(\mathbf{r}^T - \mathbf{m}_1^T)\mathbf{Q}_1(\mathbf{r} - \mathbf{m}_1) \underset{H_0}{\overset{H_1}{>}} threshold \quad (4)$$

where \mathbf{r} , \mathbf{m}_0 and \mathbf{m}_1 are vectors with N elements, \mathbf{Q}_0 and \mathbf{Q}_1 are $N \times N$ covariance matrices. \mathbf{r} represents the individual background-subtracted signals. \mathbf{m}_0 is the average of realigned no target (clutter) signals; \mathbf{m}_1 is the realigned target signals. The covariance matrices \mathbf{Q}_0 and \mathbf{Q}_1 are the inverses of the diagonal matrices of standard deviations of the clutter and target signals. Since the processed signals are not guaranteed independent of one and another, the diagonal covariance matrices used in the test are approximations. We are investigating the worse case hypothesis test here. 100 out of 500 target signals and 100 out of 500 clutter signals are grouped at test signals. The average signals \mathbf{m}_0 and \mathbf{m}_1 , and standard deviations are obtained from the remaining 400 clutter signals and 400 target signals. The likelihood ratio

test is applied to each test signal associated with a set of trial thresholds. Subsequently, the probability of *false alarm* P_F (*i.e.*, we say the target is present at certain depth when it is not) and the probability of, detection P_D (*i.e.*, we say the target is present at certain depth when it is) are obtained. The performance of the likelihood ratio test is evaluated by receiver operating characteristic (ROC) curve¹⁴. The ROC curves for a variety of rough surfaces and different depths are shown in Fig. (6). As the depth increases, the performance improves, *i.e.*, when a buried target is close to the surface, it is hard to distinguish the ground clutter from the target signal since the ground clutter signal and target signal are too close to remove the clutter signal without affecting the target signal. The detection performance degrades as the standard deviation increases. Thus, the detection becomes difficult as expected when the surface is rough and the target is shallow buried.

By relating signals associated with certain depth to several average signals for different depths, it has been found that the best performance always occurs for the correct depth estimate. An example is shown in Fig. (7). The test signals for depth 8.5cm were related to the average signals and corresponding standard deviation that are belong to the same roughness family. Thus, the correct target depth will be obtained by testing sample signals with stored statistics of various families with same surface roughness. Interestingly, the ROC curves for close but not exact depths are worse than those for much deeper or shallower depth estimates.

The previous results are based on the geometry shown in Fig. (1) -- the target is centered beneath the transmitter-receiver (TR) pair. Recall that the detection performances decrease for rougher surfaces. In order to improve the performance, multi-bistatic and multistatic geometries are investigated. Now, in addition to then single centered TR pair position, we used four more TR positions for multi-bistatic geometry: 12.2cm and 24.4cm to the left and right of the center. As the pair moves away from the target, the test performance degrades. Also, the amplitude of the average target signal drops. However, the combination of three TR positions improves the detection performance significantly. With five TR views, the performance is even better. It is also true for the multistatic case with fixed transmitter and multiple receiver views. The improvements of using three/five TR pairs and three/five receiver views are shown in Fig. (8.a) and Fig. (8.b). Although performance increases with the number of views, the combinations of three TR positions, or three receiver views, each spaced 12.2cm apart to the left and right, appear to give excellent results.

4. CONCLUSION

The computational study of 2D TM FDTD is presented to investigate a single dry TNT target at certain depth buried in dispersive soil with the presence of a random rough surface. The dispersive FDTD algorithm is used to incorporate the frequency-dependent complex conductivity media – Puerto Rican clay loam. The air/soil interface is characterized by Gaussian roughness, since the single scale roughness is well described by Gaussian process. The GPR impulse used is based on a measured transmitter pulse, with a wide-band width and the capability of exploring target in the presence of a lossy medium. The PML boundary is employed to terminate the computational grid.

The simulated result is processed using a background average procedure – PBSP, which even works for a shallow buried target. The ground clutter is mostly removed by the ensemble average and subtraction procedure. Also, shifting and scaling the individual received signal back and forth according to the transmission path in the air and the soil enhance the target signature. The statistical results of the PBSP indicate that the quality of target signals depends on the roughness of the surface and the depth of a buried target.

In addition to examining the statistics of fields scattered from a Gaussian rough surface, we investigate the worse case target detection performance by using the classic detection theory. Applying the binary hypothesis tests to the processed signals, the target can be localized at certain depths even under rough ground surfaces. Furthermore, the spatial combinations of TR position will improve the detection performance.

ACKNOWLEDGEMENT

This work is supported by the Army Research Office ground No. DAAG55-97-0013.

REFERENCES

1. C. Rappaport and D. Reidy, "Focused array radar for real time imaging and detection", in Proc. Of SPIE, vol. 2747, pp.202-213, (Orlando, FL), Apr. 1996
2. J. Hipp, "Soil Electromagnetic Parameters as Functions of Frequency, soil Density and Soil Moisture," *Proceedings of the IEEE*, vol. 62, no, 1, Jan.1974, pp. 98-103
3. C. Rappaport, S. Wu, and S. Winton, "FDTD wave propagation in Dispersive soil using a single pole conductivity model", accepted for publication in *IEEE Trans. On Magnetism*, May 1999
4. W. Weedon, and C. Rappaport, "A General Method for FDTD Modeling of Wave Propagation in Arbitrary Frequency-Dispersive Media", *IEEE Transaction of antennas and Propagation*, pp. 401-410, March 1997
5. C. Rappaport and S. Winton, " Modeling Dispersive soil for FDTD Computation by Fitting Conductivity Parameters", *12th Annual Review of Progress in Applied Computational Electromagnetics Symposium Digest*, pp. 112-118, March 1997
6. P. Beckmann and A. Spizzichino, *The Scattering of electromagnetic Waves from rough surfaces*, New York, 1963
7. M.El-Shenawee, C. Rappaport, " Quantifying the Effects of Different Rough Surface Statistics for Mine Detection Using the FDTD Technique", *Proc. SPIE*, April 2000
8. E.I.Thorson, "The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum", *J. Acoust. Soc. Amer.*, vol. 83, pp.78-92, 1988
9. A. von Hippel, *Dielectric Materials and Applications*, Wiley, New York, 1953
10. K. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media", *IEEE Trans. Antennas Propagat.*, vol. AP-14, pp302-307, May 1966
11. J. Berenger, "A Perfectly Matched Layer for the Absorption of Electromagnetic Waves", *Jour. of Comp. Phys.*, vol. 114, pp 185-200, Oct. 1994
12. C. Rappaport, M. Kilmer and E. Miller, " Accuracy considerations in using the PML ABC with FDTD Helmholtz equation computation", *International Journal of Numerical Modelling: Electronic Networks Devices and Fields* vol.13, pp.471-482, 1999
13. C.Rappaport, M.El-Shenawee, " Modeling GPR Signal degradation from random rough Ground surface ", *IGARSS*, Honolulu, HI, July 2000
14. H. Van Trees, "Detection, Estimation, and Modulation Theory", John Wiley and Sons Inc., 1968

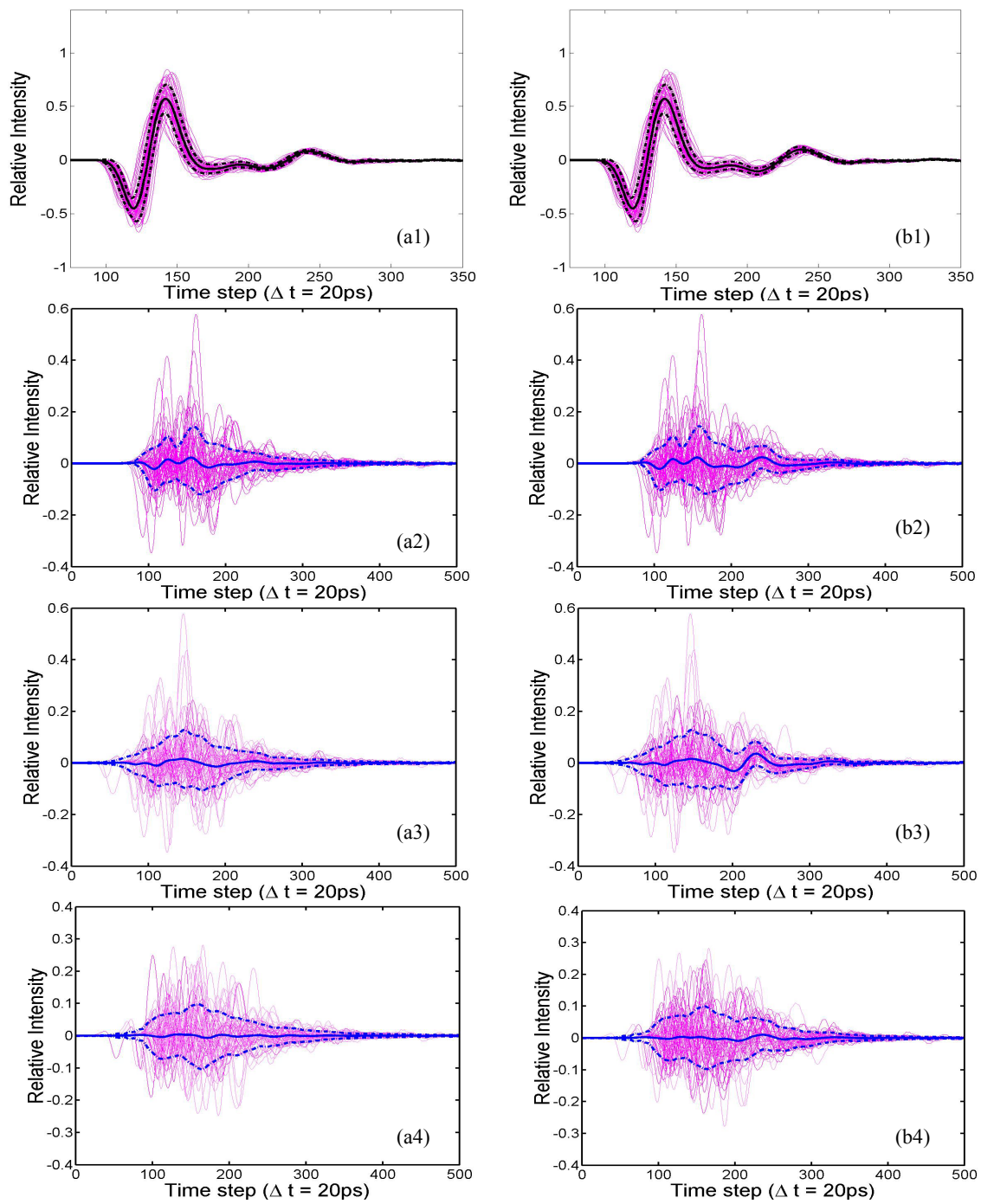


Figure 4. For case $(\sigma_h = 3, l_c = 10, z = 8.5)\text{cm}$

- (a1) target-free raw signals; (b1) target-in-ground raw signals
- (a2) ground clutter removed target-free signals; (b2) ground clutter removed target-in-ground signals
- (a3) realigned target-free signals; (b3) realigned target-in-ground signals
- (a4) double-clutter removed and realigned target-free signals; (b4) double-clutter removed and realigned target-in-ground signal

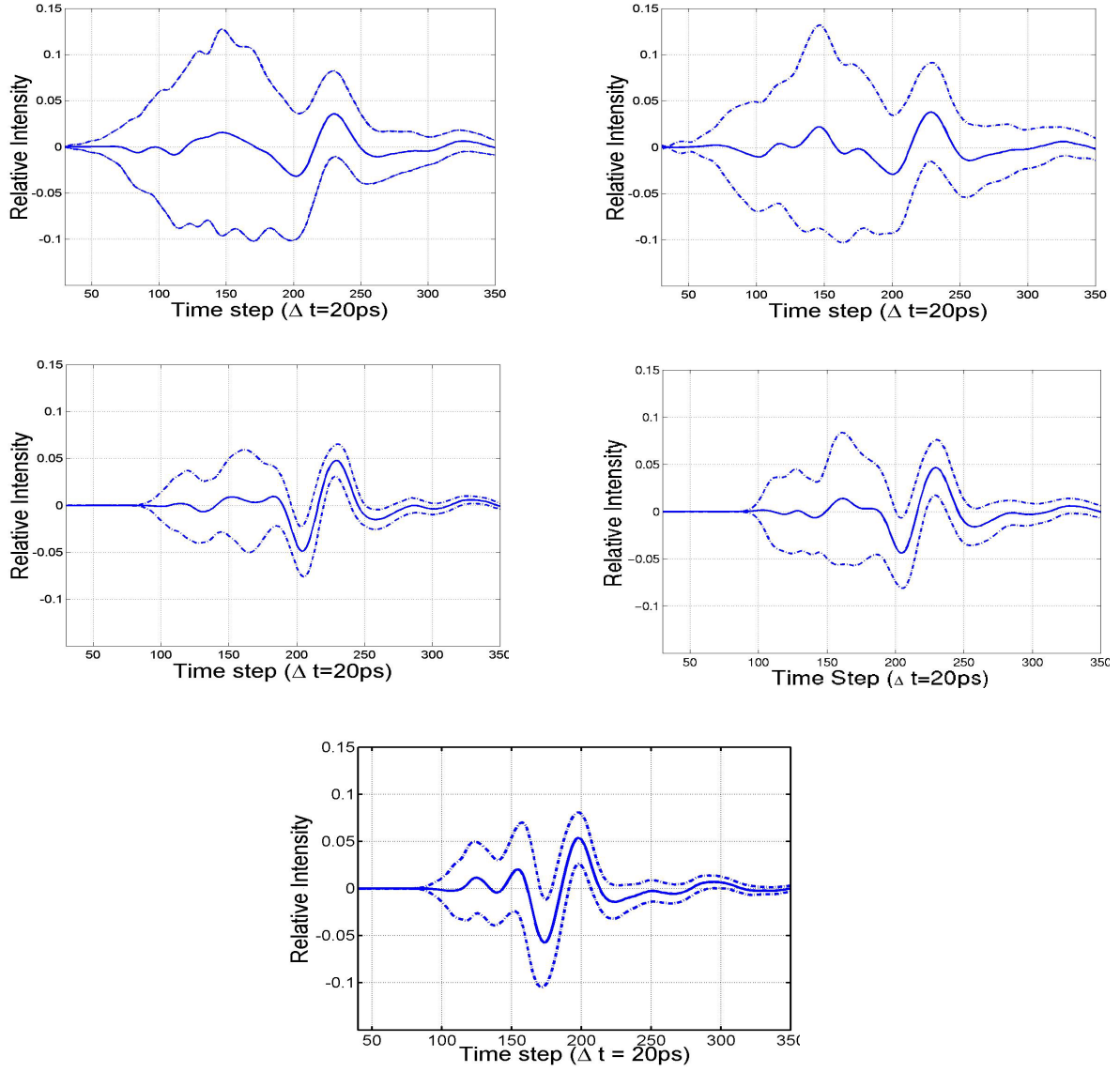


Figure 5. Mean signal and STD for several rough surfaces and various depths.
 (a) $(\sigma_h = 3, l_c = 10, z = 8.5)cm$; (b) $(\sigma_h = 3, l_c = 3, z = 8.5)cm$
 (c) $(\sigma_h = 1, l_c = 10, z = 8.5)cm$; (d) $(\sigma_h = 1, l_c = 3, z = 8.5)cm$
 (e) $(\sigma_h = 1, l_c = 10, z = 4.8)cm$

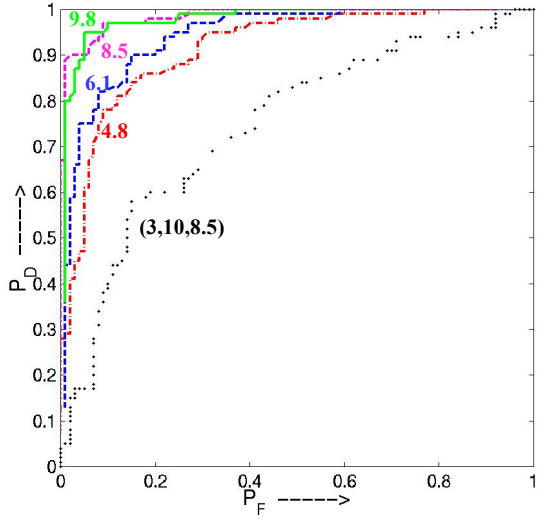


Figure 6. ROC curves for (σ_h, l_c, z) cm = (1,10,4.8), (1,10,6.1), (1,10,8.5), (1,10,9.8) and (3,10,8.5).

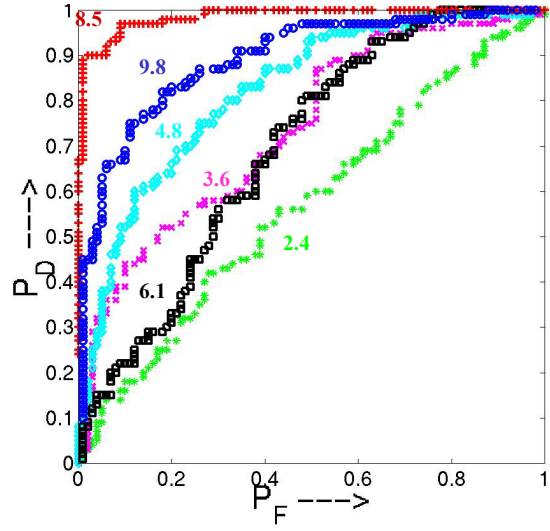


Figure 7. ROC curves, for (σ_h, l_c) cm = (1,10) with target at depth 8.5cm, using various depth test hypothesis

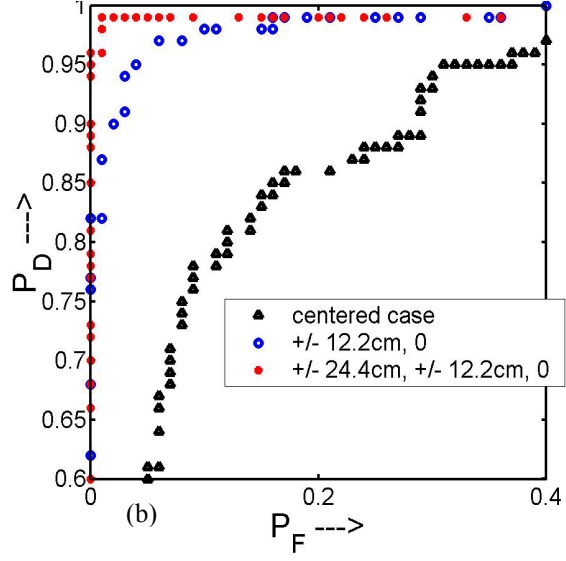
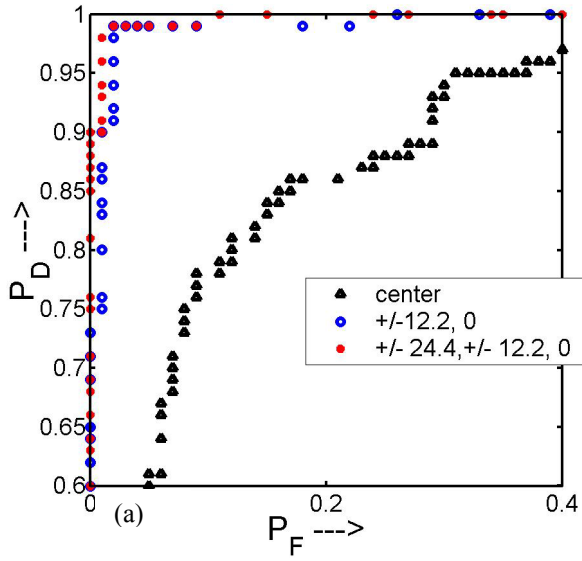


Figure 8. (a) ROC curves for combinations of 1,3,5 TR pairs; (b) ROC curves for combinations for 1,3,5 R views.

Three-dimensional FDTD model for GPR detection of objects buried in realistic dispersive soil

Panagiotis Kosmas, Yuequan Wang, and Carey Rappaport*¹
Center for Subsurface Sensing and Imaging Systems
Department of Electrical and Computer Engineering
Northeastern University, 302 Stearns Bldg
Boston, MA 02115

ABSTRACT

The use of ground penetrating radar (GPR) is one of the most popular techniques for the detection of anti-personnel mines and therefore it is desirable to accurately model such systems. For many GPR applications, FDTD models used to simulate the system are two-dimensional, because they are simple to implement and computationally inexpensive. However, a three-dimensional model is more accurate and allows complete freedom for the location of the object relative to the receivers. Instead of fully modeling the transmitter and receiver elements, and adding significant complexity, the transmitted field in this study is experimentally measured and used as the model's excitation. The model developed simulates a GPR system consisting of a parabolic reflector transmitter and a multi-static receiver array. The model is tested for both flat and rough ground with a Gaussian variation. The results are compared with experimental data and are found to be very accurate. The validation of this approach makes the model a powerful tool that can be used in different applications, where the exciting field is computationally or experimentally specified.

Keywords: FDTD, modeling, ground-penetrating radar, mine detection, parabolic reflector

1. INTRODUCTION

Land mine detection using ground-penetrating radar (GPR) has been a topic of research for many years, and research has shown that it is a challenging problem with no unique solution. The properties of the transmitting and receiving antennas used, the geometry of the system, and the signal processing algorithms employed for the processing of the data are all important aspects of the problem. A description of these design parameters and a summary of some of the most popular GPR systems can be found in [1].

Among the several different techniques, the forward-looking parabolic reflector transmitter GPR aims to reduce the clutter from the ground surface and thus make the detection of the object an easier task. This system, which was designed by Northeastern University and fabricated at Geo-Centers, Inc., has been described in previous work^{2, 3, 4}. The parabolic reflector transmitter illuminates the ground with a forward propagating, quasi-planar wave. A significant amount of the ground clutter will scatter away from the ground surface, and therefore the relatively isotropic scattered signal from a buried object has a higher probability of being detected by the receivers, positioned in the backscattered direction. The possibility of detecting the small signal scattered from the object depends on the electromagnetic properties of the material, its position relative to the receivers and how deep it is buried into the ground. Using a short-duration excitation impulse and a multi-static array receiver system, the detection performance for shallow buried targets is enhanced³.

¹ Contact authors at pkosmas@coe.neu.edu, rappaport@ece.neu.edu, Tel: 1-617-373-8387, 1-617-373-2043 Fax: 1-617-373-8627

In order to simulate this system, we can measure the transmitted field voltages and find a way to extract the electric field that will be used as an input to the model. The modeling is performed using the Finite-Difference Time-Domain numerical method. The system has been compared with previous 2D models², but here a fully three-dimensional model is developed, which simulates the incident quasi-planar wave, propagation in frequency dependent soil, the ground clutter for flat and rough cases, and the signal scattered by the object, resulting in the signal received by the four-receiver multi-static array.

2. GPR SYSTEM MODELING

The FDTD method is one of the most popular numerical techniques for electromagnetic problems. It is based on discretizing Maxwell's equations in time and space using a difference scheme for the approximation of the derivatives, on edges of the so-called Yee-cell⁵. Several issues related to the numerical implementation of the FDTD technique, such as stability and dispersion⁷, together with geometry considerations, play an important role in determining the size of this cell, which is chosen to be uniform for the present application.

A fully three-dimensional FDTD model would include the transmitting and receiving antennas, the parabolic reflector, the air/ground interface and the buried object. In addition, the transfer functions that convert voltages to electric fields are also necessary in order to compare the simulation results to measured data. Alternatively, we can use a more simplified approach if we make certain assumptions for our system. First of all, if the receiving antenna is located in the far-field of the transmitting system, there are no significant ringing effects and therefore the knowledge of the transmitting fields is sufficient for use as an excitation model. Moreover, if we consider the system to be linear and time-invariant (Fig.1), then we can interchange the order of the “boxes” in figure 1. Then, we can measure the voltages with the receiver before the channel, and use these voltages as input to get output voltages, which can be directly compared to the data.

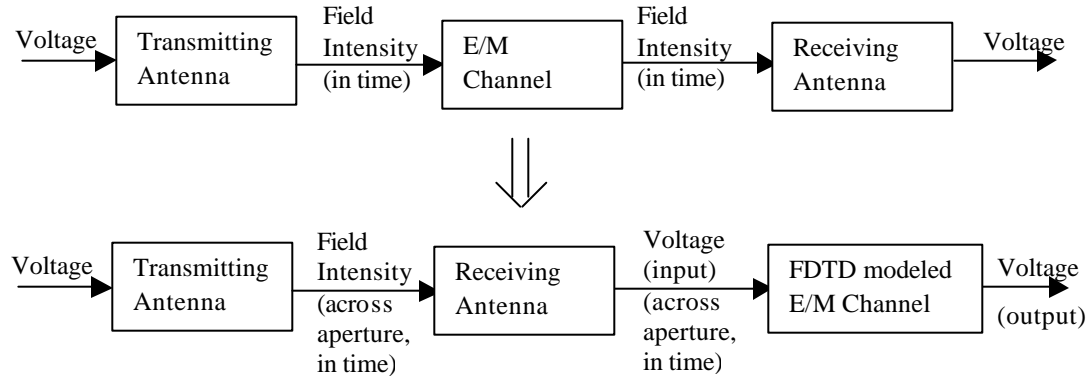


Fig. 1: Transition from the physical system to the equivalent model by interchanging the order of processes.

Based on these assumptions, we can consider a plane in the 3D grid that is excited by field values that are experimentally measured and then let this incident quasi-planar wave propagate, encounter the soil and the object, and record the signal at the points where the receivers are located. The geometry of the model is depicted in figure 2. The procedure followed to derive the input excitation by the measured values is described in the next section. Here we briefly discuss our approach towards some of the most important issues that are related to the FDTD method. These include the modeling of lossy, dispersive soil and the absorbing boundary conditions. A more detailed analysis can be found in [6].

Any FDTD analysis should account for the frequency dependence of the soil's electrical properties. There are several methods that are developed to model this dispersion based on fitting parameters to match measured data for a range of

frequencies^{10,11}. The present model uses a z-transform method based on a single-pole, two-zero conductivity model, which leads to good agreement with measured soil data over two decades in frequency⁸. This one-pole approach requires storage of only two previously computed field values and leads to very simple equations for the propagation in soil that use these past values and some additional constants.

Due to the ground's abnormalities and the presence of other possible scatterers, such as rocks or metal objects, it is also necessary to introduce ground roughness in the FDTD model. For the 3D geometry, a rough surface is generated using a height variation that has a joint, zero-mean Gaussian distribution in both x and y dimensions¹³.

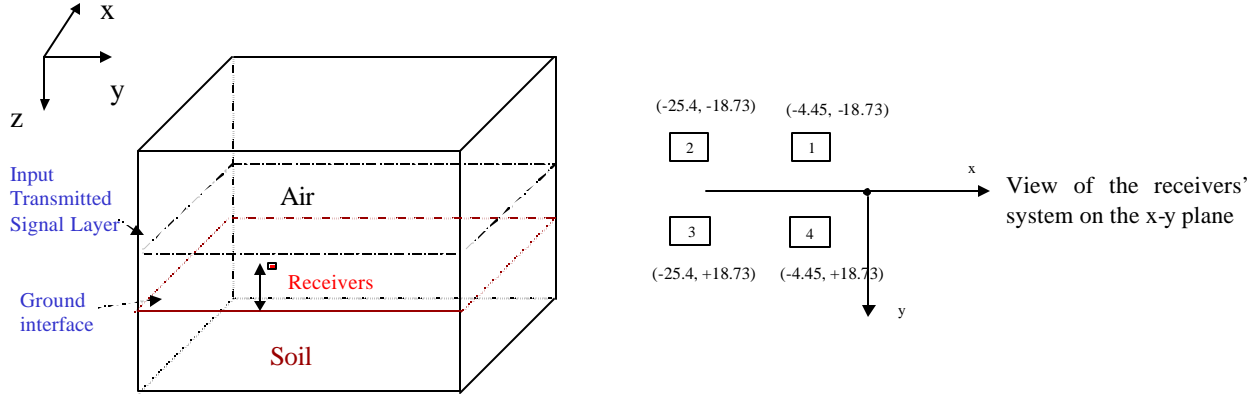


Fig. 2: Geometry of the 3D FDTD grid that simulates the GPR system and the four-receiver plane.

Terminating the grid with an effective absorbing boundary condition (ABC) that will significantly reduce unwanted reflections is one of the most important objectives for any FDTD application. In the literature, the perfectly matched layer (PML) absorbing boundary condition¹⁴ has been extensively studied and its modified versions can be used for a 3D model that includes free space and lossy, dispersive soil. An alternative approach is to use the Mur absorbing boundary conditions¹⁵. Although the PML performance in free-space is superior to the Mur ABC, in the frequency-dispersive soil it is comparable to a Mur-type ABC that uses the single-pole conductivity model mentioned above^{6,9}. Using a point-source excitation, this 3D Mur-type ABC has been tested in Puerto-Rican clay loam with 1.4 g/cc density and 10% moisture and resulted to a maximum 5% reflection on the edges of the grid⁶, and testing for 1D cases for 5% moisture with a 1GHz Gaussian exciting pulse has shown that up to a 15° angle, this Mur-type ABC behaves better than the PML⁹. For an approximately 45° incident plane wave, the angle of propagation in the soil is within this 15° range, making this Mur-type ABC preferable to the PML. In addition, using Mur-type ABC in the model is much simpler to implement and faster than a 3D air/soil PML. As for the reflection in the free-space region, the planes of excitation and receivers in the 3D grid (figure 2) are located closely enough to the ground that these reflections occur later than the ground clutter and mine signal⁶.

3. PARABOLIC REFLECTOR'S INCIDENT WAVE

The offset paraboloidal reflector antenna generates a quasi-planar wave of approximately 45° of incidence that can be considered fairly uniform in the ground region of interest (figure 3). The offset section is necessary to avoid the blockage of rays by the feed structure. The reflector collimates rays from an ultra-wideband transmitting feed. A detailed description of the necessary design considerations and the characteristics of the system can be found in references [4] and [12]. Here we are interested in describing the procedure that was followed in order to use the wave generated by this system as the excitation for the FDTD model, based on measured data. Figure 3 shows the geometry of the performed experiment.

In order to simulate this system, we experimentally measured the transmitted field. Turning the parabolic antenna upside down, we measured amplitudes and delays on the xy plane, using a Styrofoam sheet as physical support for the measurement. The measurements were taken every two inches in the x -direction for the interval $[-10'', 10'']$ and every three inches in the y -direction for the interval $[-9'', 9'']$, where the center $(0,0)$ corresponds to the beam center of the parabolic antenna, as shown in figure 3. In addition, the waveforms of the incident field were recorded for the center and the corners of the x - y plane. These nine positions are marked in the figure.

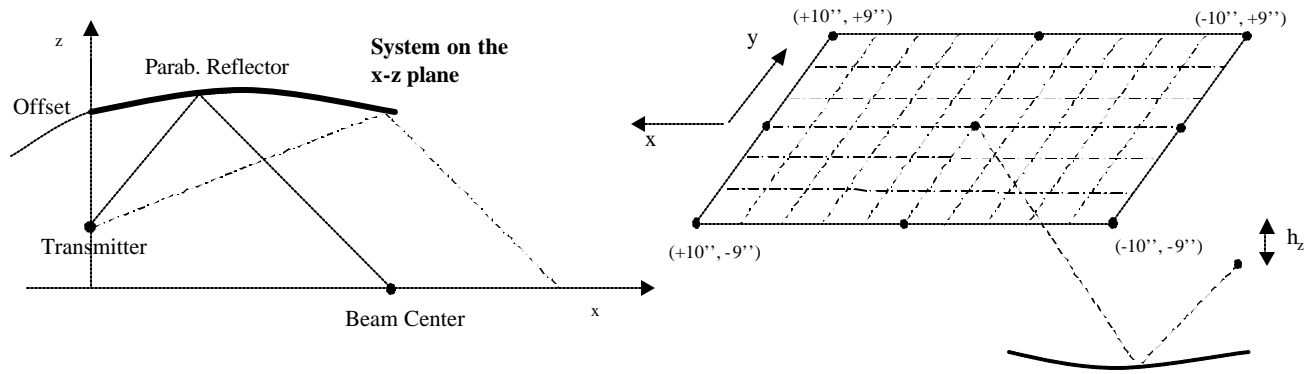


Fig. 3 Experiment for measuring the field wave generated by the parabolic reflector. On the left, the actual system facing the ground is depicted and on the right, the system is turned upside-down in order to measure the transmitted field as function of time position on the x - y plane.

In order to derive the field values for every grid cell from the recorded waveforms and the measured amplitudes and time delays, we first need to interpolate these measured values for the intermediate points according to the length of the cubic cell's sides that are used in the simulation. The time step of the recorded waveforms also needs to be equal to the one used in the simulation. Then an interpolation algorithm performs the following steps:

1. Take the waveforms at $(0, 0)$ and $(0, +9)$ to derive the waveforms for all points along $(0, y)$.
2. Repeat for $(+10, 0)$ and $(+10, +9)$ to derive waveforms along $(+10, y)$.
3. Use $(0, y)$ and $(+10, y)$ to derive the waveforms for all the points (x, y) with $x \in [0, +10]$, $y \in [0, +9]$.
4. Repeat steps 1-3 for the three remaining blocks of data.
5. Unify the results to produce the interpolated field for (x, y) , $x \in [-10, +10]$, $y \in [-9, +9]$ and for an x (y) increment that corresponds to the Δx (Δy) of the simulation.

The fields that are obtained using this method are shown in figure 4, where cuts along the x and y -axis through the center are presented. For display purposes the field values in these figures have been under-sampled both in space and time. These figures verify that the measured transmitted field possesses the characteristics desired for land-mine detection applications, being relatively planar, tilted 45° to vertical in the xz plane and having a gradual amplitude taper away from the center of the beam. The wave transmitted by the parabolic reflector is only an approximation to a 45° incident plane wave. These plots, though, illustrate that this approximation is realistic.

Figure 4 also shows the result of exciting the E_x component of the electric field with the above values in a 3D FDTD grid. The H_y magnetic field component on the xz surface that passes through the center of the grid is plotted for three different time steps, in the vicinity of the air/ground interface. The output of this simulation further illustrates the validity of the plane-wave approximation, especially locally, in the air/ground region that includes the plane of the receivers and the mine target.

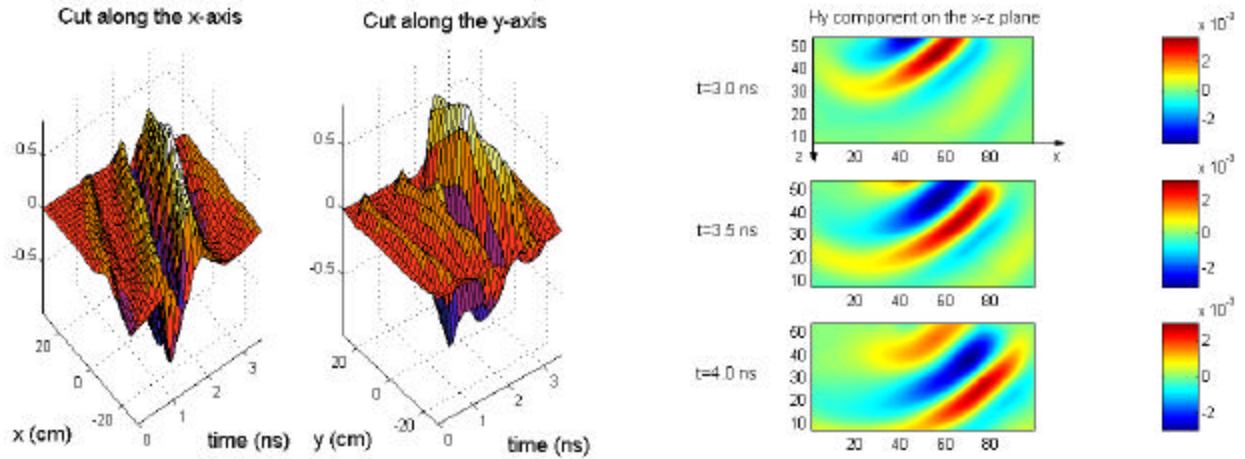


Fig. 4: Parabolic reflector's wave that excites the 3D FDTD grid. The cuts along the x and y-axis through the center illustrate the characteristics of the incident wave. As the wave propagates in the grid, it locally resembles a plane wave with 45° incident angle.

4. RESPONSE OF REALISTIC SOIL USING GPR SIMULATIONS AND DATA

The parabolic transmitter shown in figure 3 was used in an outdoor experiment in Northeastern 's Dedham Campus outdoor track. The GPR receiver array consists of two forward and two rear antenna elements, positioned close to the ground, behind the focal point of the transmitting antenna. The parabolic antenna system is put on a survey cart that can travel along-rails down-track. Several plastic and metallic objects are buried along the track (figure 5). The soil can be approximately considered as clay loam with 1.4 g/cm^3 density. We use measured soil values for Puerto-Rican clay loam¹⁶. The statistical characteristics of the roughness of the ground are not expected to change significantly along the track. As the cart moves along, the received signal at each of the four receivers is measured for every half inch. Starting with the center of the system (beam center) being in the center in the y (cross-track) direction and moving it along the x (down-track) direction, the process is repeated for every two inches left and right (y-direction) to the center.

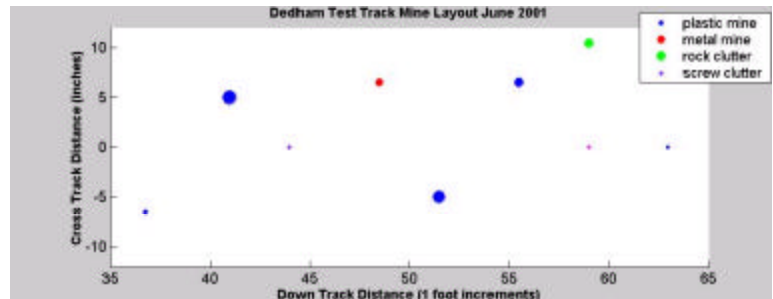


Fig. 5: Photograph of the GPR system and diagram of the outdoor track

In order to compare simulated and experimental data, a scan corresponding to a track location with no object beneath the ground was considered. To simulate the ground response recorded at this scan, the three-dimensional FDTD model was

run for both flat and rough ground cases, and the planes of excitation and receivers were placed as in figure 2, with the receivers' height equal to their distance from the ground in the outdoors experiment. The ground roughness parameters are set to be 10.0 cm for the correlation length and 1.0 cm for the mean square height. These parameters are chosen to correspond to a moderate rough case.

In order to get the ground response from the scanned signal, we first need to determine the direct signal from transmitter to receiver and subtract it. It is important and rather difficult to capture an accurate reference signal in an outdoor experiment due to reflections from the surrounding environment. Furthermore, since this signal will determine the time alignment of the data, finding the starting time is also critical. For this reason, an air-shot signal was recorded four times for each receiver, the four waveforms were aligned and their average was used as the reference direct signal for each receiver. The air-shot signals were taken by turning the system upside-down to minimize ground reflections. The reflection caused by the surrounding environment can be considered as part of the direct signal. Then the reference direct signal for each receiver was subtracted from the scanned signal and the result was normalized. For each receiver, the total number of samples in the scanned data is 512, but the sampling rate is slightly different, namely 38.2, 37.1, 33.3 and 37.5ps for receivers 1,2,3, and 4 respectively. Thus, the time window is slightly different for each receiver. The starting time is determined by time alignment based on the air-shot signal.

For the FDTD simulations the direct signal can be easily obtained by removing the ground and recording the signal at the four receivers. This stored signal is then subtracted by the ground signal and, after normalizing, the ground response for flat and rough ground is obtained. The starting time, as with the experimental data, is arbitrary. The simulated signal was normalized to the maximum value of the experimental signal for comparison. The soil was modeled to simulate the Puerto Rican clay loam that is considered for the experiment. The time step for the simulations was 10ps, so the simulated data was extrapolated to match the time steps of the four receivers in the experiment. As mentioned above, due to this difference in the sampling rate, the time windows for the four receivers are slightly different.

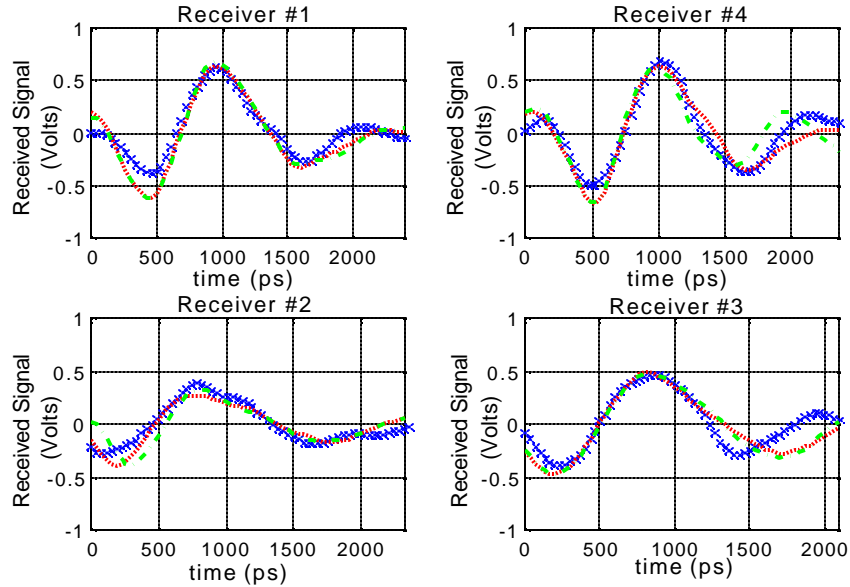


Fig. 6: Realistic soil response (measured data vs. FDTD model). The x-marked line corresponds to the measured data, the solid and dash-dotted lines to the simulation results for flat and rough ground respectively. The ground roughness follows a Gaussian distribution with parameters 10.0 cm for the correlation length and 1.0 cm for the mean square height.

The ground response of the simulation model vs. the experimental data is shown in figure 6. The figure shows a very good agreement of the three-dimensional model with the data. The ground roughness effect is mostly apparent at the fourth receiver, and the measured data lies between the flat and rough ground simulation response for the time interval where these two responses are remarkably different. For the rest of the receivers, the measured data is in very good agreement especially in the interval where the main peak occurs, and only at receiver 3 there is a time delay after the main peak for the simulation response in comparison to the data, which may be attributed to experimental errors or non-predicted roughness effects.

5. SCATTERING FROM METALLIC OBJECT

An analysis of the 3D model's predicted scattered signal from plastic and metallic mines and a comparison with two-dimensional models can be found in [6]. However, in order to further confirm the validity of the three-dimensional model, we performed an additional test for which reliable experimental data was available.

This test involves the scattered signal for an aluminum sphere placed in the center of the grid. The simulation results are compared with measurements performed inside, where the sphere was placed on the center of the Styrofoam sheet (figure 3). Here there is no air/ground interface and thus no ground signal and the waveforms were recorded in the experiment with the direct signal subtracted. The transmitting pulse is different and so is the recording time window. The rest of the geometry remains the same. As with the experimental waveforms, the direct signal is recorded and subtracted for the simulation runs. Here, the sampling rate is the same for the four receivers and the starting time is arbitrary, and it is chosen for the simulations so that the signal peaks are aligned. The results are shown in figure 7, with the simulation output having been normalized to the maximum value of the measured output.

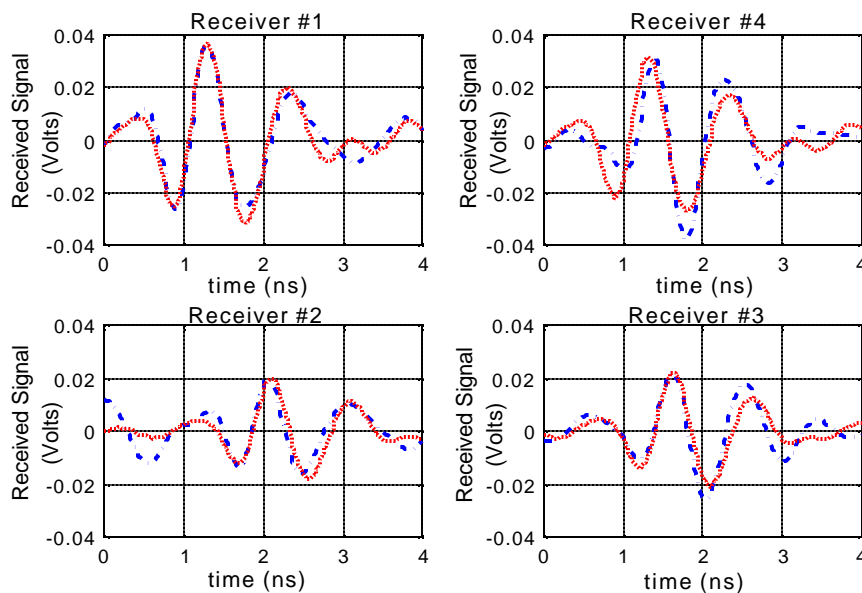


Fig. 7: Scattered signal from an aluminum sphere in free-space (measured data vs. FDTD model). The dash-dotted line corresponds to the measured data, and the solid line to the simulation results.

Figure 7 further illustrates the validity of the three-dimensional FDTD model. Using a different excitation pulse for the parabolic reflector transmitter, the agreement is very good with a small difference only for receiver #4. The scattered

signal is higher at the front receivers 1 and 4, as expected since the ball is located closer to them (beam center of figure 2). The peak occurs later for the rear receivers 2 and 3, due to the propagation time delay.

6. CONCLUSIONS

The three-dimensional FDTD model can accurately simulate a GPR application, such as land mine detection using a parabolic reflector system, which is studied in this work. Following an approach that aims to simplify the complexity of the model as much as possible, we based our excitation on experimental measurements and an interpolation procedure and used simple one-pole models for the soil equations. Mur-type absorbing boundary conditions were chosen for their simplicity and good performance in soil. Placing the receivers close to the ground, the unwanted reflections from the upper sides of the grid occur later than the examined signal.

The model was tested in two different experiments. Using the Geo-centers, Inc. radar pulse as excitation, the ground signal at the four receivers was compared to experimental data taken using the parabolic reflector system at Northeastern's outdoor track. The measured data was found to lie between the flat and rough ground simulated data. Then, for a different pulse transmitter signal, the scattered signal from an aluminum ball was compared to data taken in an inside experiment. The very good agreement for these two different cases is a very strong proof for the validity of the model.

In [6], a study of the ground response with and without plastic and metal mines is studied as a function of position, and the results are compared to two-dimensional cases and conclusions on the ability to detect a plastic non-antipersonnel mine are drawn, based on this FDTD model. It is our objective to test these conclusions with additional reliable data in the future. The validation of this model makes it a powerful tool since, if modified, it can be used for a variety of applications and different systems, once the incident wave is accurately calculated. In other words, the model can be used to break up a larger problem, provided that we are in the far-field of the transmitting antenna and effects of ringing between transmitting and receiving antennas are negligible.

ACKNOWLEDGEMENTS

This work was supported by The Army Research Office Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013 and the Center for Subsurface Sensing and Imaging Systems (CenSSIS) at Northeastern University, under the Engineering Research Centers Program of the National Science Foundation (Award Number EEC-9986821). The experiments were performed with the help of Geo-Centers, Inc., which also designed the transmitting and receiving antenna elements.

REFERENCES

1. T.R. Witten, "Present state of the art in ground-penetrating radars for mine detection", *Proceedings of SPIE*, A.C. Dubey, J. F. Harvey, and J.T. Broach, vol. 3392, pp.576-585, SPIE, Orlando FL, 1998.
2. B. Yang and C. Rappaport, "Response of Realistic Soil for GPR Applications with Two Dimensional FDTD ", *IEEE Transactions on Geoscience and Remote Sensing* **39**, pp. 1198-1205, June 2001.
3. C. Rappaport and D. Reidy, "Focused array radar for real time imaging and detection", *Proceedings of SPIE*, vol. 2747, pp.202-213, SPIE, Orlando FL, 1996.
4. C.Rappaport, D. Jin, S. Azevedo, T. Rosenbury and A. Dean, "Forward-looking handheld anti-personnel parabolic reflector-based GPR mine detection", *MURI Review Meeting*, FT. Belvoir VA, August 1999.
5. K. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Transaction on Antennas and Propagation* **14**, pp. 302-307, 1966.
6. P. Kosmas, "Three-Dimensional, Finite-Difference Time-Domain Modeling for Ground Penetrating Radar Applications", MSc Thesis, Northeastern University, March 2002.
7. K. Kunz and R. Luebbers, "The Finite Difference Time Domain Method for Electromagnetics", CRC Press, Boca Raton, FL, 1993.

8. C. Rappaport, S. Wu and S. Winton, "FDTD wave propagation in dispersive soil using a single pole conductivity model", *IEEE Transactions on Magnetics* **35**, pp. 1542-1545, May 1999.
9. J. Talbot, C. Rappaport and P. Kosmas, "An efficient Mur-type ABC for lossy scattering media", *Progress in Electromagnetic Research Symposium*, Cambridge, July 2000.
10. W. Weedon and C. Rappaport, "A general method for FDTD modeling of wave propagation in arbitrary frequency-dispersive media", *IEEE Transactions on Antennas and Propagation* **45**, pp. 401-410, March 1997.
11. R. Luebbers, F.P. Hunsberger, K.S. Kunz, R.B. Standler, and M. Schneider, "A frequency-dependent finite-difference time-domain formulation for dispersive simulations", *IEEE Transactions on Electromagnetic Compatibility* **32**, pp.222-227, Aug.1990.
12. C. Rappaport and B. Yang, S. Azevedo, T. Rosenbury and J. Gough, and A. Dean, "Detecting Antipersonnel Mines with a Handheld Parabolic Reflector Transmitter/Multistatic Receiver Impulse GPR ", *2000 UXO/Countermine Forum*, May 2000.
13. C. Rappaport, M. El-Shenawee, and H. Zhan, "Suppressing GPR Signal Degradation from Randomly Rough Ground Surfaces," in review for publication in *IEEE Trans. on Geoscience and Remote Sensing*.
14. J-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves", *Journal of Computational Physics* **114**, pp. 185-200, Oct.1994.
15. G. Mur, "Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic field equations", *IEEE Transactions on Electromagnetic Compatibility* **23**, pp. 377-382, Nov. 1981.
16. J. Hipp, "Soil electromagnetic parameters as functions of frequency, soil density and soil moisture", *Proceedings of IEEE*, vol. 62 no.1, pp.98-103, January 1974.

Mitigating Ground Clutter Effects for Mine Detection with Lightweight Artificial Dielectrics

Rob Linnehan¹ and Carey Rappaport²

Center for Subsurface Sensing and Imaging Systems
Department of Electrical and Computer Engineering
Northeastern University, 302 Stearns Building
Boston, MA 02115

Tel: ¹(617)-373-5555, ²(617)373-2043

E-mail: ¹rlinneha@ece.neu.edu, ²rappaport@neu.edu

Abstract

Ground surface roughness is problematic when using a radar impulse to detect and locate land mines. Waves scatter from a random rough ground surface in unpredictable ways, contributing to clutter that is particularly hard to suppress. This clutter has proven experimentally and computationally to distort and obscure the desired scattered field from a buried target. To overcome this effect we have developed a lightweight, artificial dielectric that can be placed over a chosen area that will mimic flat ground and mitigate clutter effects.

An artificial dielectric of close-packed array of small insulated metal-coated plastic spheres and lossless uniform plastic spheres can be formulated to match the dielectric properties soil. The ratio of these two spheres in the collection is adjusted to match a particular soil type and the moisture content. Placing them in a conformable bag and ensuring a flat upper interface with the air, ground reflections from an impulse radar can effectively be removed to reveal a target scattering signature. Furthermore, a matched filter can be used to distinguish between a landmine and a false alarm (such as a rock).

The artificial dielectric was matched by running experiments in the frequency and time domains. A 1 GHz center frequency impulse ground penetrating radar was used to collect time signals and compare different cases: flat ground, rough ground and rough ground with artificial dielectric. Results indicate excellent rough surface reflection removal and target signal enhancement.

Keywords

mine detection, artificial dielectrics, ground penetrating radar, coupling, impedance matching

Introduction

Detection of an anti-personnel land mine using ground penetrating radar (GPR) depends on recognizing a reflection from an object buried shallowly below the ground surface. Often the dielectric constant of the target materials (RTV, nylon, TNT, etc.) is in the range of dry soil¹. Mines buried in moist soil offer a slightly lower dielectric constant than their surroundings. This scattering from a low contrast object is compared to a ground signal without an object and a positive or negative decision is made.

The background reference signal that is compared to the test signal will usually be taken over flat ground without a target. If the test signal is obtained over flat ground, a reflection from an object can be seen by subtracting the background signal. However, in realistic environments the ground surface is not flat. Scatter resulting from a rough ground surface can obscure the desired scattered field from the buried target. The undesirable clutter, as it is known, is unpredictable and therefore difficult to remove in processing². Time gating the received signals is not feasible as the targets are usually buried very close to the ground surface³ (most AP mines are between 2cm and 10cm). The clutter will affect the ground reflection and disturbs the signal in the time range where the mine reflection is likely to occur.

Simulations were performed using a three dimensional finite difference time domain model³ to show the effects of rough ground on the propagating fields. A modulated Gaussian pulse was used as the source. Figure 1 displays the incident plane wave coming from the right at a 30 degree angle on flat ground, and the subsequent transmitted and reflected fields. Figure 2 illustrates how clutter is introduced into the reflected field when the ground is rough. The ground roughness follows a Gaussian distribution with a mean height of 1 cm and a correlation length of 10 cm⁷.

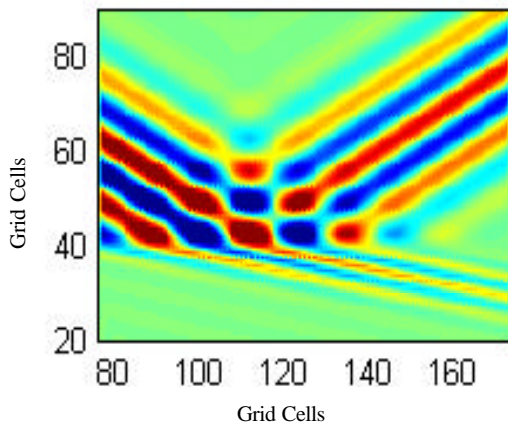


Figure 1: Propagating fields from flat ground

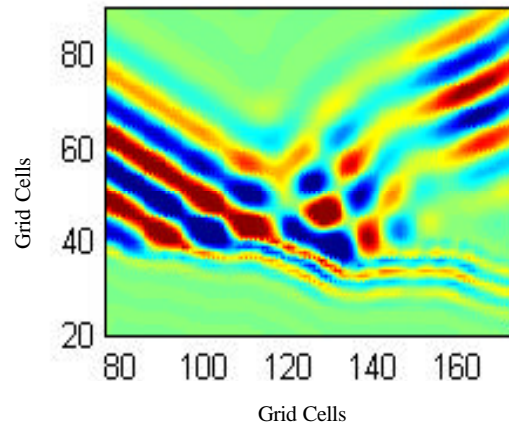


Figure 2: Propagating fields from rough ground

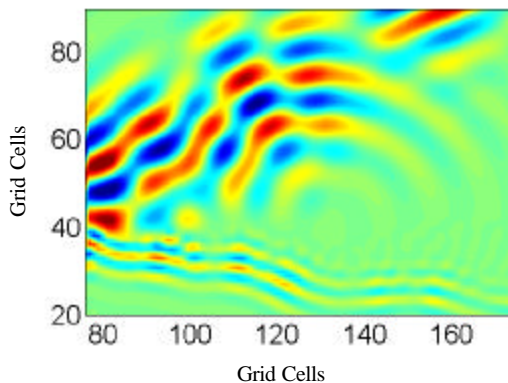


Figure 3: Isotropic effect develops from rough ground

Figure 3 shows the propagating fields of the same rough ground simulation as Figure 2 at a later time. There is no buried target in this simulation. Here we can see an isotropic effect develop near the center of the figure. This might be incorrectly interpreted as the return from a buried target. Thus rough ground clutter not only reduces the probability of detection but also increases the probability false alarms.

Although rough ground presents a complicated problem in detecting subsurface objects, a physical means is implemented to overcome this obstacle. A lightweight artificial dielectric is placed over a test area that will mimic flat ground and mitigate clutter effects. The material must be electromagnetically similar to the ground, taking into account the soil type and moisture content. Also it must physically conform well to the test area and retain a flat interface with the propagating fields from the transmitter. Experiments were performed to find a suitable artificial dielectric to match a particular soil. Subsequent experiments using ground penetrating radar have proven that the detection process is improved using the artificial dielectric.

Artificial Matching Layer

An artificial dielectric (AD) layer fills voids and lessens the effects of bumps in the ground surface. The purpose of this layer is to soften and planarize the interface rather than act as an impedance transformer between the air and the soil. The relative permittivity of the layer is for the most part constant and cannot be perfectly matched to the soil during use. The dielectric constant of soil is a function of the soil type, moisture content and density – the latter two can vary widely over a test area. The background reference signal is obtained using the AD over flat ground with a nominal moisture content and density. The background signal will have a reflection from the top, flat surface of the dielectric and a later reflection from the AD/ground interface. We will see that this second reflection is more stable under various rough surfaces when compared to the direct air/ground interface.

Conducting spheres that are spatially separated have been used as artificial dielectrics for many years⁴. However, solid metal spheres are too heavy for land mine use and separating them in a fixed matrix is clearly not feasible. The same goal is achieved with the use of lightweight plastic spheres with a thin metallic coating. The spheres also have an outer plastic film that isolates them from one another. These 6mm beads are available in bulk from an arts and crafts manufacturer as decorations.

For the GPR experiment a soil tank was built and filled with clay-loam topsoil. At 1 GHz, the complex dielectric constant of clay-loam will vary between 6 and 20 depending on the moisture content and the density. For conducting spheres, the effective impedance depends on the geometry of the array and the proximity of the spheres. The separation can be computed in terms of volume packing fraction of the spheres in space. The volume packing fraction is given as:

$$p = Kv_S / v_{TOTAL}$$

Where K is the number of spheres in a total volume (v_{TOTAL}) and v_S is the volume of a sphere ($\frac{4}{3}\pi r^3$).

The effective impedance (ϵ_r) can be approximated as a function of packing fraction by the following formulas⁴:

$$\left. \begin{matrix} \epsilon / \epsilon_0 \\ m / m_0 \end{matrix} \right\} = 1 + 3p/A(p)$$

and

$$h_r = \sqrt{\frac{m_r}{\epsilon_r}}$$

where for simple cubic (sc) packing,

$$A(p) = -1/R_1 - p + 1.3047R_3 p^{10/3} + 0.0723R_5 p^{14/3} - 0.5289R_3^2 p^{17/3} + 0.1526R_7 p^6$$

and for face centered cubic (fcc) packing,

$$A(p) = -1/R_1 - p + 0.0753R_3 p^{10/3} + 0.2420R_5 p^{14/3} + 0.0558R_3^2 p^{17/3} + 0.0231R_7 p^6$$

and,

$$R_n = \begin{cases} -1, & \text{for } \mathbf{e}_r \\ n/(1+n), & \text{for } \mathbf{m}_r \end{cases}$$

The effective impedance for both structures is plotted in Figure 4.

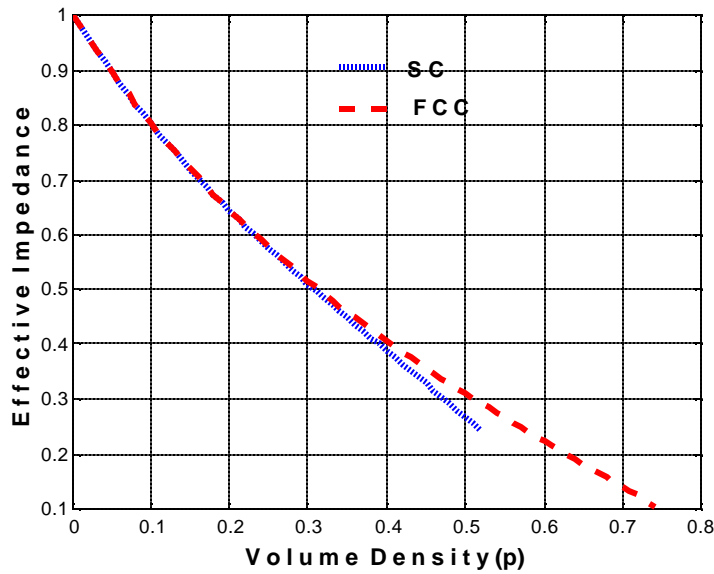


Figure 4: Effective Impedance vs. packing fraction

In practice we would expect the random distribution of spheres to favor that of an fcc structure. The volume density of the 6mm beads was found to be 0.565 which yields an approximate impedance of 0.26Ω . The impedance of soil is found using only the relative permittivity, with the permeability assumed to be that of free space. The impedance for clay loam usually varies between 0.22Ω (at 20% moisture) and $0.41\Omega^5$. We see that the impedance of the AD falls in the range of loamy soils with moisture contents of 12% to 20%.

For an impulse GPR system with a bandwidth of 700 MHz to 1.3 GHz the wavelength of the pulse in the dielectric varies between $33\text{ cm} < \lambda < 18\text{ cm}$. The values are significantly larger than the 6mm diameter spheres and therefore the impedance calculations assume frequency independence.

To better match the AD to a particular soil, the volume packing fraction is reduced, i.e. the spacing among the metallic beads is increased. This is done by uniformly distributing all-plastic beads among the conducting spheres. In order to ensure a uniform mixture of the AD, 8mm non-metallic beads were used because they are similar in weight and size. Measurements were taken in the frequency domain using a microwave network analyzer. Clay loam soil at ~15% moisture was compared to varying the ratios of the two types of beads. The contents were placed in a paraffin cavity and the S11 frequency response was recorded. Figure 5 shows the frequency response of the soil (solid dark) compared to two different ratios of the artificial dielectric. We see that a relatively good match (+) is achieved with the AD when the ratio is 1:2 (by volume 8mm/6mm). The response of dry sand is shown for contrast.

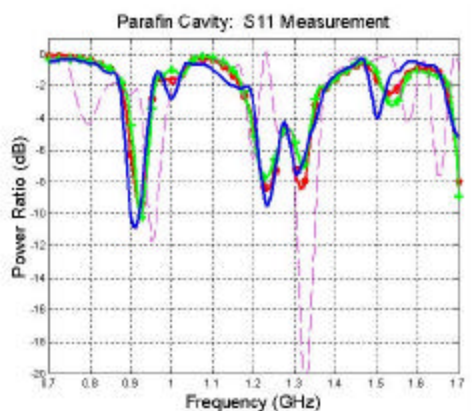


Figure 5: Frequency response of clay loam (solid dark), artificial dielectric (o,+) and sand (--)

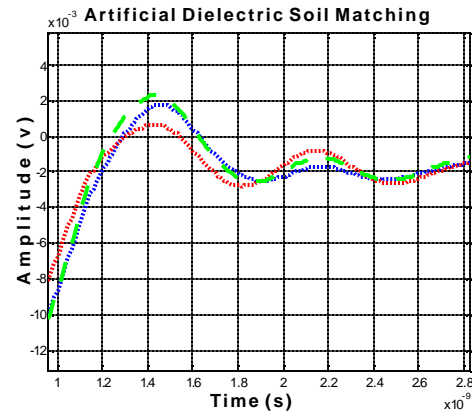


Figure 6: Flat ground (solid), roughened surface (dotted), AD fill (dashed)

A comparison was also performed in the time domain using ground penetrating radar. Looking at Figure 6 the solid line indicates flat ground, the dotted is ground with a roughened surface, and the dashed line is surface filled with AD. The ratio that best fit the flat ground response was also 1:2 (by volume 8mm/6mm).

Experiment

Ground penetrating radar experiments are performed at GeoCenters, Inc. in Newton, MA. An indoor wooden soil tank 3ft x 3ft x 2ft provides a controlled testing environment. The pulser (Picosecond Pulse Labs, Inc., model 1000D) excites the GPR system with a Gaussian pulse with a center frequency of 1 GHz and a -3dB bandwidth of 0.6 GHz. The transmitter and receiver are GeoCenters' 7 3/8" TEM horn rolled edge antennas. The rolled edge is implemented to minimize edge reflection, i.e. to better match the impedance of air. A diagram of the setup is seen below in Figure 7a along with a photograph in Figure 7b.

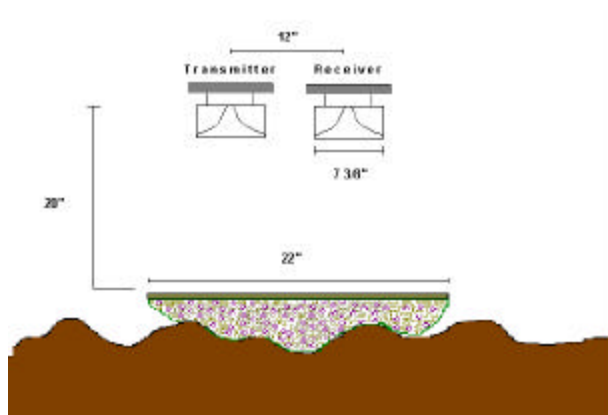


Figure 7a

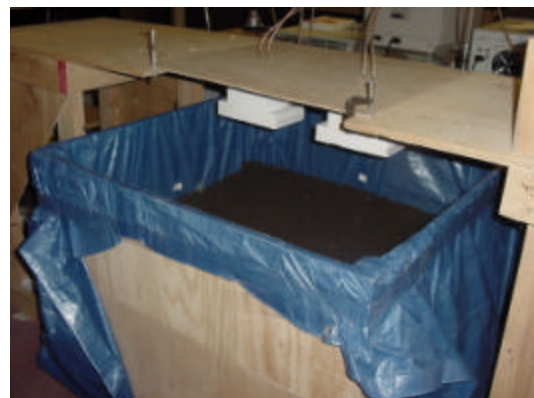


Figure 7b

The AD is placed in a conformable mesh bag and attached to a ¼” thick lossless Plexiglas (22in x 22in) to ensure a flat interface with the air. The 6” diameter target used in the experiment is an anti-personnel land mine with a material substitute for TNT that has similar electromagnetic properties.

The received signal can be modeled as follows:

$$r(t) = d(t) + n(t) + g(t-t_1) + c(t-t_2) + m(t-t_3)$$

Detection is based on deciding if the signal $m(t)$ is present or not in $r(t)$ among the other contributions. The targets are positioned below the antennas at varying depths and ground surfaces. The direct signal $d(t)$ is acquired by storing an air shot on the oscilloscope, making sure all objects that will reflect energy in the time range (a 4 ns window) of the experiment are removed. This is then subtracted from all test signals. The high frequency noise $n(t)$ is numerically suppressed using an FIR low-pass filter⁶ with a cutoff frequency $\omega_h = 0.06$. The flat ground signal $g(t)$ is a density-dependent version of the impulse that varies in delay and amplitude according to how the soil is packed (assuming that the moisture content is constant over a test period). The clutter $c(t)$ is dependent on the surface roughness and thus will vary in unpredictable ways. This distorts the return from the mine, $m(t)$, which, like $g(t)$, is also an attenuated and delayed version of the impulse.

The ground signal is removed by first obtaining a reference flat ground, no target signal. The permittivity and conductivity of the soil will change with density, which affects the amplitude of the ground impulse. Also the nominal distance between the antennas and the ground changes slightly throughout the test, affecting delay. As a result, a series of flat ground signals are acquired. The signals are aligned and scaled according to the average time reference and average amplitude of their peak values (the reflection of the impulse from the ground). These normalized signals are then averaged to produce a single background reference signal that will be compared to subsequent test signals.

This process is also performed to produce one background reference signal using the artificial dielectric. This signal is compared to test signals that use the AD.

Figure 8a shows samples of flat, no target signals (dotted) with the average background reference overlaid (solid). Figure 8b is the same for the AD background. The first peak in Figure 8b is the impulse return from the top of the AD. The second peak is the addition of the first pulse shape and the return from the AD/soil interface (resulting from the dielectric constants not being perfectly matched).

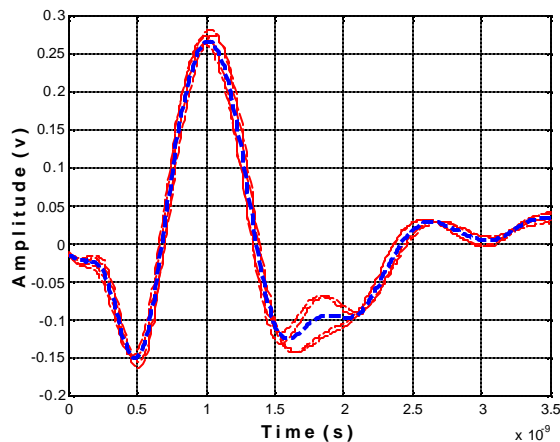


Figure 8a: Various flat ground reference signals (dotted); averaged background reference (solid)

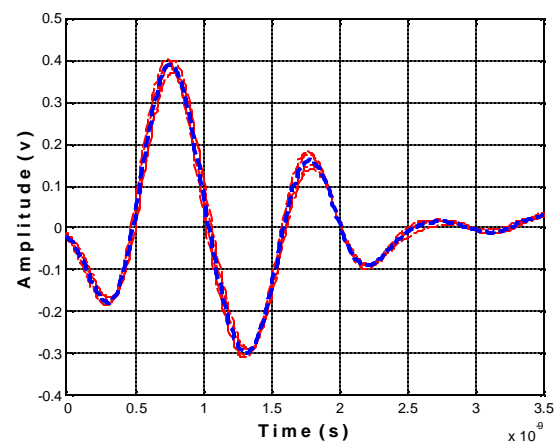


Figure 8b: Using the AD

A new test signal will be compared to the background signal. To account for soil density and height changes, the test signal is aligned to the background signal's peak amplitude. If clutter from the ground roughness is very low, subtracting the background reference should reveal a target reflection. Using the artificial dielectric we will be able to reduce this clutter enough on moderately rough surfaces to still yield a correct decision. The following figures illustrate the effect of two different rough ground cases. We see in Figure 9a the variation in the received rough ground signals (dashed) compared to the flat ground reference (solid). Signals of the same rough ground cases were compared to the reference using the AD and a significant reduction in variance is noted. The markers on the plots indicate the time range of interest, i.e. when a target reflection will show up if it is buried from 2 cm to 10 cm.

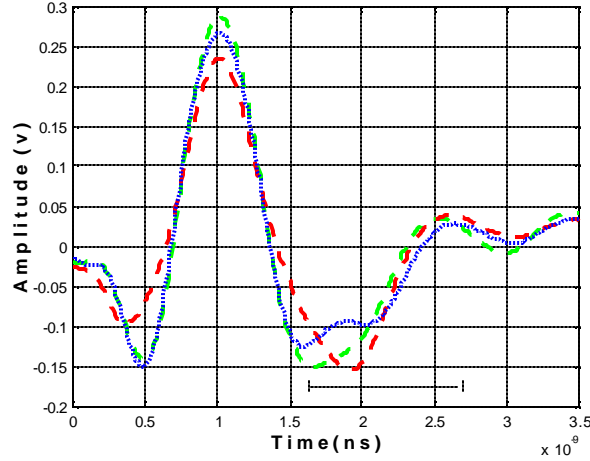


Figure 9a: Flat background reference (solid), two different rough ground signals (dashed)

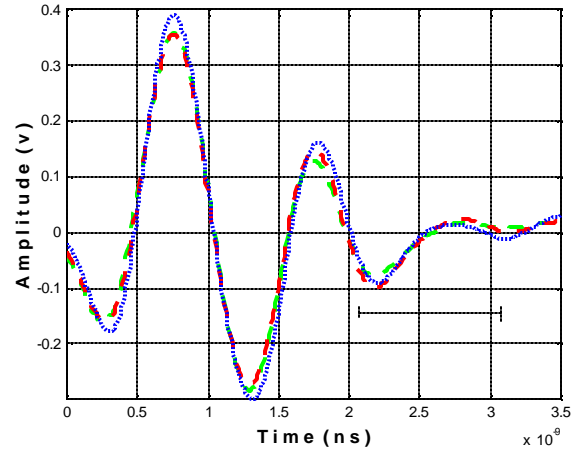


Figure 9b: Using the AD reduces the variance of rough ground signals compared to the reference

The cross correlation of the flat reference signal with the test signal, normalized to the square root of its energy can be found as follows⁷:

$$C_{fi}(m) = \frac{\sum_{n=1}^{N-|m|} S_f(n) S_i(n+m)}{\sqrt{\sum_{n=1}^N |S_f(n)|^2} \sqrt{\sum_{n=1}^N |S_i(n)|^2}}, \text{ for } m \geq 0$$

$$C_{fi}(m) = C_{if}(-m), \text{ for } m < 0$$

The maximum of the flat reference autocorrelation normalized by its energy will be unity. The difference between the maximum value of $C_{fi}(m)$ and 1 is a good metric to use for characterizing how well the rough ground test signal resembles the flat reference, i.e. the error between the reference and test signals. The average error when using the AD for 9 signals of varying ground roughness is found to be 0.0096, with a variance of 7.365×10^{-5} . When not using the dielectric the average is 0.0222 with a variance of 1.533×10^{-4} . The lower mean value and variance of this metric illustrates an improvement in the stability of the test signals when using the artificial dielectric. Thus the probability of detecting a slight signal variation due to a small, plastic object is enhanced.

A simple method for detection without the AD is given in the following diagram.

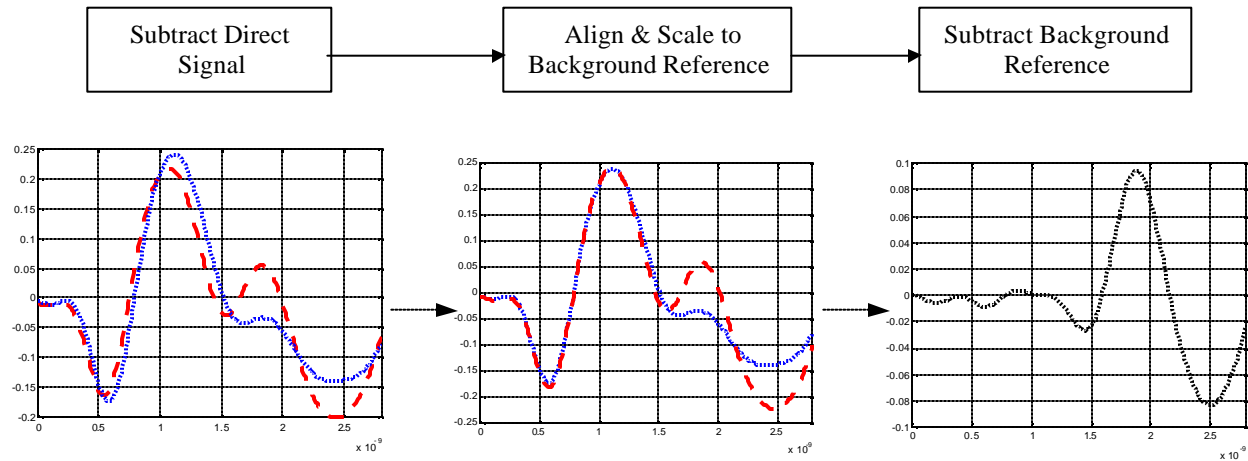


Figure 10: The impulse from a metal target is visible after background subtraction

The plots in Figure 10 show an example detecting a subsurface object using this method. The object is a metal sphere buried 5cm below flat ground. The average background signal (solid) is seen in the first two plots on the left along with the test signal (dashed). After scaling and aligning the background signal is subtracted, revealing an version of the impulse scattered from the object (scaled in the plot on the right).

The method becomes less effective when the object's dielectric constant is similar to the soil and when the ground surface is not flat. In Figure 11a we see a signal with a buried dielectric target (dashed) in flat ground, a buried target signal in rough ground (dotted) and a no-target signal in rough ground (dash-dot). Subtraction in Figure 11b reveals the target in flat ground but a determination of the same target in rough ground is much more challenging. Furthermore the no-target and target signals in rough ground are difficult to distinguish. In this case the chance of a false alarm is increased.

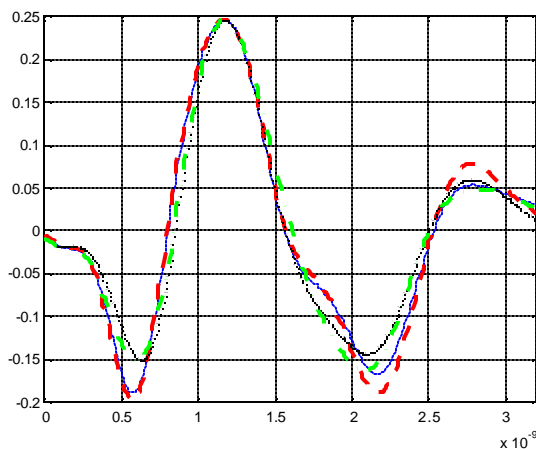


Figure 11a: Target signals in flat (dashed) and rough (dotted) ground; no-target signal in rough (dash-dot); background (solid)

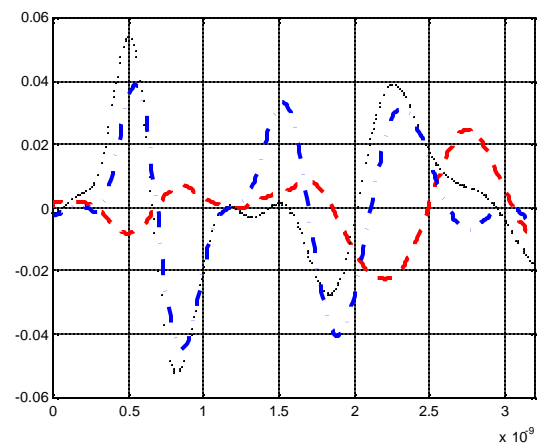


Figure 11b: Subtraction shows the plastic target at 2.75 ns in flat ground (dashed), but roughness obscures the target (dotted) and no-target (dash-dot) signals

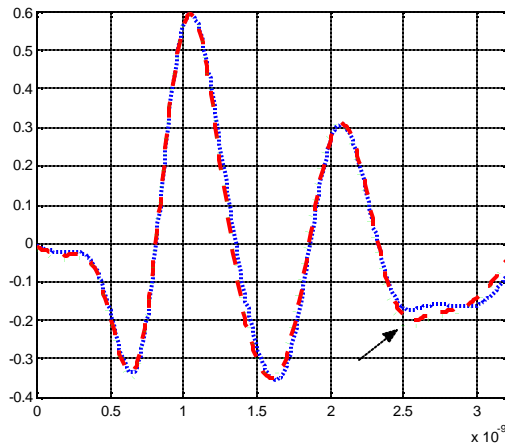


Figure 12: AD target signals; flat (dashed) and rough (dotted) compare similarly to the reference

Figure 13 displays the result of this algorithm with the object seen at 2.6 ns in flat (solid) and in rough ground (dashed and dotted). A test signal without a target (dash-dot) is also displayed to show a case where a false alarm can be avoided.

The procedure was tried for mines at various depths and preliminary assumptions show that the artificial dielectric offers improvement during cases of rough ground, especially at the more shallow target depths (2 cm – 5cm)

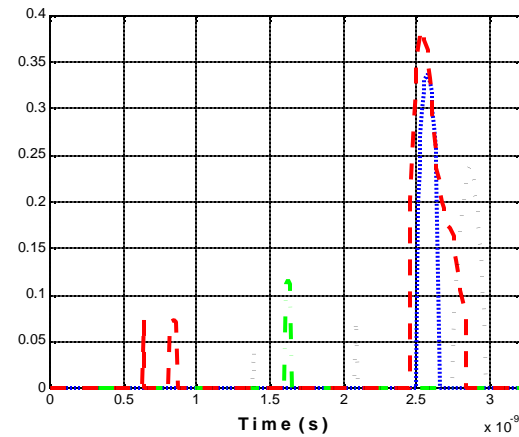


Figure 14: Target detected in flat ground (solid) and two rough ground cases (dashed, dotted); a no-target response is also visible (dash-dot)

Conclusions

The artificial dielectric placed over the ground area to be tested has proven to facilitate the mine detection process. This is true for several cases examined with the 6" diameter plastic AP mine buried between 2cm and 8cm in various rough ground interfaces. Even for flat ground it appears that the AD reduces whatever clutter is added to the received signal due to soil density changes.

As more data is collected one might expect improvements in the detection algorithm as the reduced clutter is better characterized and removed. Also receiver operating characteristics can be developed to optimize the probability of detection for a given false alarm rate.

Acknowledgements

This work was supported by the Army Research Office Multidisciplinary University Research Initiative Grant No. DAAG55-97-0013 and the Center for Subsurface Sensing and Imaging Systems (CenSSIS) at Northeastern University, under the Engineering Research Centers Program of the National Science Foundation (Award Number EEC-9986821). The experiments were performed with the help of Geo-Centers, Inc., which also designed the transmitting and receiving antennas. The FDTD simulations were provided by P. Kosmas of CenSSIS at Northeastern University.

References

1. J.T. Johnson, J. Jenwatanavet and N. Wang, "Soil modification studies for enhanced mine detection with ground penetrating radar", *Proceedings of SPIE*, A.C. Dubey, J.F. Harvey, J.T. Broach, **vol. 3710**, pp. 739-748, SPIE, Orlando, FL, 1999.
2. C. Rappaport, F. Biehold, R. Linnehan, "Mitigating Ground Clutter Effects with Lightweight Artificial Dielectrics", *Proceedings of SPIE*, A.C. Dubey, J.F. Harvey, J.T. Broach, V. George, **vol. 4394**, pp. 503-514, SPIE, Orlando, FL, 2001
3. P. Kosmas, "*Three-Dimensional Finite-Difference Time-Domain Modeling for Ground Penetrating Radar Applications*", MSc Thesis, Northeastern University, March 2002.
4. J. Lam, "Effective dielectric constant and magnetic permeability of a simple cubic lattices populated by two species of conducting magnetic spheres", *Journal of Applied Physics*, **vol. 66(8)**, pp. 3741-3749, October 1989
5. C. Rappaport, S. Winton, "Modeling Dispersive Soil for FDTD Computation by Fitting Conductivity Parameters", 12th Annual Review of Progress in Applied Computational Electromagnetics Symposium Digest, pp. 112-118, March 1997
6. J. Proakis, D. Manolakis, *Digital Signal Processing – Principles, Algorithms and Applications*, Ch. 8, Prentice Hall, Upper Saddle River, NJ, 1996
7. C. Rappaport, M. El-Shenawee, H. Zhan, "Suppressing GPR Clutter from Randomly Rough Ground Surfaces to enhance Nonmetallic Mine Detection"

Scattering from Multilayered Random Rough Surfaces Using the Steepest Descent Fast Multipole Method (SDFMM) and the Multiple Interaction Model

Magda El-Shenawee
Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701
Tel: 501-575-6582, Fax: 501-575-7967
magda@uark.edu

Abstract

Scattering of electromagnetic waves from multilayered random rough surfaces is crucial for subsurface sensing applications. A multiple interaction method of moments (MoM) model is used in this work to analyze scattering from two-dimensional multilayered random rough ground (3-D scattering problem) especially when the underground layer is deeply buried under the air/ground interface. The presented model removes a barrier and enables the application of the Steepest Descent Fast Multipole Method (SDFMM) to certain 3-D non-quasi-planar structures. The conventional SDFMM has been used to analyze electromagnetic wave scattering from quasi-planar structures where the scatterer's height is a fraction of a free-space wavelength. The presented model is based on multiple interactions mechanism between the air/ground interface and the buried underground layer. The basic idea of the proposed multiple scattering model is to decompose the non-quasi-planar multilayered ground into two quasi-planar scatterers where the conventional SDFMM can be applied separately to each one. The interactions between the sub-quasi-planar scatterers are calculated using the electromagnetic vector potentials near-field expressions. This model is tested and validated with the MoM on a variety of geometries. The results show that the strongest signature of the buried scatterer is mainly due to the first multiple interaction mechanism (ground-object-ground) while the contributions from repeating this mechanism become insignificant even for lossless and/or slightly lossy underground.

I. INTRODUCTION

Electromagnetic subsurface sensing has recently become an attractive research area due to its wide civil and defense applications. There are numerous unfound buried objects; these objects could be, but are not limited to, the location of underground water, gas and/or water lines, identifying cracks in asphalt roads, cancerous tumors in the human breast, plastic anti-personnel or metallic anti-tank mines, or the location of hazardous environmental wastes, etc. In reality, these complicated scattering problems have three-dimensions (x , y and z) that cannot be solved in closed forms but rather can be solved using computational techniques, in particular using fast algorithms. Few fast computational techniques were developed in the literature, e.g., the Fast Multipole Method (FMM) [1]-[3], the Steepest Descent Fast Multipole Method (SDFMM) [4]-[6], the Sparse Matrix/Canonical Grid Method (SMCG) [7], [8], and the Spectral Algorithm combined with the Forward-Backward Method (FB/NSA) [9]. These fast techniques have shown superiority regarding the CPU time and memory requirements over the conventional methods (e.g. the method of moment (MoM), the finite-difference methods (FD) and the finite-element method (FEM)). Recently, the SDFMM has been adopted to analyze the scattering from penetrable shallow objects buried under two-dimensional random rough ground [10]-[11]. The SDFMM has the great advantage of $O(N)$ computational complexity for both the CPU time and computer memory, where N is the total number of electric and magnetic surface current unknowns. However, there is a barrier that prohibits using the SDFMM in some potential applications; the scatterer should have a quasi-planar structure with total height equal to a fraction of a free-space wavelength. On the other hand, there are several potential geometries that have non-quasi-planar structures such as (i) multilayered ground where the burial depth of

the underground layer is larger than the wavelength [12], (ii) a non-quasi-planar object buried under the quasi-planar rough ground, (iii) utility cylindrical pipes deeply buried under the rough ground, etc.

The basic idea of the multiple scattering model is to decompose a non-quasi-planar structure into two quasi-planar scatterers as shown in Fig. 1 where the conventional SDFMM can be applied separately to each one. The interactions between the sub-quasi-planar scatterers, e.g. between the ground and the buried object in Fig. 1, are calculated directly using the electromagnetic vector potential expressions for the electric and magnetic near-fields [13]. The proposed model sheds light on the physics involved in the subsurface scattering mechanism [12], [14]. In this work, we are emphasizing on (i) presenting and validating the multiple scattering model on a variety of geometries as shown in Fig. 2 and (ii) using the SDFMM in the proposed multiple scattering model for non-quasi-planar structures aiming to speed up the calculations by exploiting the superior $O(N)$ computational complexity of the SDFMM.

The proposed model algorithm will be described in Section II, numerical results are presented in Section III and concluding remarks are given in Section IV.

II. FORMULATION

The four integral equations describing the unknown equivalent electric and magnetic surface currents for the problem of a single object buried beneath two-dimensional rough ground were derived in [10]-[11]. Upon applying Galerkin's method for testing and using the RWG vector basis functions for approximating the surface currents [15], these integral equations are transformed into a set of linear system of equations $\bar{\bar{Z}}\bar{\bar{I}} = \bar{\bar{V}}$ given by [10]:

$$\begin{pmatrix} \bar{Z}_{g.g} & \bar{Z}_{g.obj} \\ \bar{Z}_{obj.g} & \bar{Z}_{obj.obj} \end{pmatrix} \begin{pmatrix} \bar{I}_g \\ \bar{I}_{obj} \end{pmatrix} = \begin{pmatrix} \bar{V}_g \\ 0 \end{pmatrix} \quad (1)$$

where $\bar{Z}_{g.g}$ is a submatrix representing interactions between elements only on the ground surface; $\bar{Z}_{g.obj}$ is a submatrix representing interactions between elements on the ground surface and elements on the object surface; $\bar{Z}_{obj.g}$ is a submatrix representing interactions between elements on the object surface and elements on the ground surface; and $\bar{Z}_{obj.obj}$ is a submatrix representing interactions between elements only on the object surface. It was shown in [10] that the total matrix $\bar{\bar{Z}}$ has order of $2(N+P) \times 2(N+P)$, where N is the number of vector basis functions on the ground and P is the number of vector basis functions on the object. The factor of two is to account for both the electric and magnetic surface currents. Moreover, $\bar{Z}_{g.g}$ (and $\bar{Z}_{obj.obj}$) is exactly the impedance matrix obtained in the PMCHW integral equations [16]. The vector \bar{V}_g is composed of the tested tangential incident electric field \bar{E}_g^{inc} and the tested normalized magnetic field $\eta_1 \bar{H}_g^{inc}$ on the ground surface. The unknown current coefficients \bar{I}_g and \bar{I}_{obj} were solved for in [10] and [11] by applying the SDFMM directly to Eq. 1. Conversely, in this work, the multiple scattering model will be used to iteratively solve for the unknown current coefficients in

$$\bar{Z}_{g.g} \bar{I}_g^{(n)} = \bar{V}_g^{(n)} \quad (2a)$$

$$\bar{Z}_{obj.obj} \bar{I}_{obj}^{(n)} = \bar{V}_{obj}^{(n)} \quad (2b)$$

where $n = 1, 2, 3, \dots$ is the number of iterations or the number of multiple scattering mechanisms between the ground and the buried object as shown in Fig. 2b. The mechanism of the multiple scattering model begins by calculating the incident waves from the source (transmitting antenna) on the rough ground $\bar{V}_g^{(0)}$ with the assumption that there are no buried objects under the ground. Then the SDFMM is used to calculate the induced electric and magnetic surface currents on the ground as $\bar{J}_g^{(0)}$ and $\bar{M}_g^{(0)}$, respectively. In return, these currents are used to induce incident electric and magnetic fields on the surface of the

buried scatterer $\bar{V}_{obj}^{(1)}$, which are calculated directly using the near-field expressions [13]. Depending on the geometry of the buried object, the SDFMM or the MoM is used to calculate the induced electric and magnetic surface currents on the buried object, $\bar{J}_{obj}^{(1)}$ and $\bar{M}_{obj}^{(1)}$. Finally, these currents are used to induce electric and magnetic fields on the ground surface as $\bar{V}_g^{(1)}$ using [13], and the SDFMM can be used to solve for the induced electric and magnetic currents on the ground, $\bar{J}_g^{(1)}$ and $\bar{M}_g^{(1)}$, respectively. This completes one multiple scattering mechanism (e.g. ground-object-ground). The process is to be repeated till the solution of currents on the ground and on the buried object converges.

As will be shown in the following section, a significant reduction in the CPU and memory requirements is anticipated upon using the multiple scattering mode where only $\bar{Z}_{g,g}$ and $\bar{Z}_{obj,obj}$ are used in the iterative calculations while in the conventional SDFMM and/or the conventional MoM, the four matrices $\bar{Z}_{g,g}$, $\bar{Z}_{obj,obj}$, $\bar{Z}_{g,obj}$ and $\bar{Z}_{obj,g}$ are to be used. In many applications $\bar{Z}_{g,obj}$ and $\bar{Z}_{obj,g}$ have the same order as $\bar{Z}_{g,g}$ and $\bar{Z}_{obj,obj}$ as will be demonstrated in Section III.

III. NUMERICAL RESULTS

A variety of geometries represented by five examples are used in this section to test and validate the proposed multiple scattering model with the conventional MoM. The numerical results will present (i) the electric and magnetic surface currents on the ground, (ii) the electric and magnetic surface currents on the buried scatterer, and (iii) the scattered electric field above the ground due to just the buried scatterer (i.e., scatterer's signature). In all results presented in this section, the incident wave is assumed a Gaussian beam tapered towards the edges of the ground [17] with horizontally polarized incident electric field (i.e., in the y -direction) and due to just the air/ground interface as well. The half beam width of the beam is $L/5$ where the ground has dimensions $L \times L$.

In Example 1 shown in Fig. 2a-b, the flat ground has dimensions $3.04\lambda_o \times 3.04\lambda_o$ with an x -directed horizontal cylinder of length $b = 3.04\lambda_o$, radius $a = 0.15\lambda_o$, and burial depth $z = -0.65\lambda_o$ measured from its center, where λ_o is the free-space wavelength. The relative dielectric constant of the ground and the buried cylinder are assumed $\epsilon_{2r} = 2.5 - j0.18$ and $\epsilon_{3r} = 7.9 - j0.0029$, respectively. The incident angle in this example is $\vartheta^i = 0$. Excellent agreement between the multiple scattering model and the MoM for the electric and magnetic surface currents on the flat ground and on the buried cylinder is shown in Figs. 3a-d. For qualitative comparisons, the magnitudes of the surface currents are plotted versus the y -direction at $x = 1.52\lambda_o$ in these figures. Four multiple scattering mechanisms, as described in Fig. 2b, are used to obtain these results. Similar agreement is observed for results plotted versus the x -direction at $y = 1.52\lambda_o$. An insignificant error is observed in Figs. 3c and 3d that could be due to the error introduced in approximating the cylindrical surface with triangular patches. As shown in these figures, the presence of the buried cylinder caused a significant change in the initial ground currents $\bar{J}_g^{(0)}$ and $\bar{M}_g^{(0)}$ only after the first multiple scattering mechanism. Similarly, Fig. 4a shows the convergence of the magnetic surface current on the upper half of the buried cylinder, plotted versus the x -direction at $y = 1.52\lambda_o$, while Fig. 4b shows the convergence of the magnetic surface current on both the upper and lower halves of the buried cylinder, plotted versus the y -direction at $x = 1.52\lambda_o$. The results confirm that the surface currents converge after only one ground-object-ground scattering mechanism, as described in Fig. 2a.

In Example 2, the geometry is a multilayered flat ground as shown in Fig. 2c, where the ground has dimensions $3.04\lambda_o \times 3.04\lambda_o$ and the underground layer has burial depth of $z = -0.45\lambda_o$. The relative dielectric constant of the ground and the underground layers are assumed $\epsilon_{2r} = 2.5 - j0.18$ and

$\epsilon_{3r} = 4.2 - j0.29$, respectively. The incident angle in this example is $\vartheta^i = 10^\circ$. All results show excellent agreement with the MoM (not presented here).

In Example 3, the geometry is a multilayered rough ground with flat underground layer as shown in Fig. 2d. The ground dimensions, the burial depth of underground layer, the relative dielectric constants, and the incident angle are the same as in example 2. The roughness parameters of the upper rough air/ground interface are given by the rms height and the correlation length of the random rough surface as $\sigma = 0.08\lambda_0$ and $l_c = 0.5\lambda_0$, respectively. All comparisons with the MoM strongly validate the multiple scattering model used in this example similar to previous examples (not presented here).

In Example 4, the geometry is flat ground with a buried sphere as shown in Fig. 2e where the dimensions of the ground are $4.08\lambda_0 \times 4.08\lambda_0$. The sphere has radius of $a = 0.5\lambda_0$ and is buried at $z = -0.75\lambda_0$ measured from its center. The relative dielectric constants of the ground and the buried sphere are assumed $\epsilon_{2r} = 2.5 - j0.18$ and $\epsilon_{3r} = 4.5 - j0.029$, respectively. The incident angle in this example is $\vartheta^i = 0$. In Figs. 5a and 5b, the magnitude of magnetic surface current on the air/ground interface is plotted versus the x -direction at $y = 2.04\lambda_0$. Three solutions are presented in Figs. 5a; (i) solution obtained using the conventional MoM for the whole scatterer (hollow circles), (ii) solution obtained by employing the MoM in the multiple scattering model (solid circle), (iii) solution obtained by employing the SDFMM in the multiple scattering model (+ symbol). The results show excellent agreement between these three methods. Moreover, validations based on the scattered electric fields observed above the ground at $z = 0.5\lambda_0$ and due just to the buried sphere are shown in Fig 5b where results are plotted versus the x -direction at $y = 2.04\lambda_0$. Excellent agreement between the three methods is also observed in these figures. However a slight difference between the multiple scattering model and the conventional MoM results is observed in Fig. 5b. This slight difference can be attributed due to the definition of the scattered fields due to just the buried object in both methods. In the conventional MoM method we calculated the total scattered electric fields twice; once with the buried sphere and once without the buried sphere, then the results are subtracted from each other using complex vectors [10]-[11]. On the other hand, for the multiple scattering model we use the obtained surface electric and magnetic currents on the ground due to only the presence of the sphere, i.e., $\bar{J}_g^{(1)} + \bar{J}_g^{(2)} + \bar{J}_g^{(3)}$ and $\bar{M}_g^{(1)} + \bar{M}_g^{(2)} + \bar{M}_g^{(3)}$, and use the near-field expression in [13] to compute the scattered electric fields. Notice that the quantities $\bar{J}_g^{(0)}$ and $\bar{M}_g^{(0)}$ represent the surface currents on the ground with the assumption that there is no buried objects, i.e., only the ground is present and are not used to calculate the signature of the buried sphere.

In Example 5, the geometry is a multilayered rough ground as shown in Fig. 2f where the dimensions of the ground and the incident angle are the same as in example 4. The underground rough layer has a burial depth of $z = -0.95\lambda_0$ measured from its mean plane to the rough air/ground interface mean plane. The relative dielectric constants of the ground and the underground layer are assumed $\epsilon_{2r} = 2.5 - j0.18$ and $\epsilon_{3r} = 3.7 - j0.2$, respectively. The roughness parameters of the air/ground interface are $\sigma_1 = 0.06\lambda_0$ and $l_{c1} = 0.5\lambda_0$ while they are $\sigma_2 = 0.05\lambda_0$ and $l_{c2} = 0.4\lambda_0$ for the underground layer. In Fig. 5a, the magnitudes of the magnetic surface currents on both the rough air/ground interface and the underground rough layer are plotted versus the y -direction at $x = 2.04\lambda_0$. The results show an excellent validation between the model using the MoM and the model using the SDFMM. In Fig. 5b, the scattered electric fields observed above the ground at $z = 0.5\lambda_0$ are plotted versus the y -direction at $x = 2.04\lambda_0$. The scattered fields due to just the rough air/ground interface are calculated using only the surface currents $\bar{J}_g^{(0)}$ and $\bar{M}_g^{(0)}$ while the scattered fields due to just the underground layer are calculated using only the

surface currents $\bar{J}_g^{(1)} + \bar{J}_g^{(2)} + \bar{J}_g^{(3)}$ and $\bar{M}_g^{(1)} + \bar{M}_g^{(2)} + \bar{M}_g^{(3)}$. As expected, the results in Fig. 5b show that the signature of the air/ground interface is significantly larger than that of the underground layer. Moreover, the results of this example clearly confirm the previous observations that the solution converges after the first multiple scattering mechanism (i.e., air/ground interface-underground layer-air/ground interface).

It should be mentioned that the relative residual error used in the TFQMR solver is 10^{-5} in all results of this section. For efficient results, the multiple scattering model should not be used when the buried objects are very shallow due to the inaccuracy of the near-field expressions in [13], but instead, the complete SDFMM, which has been successfully used in analyzing these structures [10]-[11], should be used.

IV. CONCLUSIONS

A new multiple scattering model to compute the signature of non-shallow objects buried under the rough ground (3-D scattering problem) is presented and validated in this work. The advantages of this model is (i) superior computational requirements than using the conventional MoM especially when the SDFMM is employed in the model and (ii) removing a barrier in the conventional SDFMM since it should be only applied to quasi-planar structures. It has been shown in this work that upon using the proposed multiple scattering model, certain potential non-quasi-planar structures, e.g., three multilayered ground, are decomposed into two quasi-planar layers where the SDFMM is separately applied. Interestingly, all obtained results show that both the electric and magnetic surface current solutions converge after only a single multiple scattering mechanism (i.e., from air/ground interface to buried object and then back to air/ground interface). This assures that no extra CPU time is needed to achieve the correct solutions.

ACKNOWLEDGMENTS

This research was sponsored by the Northeastern University's Demining MURI grant # DAA 0-55-97-0013 and in part by the College of Engineering at the University of Arkansas. The SDFMM was originally developed by V. Jandhyala, E. Michielssen and W. Chew at the UIUC.

References

- [1] V. Rokhlin, "Rapid solution of integral equations of scattering theory in two dimensions," *J. Comput. Phys.*, vol. 36, pp. 414-439, 1990.
- [2] C. C. Lu and W. C. Chew, "A multilevel fast-algorithm for solving a boundary integral equation of wave scattering," *Microwave Opt. Tech. Let.*, vol. 7, pp. 466-470, July 1994.
- [3] N. Geng, A. Sullivan and L. Carin, "Multilevel fast-multipole algorithm for scattering from conducting targets above or embedded in a lossy half space," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 4, pp. 1561-1573, July 2000.
- [4] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [5] V. Jandhyala, E. Michielssen, B. Shanker and W.C. Chew, "A combined steepest descent-fast multipole algorithm for the fast analysis of three-dimensional scattering by rough surfaces," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, no. 3, pp. 738-748, May 1998.
- [6] M. El-Shenawee, V. Jandhyala, E. Michielssen and W. C. Chew, "The steepest descent fast multipole method (SDFMM) for solving combined field integral equation pertinent to rough surface scattering," *Proc. of the IEEE APS/URSI '99 conf.*, Orlando, Florida, pp. 534-537, July 1999.
- [7] S. Li, C. H. Chan, L. Tsang, Q. Li, and L. Zhou, "Parallel implementation of the sparse matrix/canonical grid method for the analysis of two-dimensional random rough surfaces (three-dimensional scattering problem) on a Beowulf system," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 4, pp. 1600-1608, July 2000.

- [8] G. Zhang, L. Tsang and K. Pak, "Angular correlation function and scattering coefficient of electromagnetic waves scattered by a buried object under a two-dimensional rough surface," *J. Opt. Soc. Am. A*, vol. 15, no. 12, pp. 2995-3002, December 1998.
- [9] D. Torrungrueng, H. Chou and J. T. Johnson, "A novel acceleration algorithm for the computation of scattering from two-dimensional large scale perfectly conducting random rough surfaces with the forward-backward method," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 1656-66, July 2000.
- [10] M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, "3-D subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces: Use of the steepest descent fast multipole method (SDFMM)," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 6, pp. 1174-1182, June 2001.
- [11] M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo simulations of electromagnetic wave scattering from random rough surface with 3-D penetrable buried object: Mine detection application using the SDFMM," *J. Opt. Soc. Am. A*, to appear in December 2001.
- [12] E. Bahar and Y. Zhang, "Diffuse like and cross-polarized fields scattered from irregular layered structures-full wave analysis," *IEEE Trans. Anten. and Propag.*, vol. 47, no. 5, pp. 941-948, 1999.
- [13] C. A. Balanis, *Advanced engineering electromagnetics*, John Wiley&Sons Inc., pp. 254-309, 1989.
- [14] J. T. Johnson, "A study of the four-path model for scattering from an object above a halfspace," to appear, *Microwave Opt. Tech. Letters*, July 2001.
- [15] S. M. Rao, D. R. Wilton and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Anten. and Propag.*, vol. AP 30, no. 3, pp.409-418, May 1982.
- [16] L. Medgyesi-Mitschang, J. Putnam and M. Gedera, "Generalized method of moments for three-dimensional penetrable scatterers," *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1383-1398, April 1994.
- [17] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.

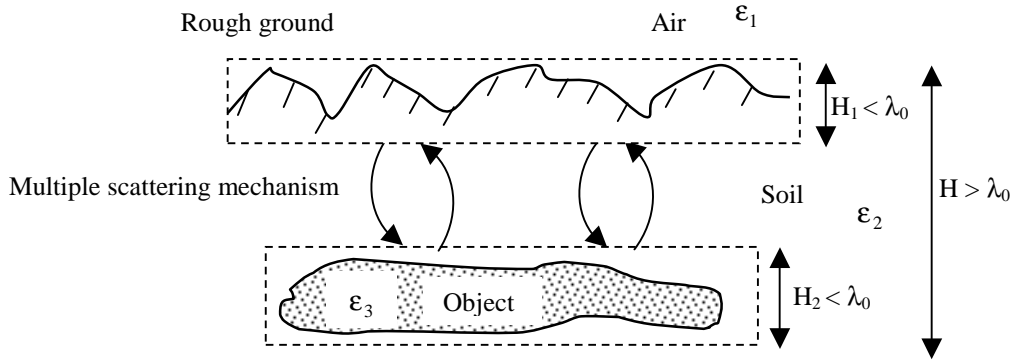


Figure 1. Decomposition of non-quasi-planar structure into two quasi-planar structures showing their interactions.

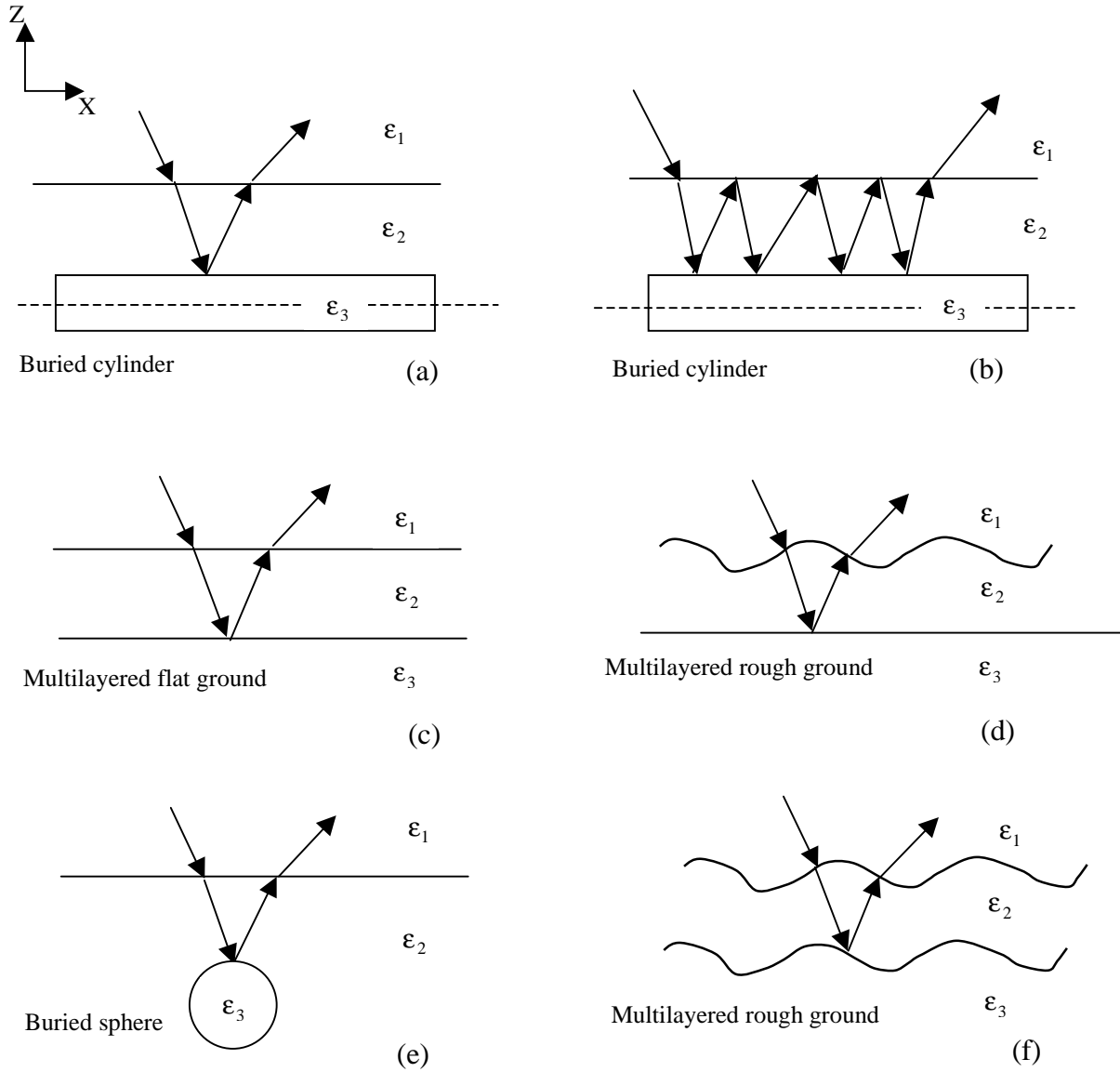


Figure 2. Multiple scattering mechanism between (a) the air/ground interface and a buried horizontal cylinder showing one ground-object-ground mechanism, i.e., $n = 1$ (b) the same as in (a) but showing four ground-object-ground mechanisms, i.e., $n = 4$, (c) the air/ground flat interface and an underground flat layer, (d) the air/ground rough interface and an underground flat layer, (e) the flat air/ground interface and a buried sphere, (f) the air/ground rough interface and an underground rough layer.

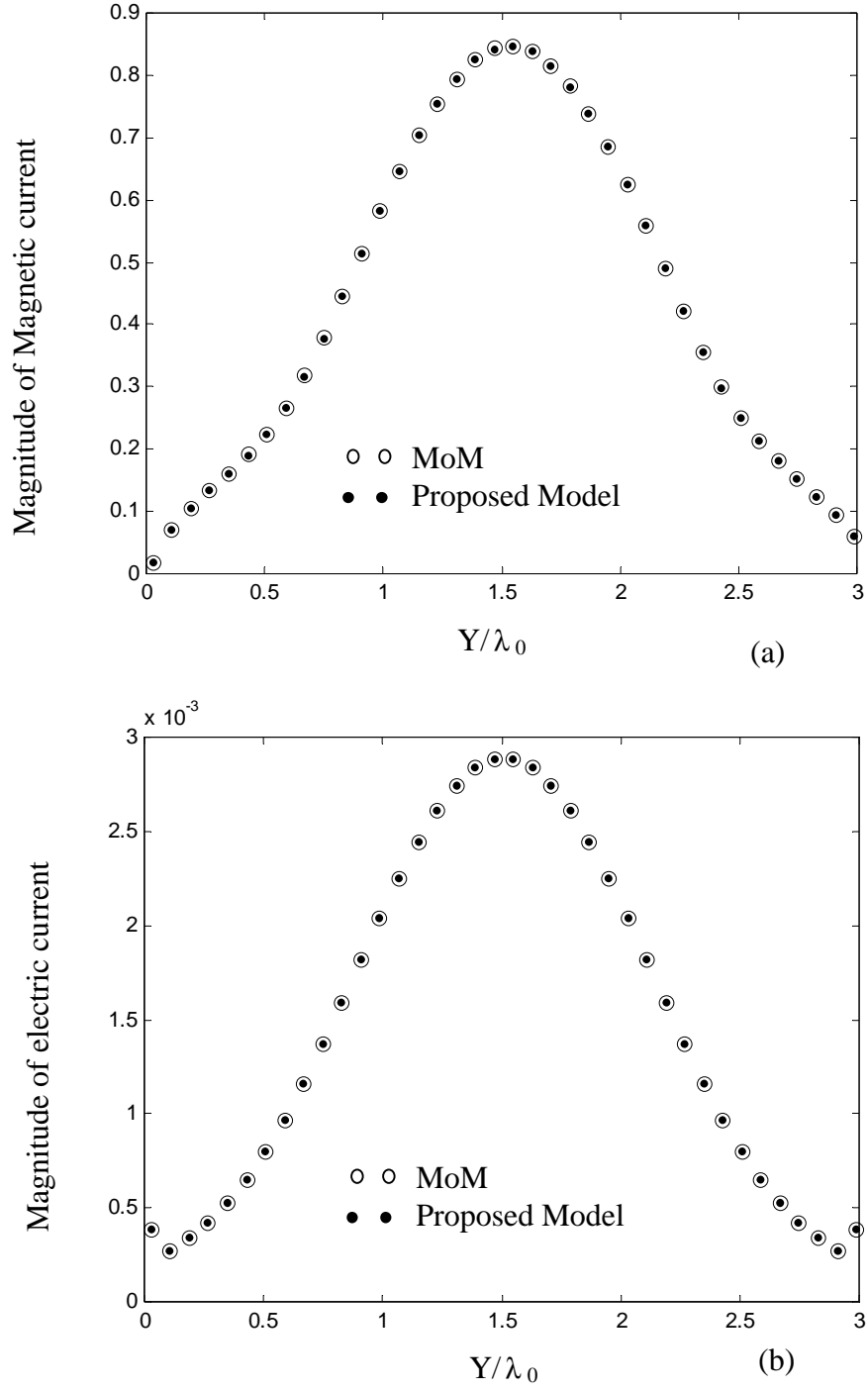


Figure 3. Magnitude of surface current on the flat air/ground interface shown at $x=1.52\lambda_0$ for (a) magnetic current $|\overline{M}|$, (b) electric current $|\overline{J}|$. Comparison between the proposed model and the conventional MoM. Data are for Example 1 (Figs.2a and b).

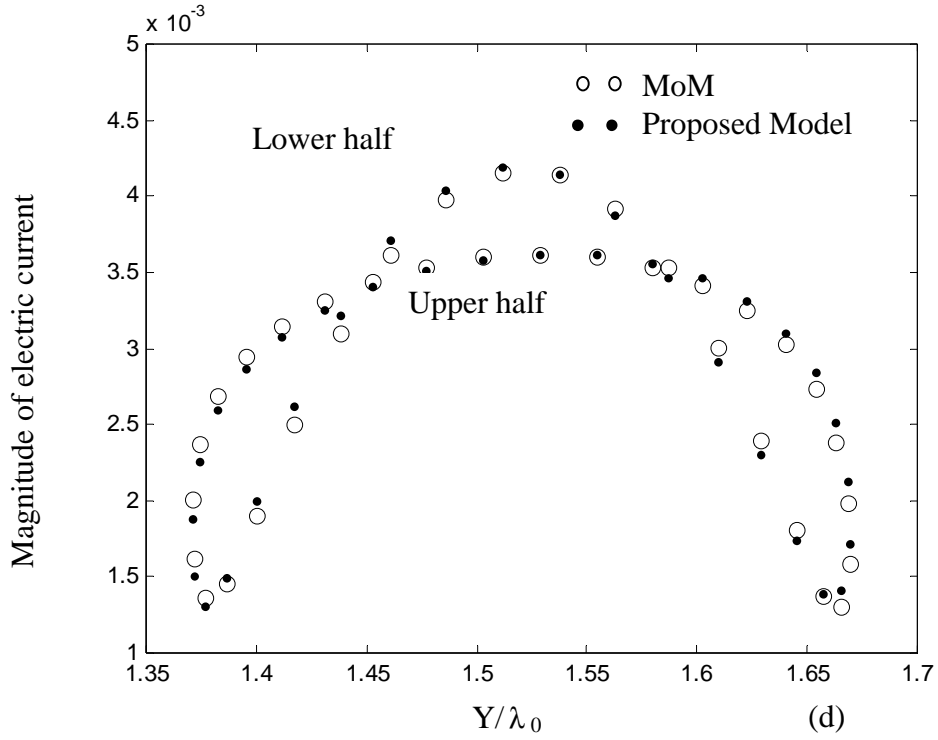
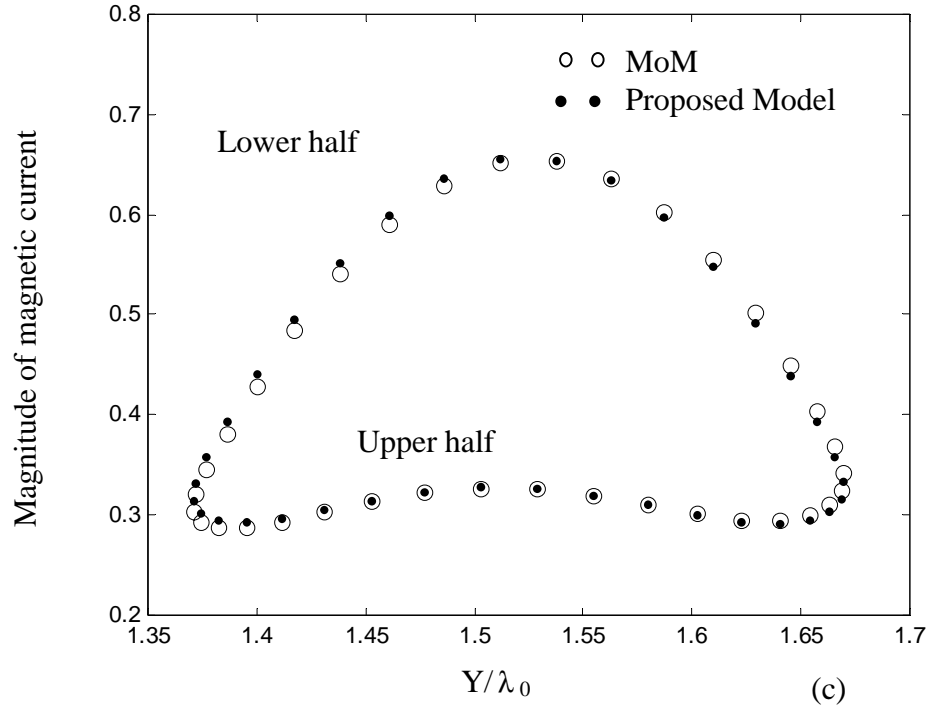


Figure 3. Magnitude of surface current on the buried cylinder shown at $x = 1.52\lambda_0$ for (c) magnetic current $|\overline{M}|$, (d) electric current $|\overline{J}|$. Comparison between the proposed model and the conventional MoM. Data are for example 1.

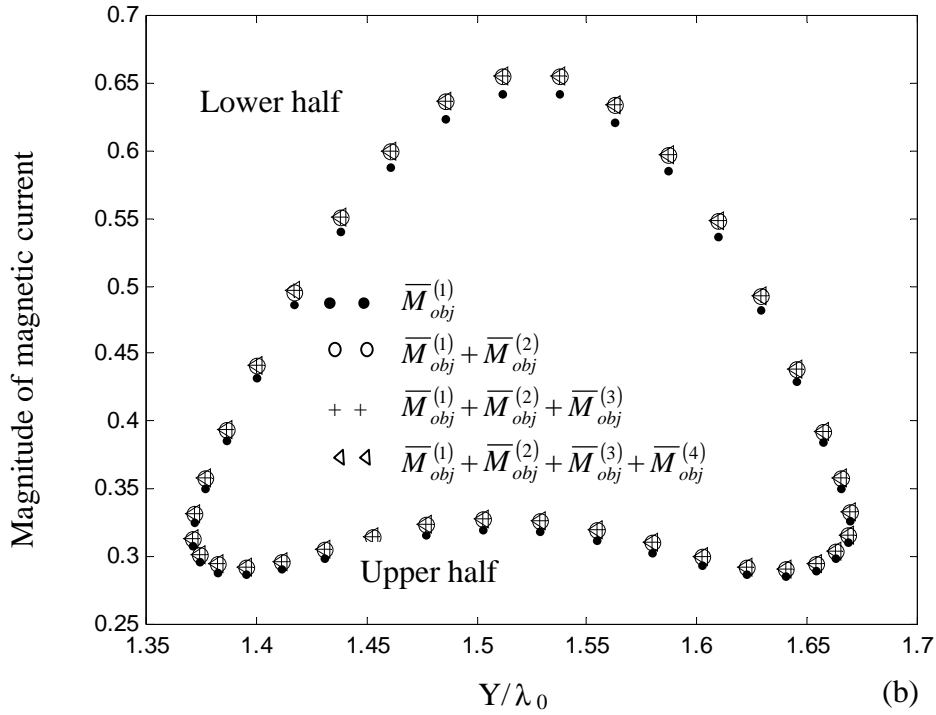
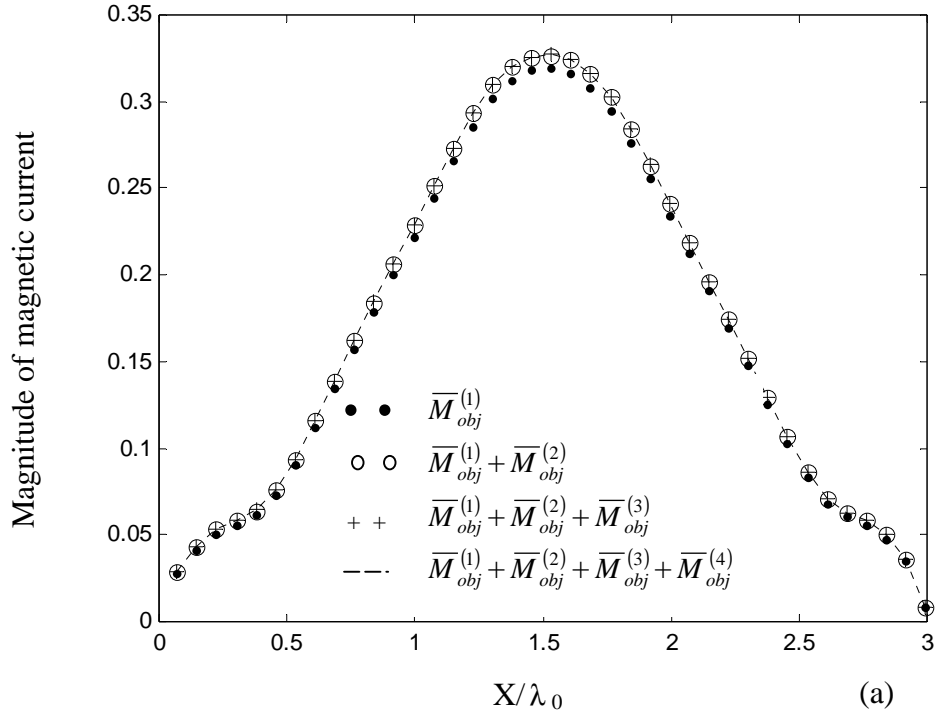


Figure 4. Convergence of magnetic surface current solution on (a) the upper half of the buried cylinder shown at $y = 1.52\lambda_0$, (b) on the buried cylinder shown at $x = 1.52\lambda_0$. Data are for Example 1.

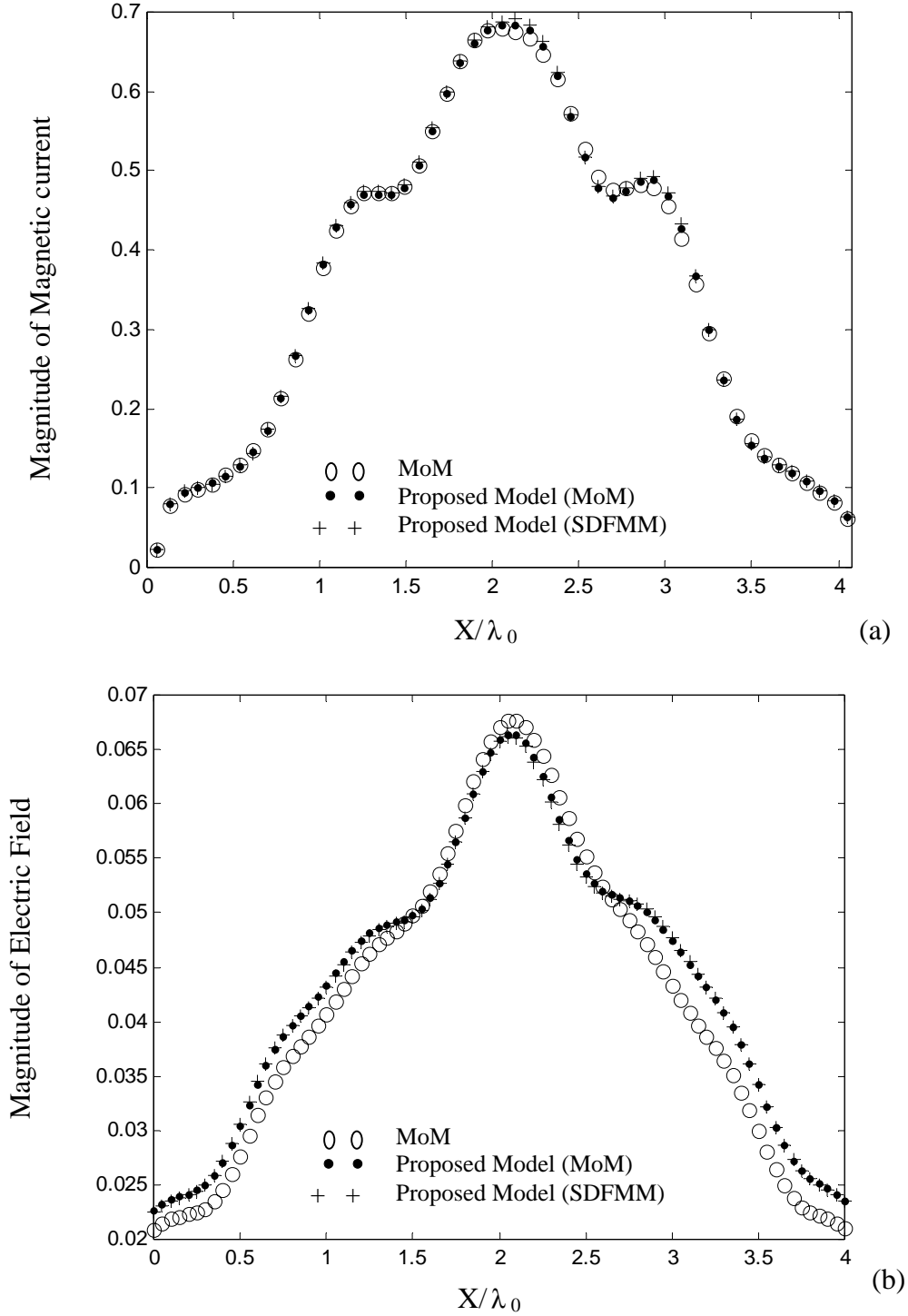


Figure 5. (a) Magnitude of magnetic surface current $|\overline{M}|$ on the flat air/ground interface, (b) Magnitude of scattered electric field $|\overline{E}|$ due to just the buried sphere observed at $z = 0.5\lambda_0$ above the ground. All are shown at $y = 2.04\lambda_0$. Comparison between the conventional MoM, the proposed model using both the MoM and the SDFMM. Data are for Example 4 (Fig. 2e).

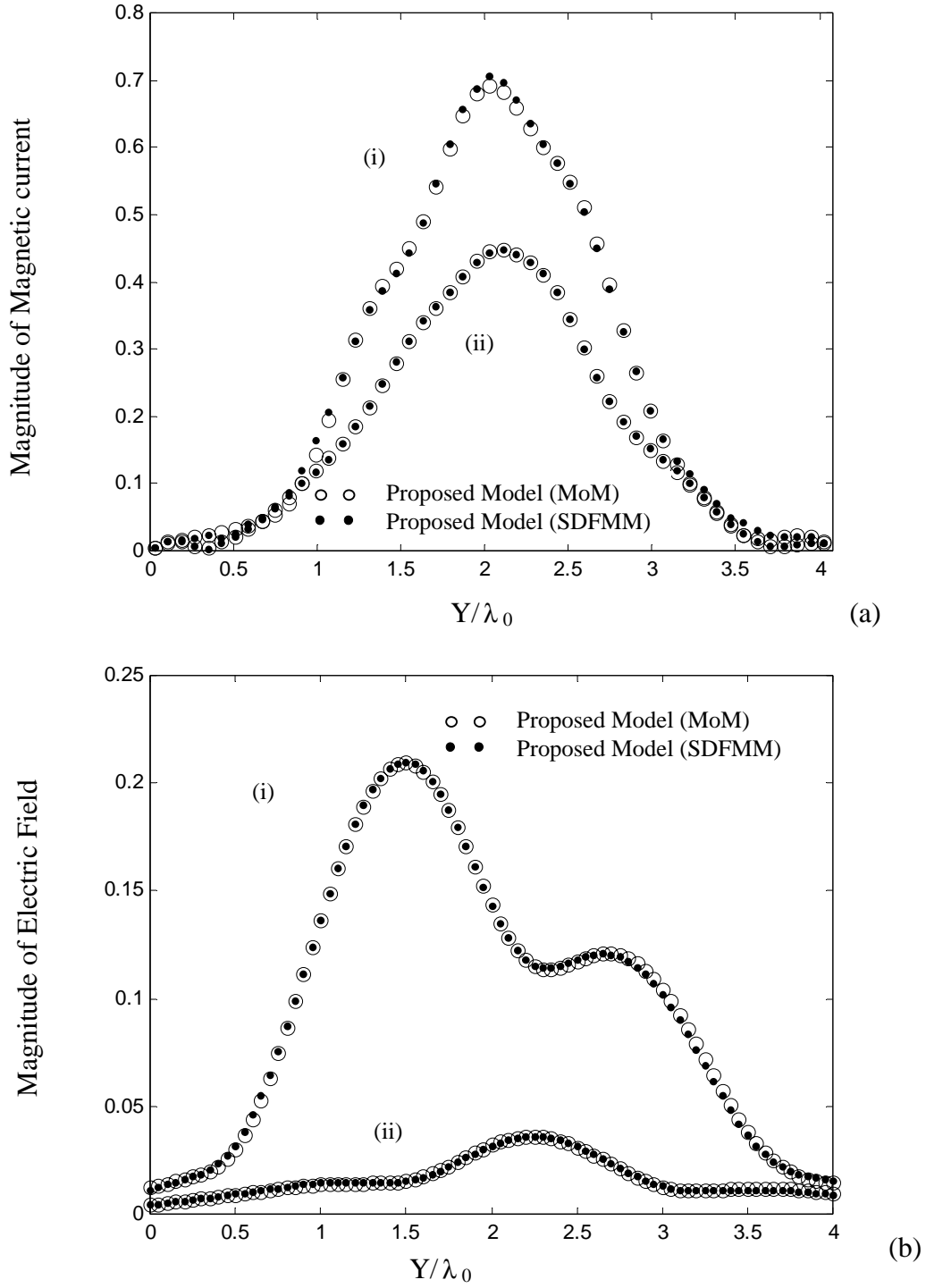


Figure 6. (a) Magnitude of magnetic surface current $|\overline{M}|$ on (i) the rough air/ground interface and (ii) the underground rough layer buried at $z = -0.95\lambda_0$, (b) Magnitude of scattered electric field $|\overline{E}|$ at $z = 0.5\lambda_0$ due to (i) just the rough air/ground interface (ii) just the rough underground layer. Comparison between the proposed model using both the MoM and the SDFMM. All are shown at $x = 2.04\lambda_0$. Data are for Example 5 (Fig. 2f).

Electromagnetics Computations Using the MPI Parallel Implementation of the Steepest Descent Fast Multipole Method (SDFMM)

M. El-Shenawee¹, C. Rappaport², D. Jiang², W. Meleis² and D. Kaeli²

¹Department of Electrical Engineering

University of Arkansas

Fayetteville, AR 72701

magda@uark.edu

²Department of Electrical and Computer Engineering

Northeastern University

Boston, MA 02115

rappaport@neu.edu

ABSTRACT

The computational solution of large-scale linear systems of equations necessitates the use of fast algorithms but is also greatly enhanced by employing parallelization techniques. The objective of this work is to demonstrate the speedup achieved by the MPI (Message Passing Interface) parallel implementation of the Steepest Descent Fast Multipole Method (SDFMM). Although this algorithm has already been optimized to take advantage of the structure of the physics of scattering problems, there is still the opportunity to speed up the calculation by dividing tasks into components using multiple processors and solve them in parallel. The SDFMM has three bottlenecks ordered as (1) filling the sparse impedance matrix associated with the near-field Method of Moments interactions (MoM), (2) the matrix vector multiplications associated with this sparse matrix (3) the far field interactions associated with the fast multipole method. The parallel implementation task is accomplished using a thirty-one node Intel Pentium Beowulf cluster and is also validated on a 4-processor Alpha workstation. The Beowulf cluster consists of thirty-one nodes of 350MHz Intel Pentium IIs with 256 MB of RAM and one node of a 4x450MHz Intel Pentium II Xeon shared memory processor with 2GB of RAM with all nodes connected to a 100 BaseTX Ethernet network. The Alpha workstation has a maximum of four 667MHz processors. Our numerical results show significant linear speedup in filling the sparse impedance matrix. Using the 32-processors on the Beowulf cluster lead to achieve a 7.2 overall speedup while a 2.5 overall speedup is gained using the 4-processors on the Alpha workstation.

INTRODUCTION

The calculation of the scattered electric and magnetic fields from a three-dimensional problem using the conventional techniques (e.g., the Method of Moments

(MoM), the Finite Element Method (FEM), or the Finite Differences in the time or frequency domains (FDTD or FDFD)) is a computationally intensive undertaking, especially for soils having a large dielectric constant. Moreover, the computational complexity of the problem dramatically increases upon inserting penetrable objects under the rough ground. Therefore, there was a necessity to use the fast computational algorithms to deal with this complex scenario. There exist few fast algorithms in the literature: the Fast Multipole Method (FMM) [1]-[3]; the SDFMM [4]-[6]; and the Sparse Matrix/Canonical Grid Method (SMCG) [7]-[8]. Basically, the standard FMM, the SDFMM, and the SMCG fast methods have the great advantage of converting the dense matrix obtained using the MoM into a sparse matrix leading to a dramatic reduction in the CPU time and computer memory requirements. In addition the fast algorithm, the Spectral Algorithm combined with the Forward-Backward Method (FB/NSA) [9], has shown to be an efficient iterative MoM for 3-D scattering problems.

In this work we adopted the SDFMM due to its superiority over the other fast algorithms in treating quasi-planar structures. The SDFMM is an integral equation-based fast algorithm that is a hybridization of (1) the Method of Moments (MoM), (2) the Fast Multipole Method (FMM), (3) the Steepest Descent Integration path (SDP) [4]-[6]. Recently the SDFMM has been successfully implemented to handle subsurface sensing applications, in particular, the scattering from a landmine modeled as a PEC and/or penetrable spheroid buried under a two dimensional randomly rough ground [10]-[11]. The SDFMM has computational complexity for the CPU time and for the memory requirement equal to only $O(N)$ per iteration versus $O(N^2)$ for the MoM, where N is the total number of the unknowns [4]-[6]. The reduced complexity of the SDFMM over several other computational electromagnetics techniques has helped in achieving a

fast and successful running for the Monte Carlo simulations [11]. However, the Monte Carlo sample needs in some cases to be greatly increased, e.g. when the ground random roughness increases the size of the Monte Carlo sample needs to be increased to achieve a converging solution. This could dramatically increase the required run time, especially when the dielectric constant of the ground is large and/or the penetrable buried object is electrically large. This necessitates more acceleration to the SDFMM computer code by using the MPI parallel implementation [8],[12],[13].

In this work, we used the MPI library for the parallel implementation of the SDFMM code [14]-[15]. The advantage of using the Beowulf cluster is that the system can be completely dedicated to the parallelization task, which is demonstrated in this work by executing small-scale cases due to memory limitations. Our emphasis is to demonstrate the overall speedup that can be achieved using the thirty-two processors. Porting the parallelized code to the national supercomputers, where hundreds of processors and adequate RAM are available, will potentially facilitate the computations of large-scale problems.

PARALLELIZATION METHODOLOGY

The SDFMM makes use of the equivalence theorem to calculate the electric and magnetic fields inside and outside a 3-D penetrable object buried under the rough surface interface [10]-[11]. The 3-D arbitrary object is modeled by scatterer R_3 that is immersed in scatterer R_2 which represents the rough ground which is immersed in the free space region represented by R_1 . The three regions, R_1 , R_2 and R_3 have permittivity and permeability given by ϵ_1 and μ_1 , ϵ_2 and μ_2 , and ϵ_3 and μ_3 , respectively, representing free space, soil medium and penetrable buried object. There are two final sets of unknown equivalent electric and magnetic surface currents in the following formulations. They are \bar{J}_1, \bar{M}_1 on the exterior of the rough ground interface between R_1 and R_2 , and \bar{J}_3, \bar{M}_3 on the exterior of the buried object interface between R_2 and R_3 . Upon applying the boundary conditions, continuity of tangential components of the electric and magnetic fields on these interfaces, new integral equation formulations are obtained as [10]-[11]:

$$\bar{E}^{inc}(\bar{r})\Big|_{\text{tang.}} = [(L_1 + L_2)\bar{J}_1 - (K_1 + K_2)\bar{M}_1 - L_3\bar{J}_3 + K_3\bar{M}_3]_{\text{tang.}} \quad (1a)$$

$$\bar{H}^{inc}(\bar{r})\Big|_{\text{tang.}} = \left[(K_1 + K_2)\bar{J}_1 + \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{M}_1 - K_3\bar{J}_3 - \frac{L_3}{\eta_2^2} \bar{M}_3 \right]_{\text{tang.}} \quad (1b)$$

$$0 = [-L_2\bar{J}_1 + K_2\bar{M}_1 + (L_3 + L_4)\bar{J}_3 - (K_3 + K_4)\bar{M}_3]_{\text{tang.}} \quad (1c)$$

$$0 = \left[-K_2\bar{J}_1 - \frac{L_2\bar{M}_1}{\eta_2^2} + (K_3 + K_4)\bar{J}_3 + \left(\frac{L_3}{\eta_2^2} + \frac{L_4}{\eta_3^2} \right) \bar{M}_3 \right]_{\text{tang.}} \quad (1d)$$

In which the integro-differential operators L_i and K_i , $i=1, 2, 3$ and 4, are given in detail in [11]. In Eqs. 1a-d, the unknown surface electric and magnetic currents are \bar{J}_1 , \bar{M}_1 , \bar{J}_3 , and \bar{M}_3 , while the tangential components of the incident electric and magnetic fields on the rough surface are given by $\bar{E}^{inc}(\bar{r})\Big|_{\text{tang.}}$ and

$\bar{H}^{inc}(\bar{r})\Big|_{\text{tang.}}$, respectively. The intrinsic impedance in

each region is $\eta_i = \sqrt{\mu_i / \epsilon_i}$, $i=1, 2$, and 3, where the dielectric permittivity and permeability in each region are ϵ_i and μ_i , respectively. The equivalent electric and magnetic currents are approximated using the Rao, Wilton and Glisson (RWG) vector basis functions [16]-[17]. Upon applying Galerkin's method for testing and substituting the RWG surface current approximations

in 1a-d, the original integral equations are transformed into a set of linear system of equations given by [10]-[11]:

$$\bar{\bar{Z}} \bar{\bar{I}} = \bar{\bar{V}} \quad (2a)$$

The impedance matrix $\bar{\bar{Z}}$ has order of $2(N+P) \times 2(N+P)$. The vector $\bar{\bar{V}}$ is a matrix of order $2(N+P) \times 1$ and composed of a submatrix of the tested tangential incident electric field \bar{E}^{inc} of order $N \times 1$ and a submatrix of the tested normalized magnetic field $\eta_1 \bar{H}^{inc}$ of order $N \times 1$, and a null submatrix of order $2P \times 1$. The quantities N and P are the numbers of basis functions (total number of edges of the triangular patches) on the surfaces of the rough ground and the buried object, respectively. If the MoM

is used to formulate and solve Eq. (2a), the computation and storage of all the elements of the matrix $\bar{\bar{Z}}$ are required. Furthermore, if an iterative solution process is used, matrix-vector products involving multiplying $\bar{\bar{Z}}$ by a vector \bar{I} are required. However, upon using the SDFMM [4]-[6],[10],[11] the matrix $\bar{\bar{Z}}$ becomes sparse and the system of equations in (2a) can be written as:

$$\bar{\bar{Z}}' \bar{I} + \bar{\bar{Z}}'' \bar{I} = \bar{V} \quad (2b)$$

where the matrix $\bar{\bar{Z}}'$ is a sparse matrix whose non-zero elements need to be calculated and stored using the conventional MoM and then multiplying them by the vector \bar{I} (near field interactions) while the matrix-vector multiply $\bar{\bar{Z}}'' \bar{I}$ is computed in one step without calculating or storing any elements of the matrix $\bar{\bar{Z}}''$. This was achieved by hybridizing the FMM with the SDP leading to the SDFMM [4]-[6]. There are three bottlenecks in the SDFMM computer code: (i) the subroutines that calculate the elements of the sparse matrix $\bar{\bar{Z}}'$; (ii) the subroutine that executes the matrix vector multiplication $\bar{\bar{Z}}' \bar{I}$ in every iteration; (iii) the subroutine that executes the fast multipole method for $\bar{\bar{Z}}'' \bar{I}$ (far-field fast multipole interactions). These three bottlenecks in the serial SDFMM computer code are separately parallelized in the current work as pictorially described in Fig. 1 [18].

The key data structure of bottleneck (i) is the sparse matrix $\bar{\bar{Z}}'$ which is originally stored in the serial SDFMM computer code as blocks of nonzero elements. These elements represent the near-field interactions in the conventional MoM. The computations of these blocks are independent and therefore are parallelized in a straightforward manner by distributing them among all processors with no additional communication. When this routine is parallelized we achieved almost a linear speedup for bottleneck (i) on 32 processors. It is necessary to mention that the elements of $\bar{\bar{Z}}'$ remain distributed among the processors at the end of parallelizing bottleneck (i).

In the second bottleneck (ii) in the serial SDFMM computer code, the matrix-vector multiplication $\bar{\bar{Z}}' \bar{I}$ is executed every iteration of the transpose-free quasi-minimal residual (TFQMR) iterative solver [19]. This multiplication is parallelized by distributing the vector \bar{I} to all processors similar to the elements of the sparse matrix $\bar{\bar{Z}}'$ in bottleneck (i). Therefore, the multiplication proceeds in parallel without additional communications and the vector components that result

from the multiplication are then distributed to all processors.

In the third bottleneck (iii) in the serial SDFMM computer code, the fast multipole part for the matrix vector multiplication represented by $\bar{\bar{Z}}'' \bar{I}$ in (2b) is computed for every iteration of the TFQMR iterative solver. This part of the serial code includes the computations of the Green's function approximations for the air and for the medium (e.g. soil). These two approximations of the Green's function are independent and are represented in the serial computer code by two separate subroutines; therefore they are executed concurrently as a first parallelization phase of bottleneck (iii). The load balance between these two subroutines is achieved using a detailed performance model based on the serial execution time of each routine, the time required for collective communication operations, and the amount of communication overhead needed. Moreover, in the serial computer code each one of these two subroutines includes all the multi-level FMM computations such as the inhomogeneous plane wave expansions and the dipole interactions at the finest level (aggregation), all interactions going up the tree, all the multi-level translation operations, all interactions going down the tree and finally the disaggregation process at the finest level which concludes the far-field interactions producing $\bar{\bar{Z}}'' \bar{I}$ [1]-[6]. As a second phase of parallelizing bottleneck (iii), each Green's function approximation subroutine is parallelized but only at the finest level. In this work, no parallelization was conducted for the multi-level portion in these subroutines due to the existing complex interdependencies in the serial computer code. More parallelization future work is needed for this part.

NUMERICAL RESULTS

We evaluated the parallel implementation of the SDFMM computer code on a 32-node Intel Pentium-based Beowulf cluster. Thirty one nodes of the Beowulf cluster are 350MHz Intel Pentium IIs with 256 MB of RAM in addition to one node of a 4x450MHz Intel Pentium II Xeon shared memory processor with 2GB of RAM. The nodes are connected to a 100 BaseTX Ethernet network and they use the SuSE 6.1 operating system with Linux kernel 2.2.13, and the MPICH 1.2.1 implementation of the MPI library. Moreover, we tested the parallelized code on a 4-node shared memory Compaq Alpha-based workstation (667Mhz Alpha 21264) of 16GB total RAM. The processor uses the UNIX OSF/1 V5.1 operating system with the MPICH 1.1.2 MPI library. Our benchmark includes three small-scale cases executed on the 256MB Intel cluster, and in addition

one moderate-scale case that is executed on the Alpha workstation. To evaluate the speedup achieved by the parallel code, we considered a range of values for the ground roughness and/or for the buried object. All results obtained by executing the parallel version of the code are validated with those computed by the serial version of the code [10]-[11]. In all computations a 10^{-3} tolerance is assumed for the TFQMR iterative solver [19]. The scattering problem configurations used in [11] are employed here, but for only one rough surface realization as shown in Fig. 2. The rough ground is characterized by Gaussian statistics with zero mean for the height, thus the roughness parameters can be described by the rms height σ and the correlation length l_c . In all cases, the relative dielectric constant of the ground soil and the penetrable buried object (anti-personnel mine) are $\epsilon_r = 2.5 - j0.18$ and $\epsilon_r = 2.9 - j0.0092$, respectively. A Gaussian beam with horizontal polarization is employed for the incident waves at normal incidence for Cases 1-3 and at 10° from normal direction for Case 4 [11].

In the small-scale Cases 1-3, the dimensions of the modeled ground are assumed to be $3\lambda_0 \times 3\lambda_0$ leading to almost 8,800 of total number of surface current unknowns, while these dimensions are increased to be $8\lambda_0 \times 8\lambda_0$ for the moderate-scale Case 4 leading to 60,320 unknowns, where λ_0 is the free space wavelength. In Case 1, the scattered electric fields from a rough ground alone (no buried target) with $\sigma = 0.3\lambda_0$ and $l_c = 0.5\lambda_0$ are calculated at height of $1.2\lambda_0$ above the ground. In Case 2, the scattered electric fields from a rough ground with a buried penetrable sphere are calculated at height of $0.5\lambda_0$ above the ground. The ground roughness is assumed to be $\sigma = 0.1\lambda_0$ and $l_c = 0.5\lambda_0$ and the sphere has radius of $a = 0.16\lambda_0$ with burial depth equal to $z = -0.32\lambda_0$ measured from its center to the mean plane of the ground. The sphere in Case 2 is replaced by a spheroid of dimensions $a = 0.3\lambda_0$ and $b = 0.15\lambda_0$ in Case 3 that is buried at $z = -0.3\lambda_0$ with ground roughness equal to $\sigma = 0.04\lambda_0$ and $l_c = 0.5\lambda_0$.

Both the overall speedup and the initial speedup (filling matrix $\bar{\bar{Z}}'$) are plotted versus the number of processors for Cases 1, 2 and 3 in Figs. 3a, 3b and 3c, respectively. The speedup is defined as the ratio of the serial runtime to the parallel runtime. The results in these figures show the significant speedup in the initial time (set up) that is consumed to fill the sparse matrix $\bar{\bar{Z}}'$ as explained in Section II. This initial speedup

dramatically affects the overall speedup of the code as shown in these figures. In addition, the results show that almost the same overall speedup can be achieved by employing only twelve instead of thirty-two processors.

The efficiency for a given number of processors is defined as the ratio of the speedup to the number of processors. In each case, the peak speedup is achieved when running on 32 processors, where for case 1, the peak speedup is 7.1 as shown in Fig. 3a, with a reduction in runtime from 99 minutes on one processor to 14 minutes on 32 processors. For Case 2, the peak speedup is 6.2 as shown in Fig. 3b, with a reduction in runtime from 90 minutes to 14 minutes while for Case 3, the peak speedup is 7.2 as shown in Fig. 3c, with a reduction in runtime from 88 minutes to 12 minutes. Over these three cases, the average speedup on 32 processors is 6.8, giving an efficiency of 0.21. Based on the serial runtimes, 88% of the code is executed in parallel. Therefore by Amdahl's Law [20], the peak speedup achievable is 8.3. We conclude that communication overhead and load imbalance among the processors account for the reduction in speedup from 8.3 to 6.8. An interesting comparison between the speedup achieved in each one of the bottlenecks (i)-(iii) mentioned in Section II, is shown in Fig. 4. These results show that the matrix-vector multiplication $\bar{\bar{Z}}'\bar{T}$ (that is the bottleneck (ii)) governs the overall speedup of the parallelized computer code.

In the second set of experiments, we solved the moderate-scale problem of Case 4 (60,320 unknowns) on the Alpha SMP using all four available processors. The penetrable spheroid of dimensions $a = 0.3\lambda_0$ and $b = 0.15\lambda_0$ is buried at $z = -0.3\lambda_0$ under the $8\lambda_0 \times 8\lambda_0$ rough ground with $\sigma = 0.04\lambda_0$ and $l_c = 0.5\lambda_0$. The magnitude of the total scattered electric field from the ground with the buried target is shown in Fig. 5. The magnitude of the scattered electric fields for just the buried spheroid is computed by subtracting the return from the rough ground using complex vector representation from the total return from the ground with the buried target [10]-[11]. The output is shown in Fig. 5b. The results of Fig. 5a and 5b clearly demonstrate that the signature of the buried plastic landmine is relatively small compared with the return from the ground which is considered a major source of clutter in landmine detection application. Moreover, the distortion observed in Fig. 5b is due to the roughness of the ground which is modeled here as only one random rough surface realization, however the Monte Carlo simulations case was presented in [11]. The serial version took 96 minutes to run this case

while the parallel version took 37 minutes, giving a speedup of 2.5 and an efficiency of 0.63. The predicted peak speedup on the four processors is 2.9. This implies that executing the parallel code on the 4-Alpha 667MHz processor gives a remarkable reduced absolute runtime for this moderate-scale case. This achievement can be exploited to execute large-scale

scattering problems as mentioned in Section II. For the memory requirements, the serial version of the code requires 950MB of RAM while the parallel version requires 1154MB of RAM distributed over the four processors as 288, 290, 289 and 287MB, respectively. Table I summarizes the parameters and output results for all cases presented in this section.

Table I Overall speedup

Case #	Number of Unknowns	σ	Object	System	Number of Processors	Serial/Par. time (min.)	Speedup (overall)
1	8,800	$0.3\lambda_0$	None	Cluster	32	99/14	7.1
2	8,800	$0.1\lambda_0$	Sphere	Cluster	32	90/14	6.2
3	8,800	$0.04\lambda_0$	Spheroid	Cluster	32	88/12	7.2
4	60,320	$0.04\lambda_0$	Spheroid	Alpha Server	4	96/37	2.5

The results described in this section demonstrate that by implementing the fine grained parallelism, we have achieved good speedups when using a single rough surface realization (one run of the code). This achievement is suitable for some subsurface scattering configurations where we may need to obtain multiple views of a target buried under the same rough surface realization [10]. This requires running the code several times. However, the current speedup is not suitable when the number of rough surface realizations is much larger than the number of available processors, e.g. Monte Carlo simulations, due to the saturation occurs in the speedup of the second and third bottlenecks.

CONCLUSIONS

Good overall speedup has been achieved as the SDFMM computer code is parallelized using the MPI library. The linear speedup obtained for the first bottleneck associated with filling the sparse impedance matrix is significant. Sensible speedups are obtained for the second and third bottlenecks associated with the matrix vector multiplication in the near-field and the far-field FMM approximations, respectively. However, the later speedups saturate upon using only 12 processors out of the 32 nodes available on the system. This saturation affects the overall speedup of the computer code and limits its application. More parallelization work is needed to enhance the speedup to be used for large Monte Carlo simulations.

ACKNOWLEDGMENTS

This research was sponsored in part by the Army Research Office Demining MURI grant # DAA 0-55-97-0013, in part by the Engineering Research Centers

Program of the NSF under award number EEC-9986821, and in part by the College of Engineering at the University of Arkansas. This work benefited from the allocation of time at the Joulian Cluster and at the Advanced Scientific Computation Center (NU-ASCC) at Northeastern University. The SDFMM was originally developed by V. Jandhyala, E. Michielssen and W. Chew at the UIUC.

REFERENCES

- [1] V. Rokhlin, "Rapid solution of integral equations of scattering theory in two dimensions," *J. Comput. Phys.*, vol. 36, pp. 414-439, 1990.
- [2] C. C. Lu and W. C. Chew, "A multilevel fast-algorithm for solving a boundary integral equation of wave scattering," *Microwave Opt. Tech. Let.*, vol. 7, pp. 466-470, July 1994.
- [3] N. Geng, A. Sullivan and L. Carin, "Multilevel fast-multipole algorithm for scattering from conducting targets above or embedded in a lossy half space," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 4, pp. 1561-1573, July 2000.
- [4] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [5] V. Jandhyala, B. Shanker, E. Michielssen and W. C. Chew, "A fast algorithm for the analysis of scattering by dielectric rough surfaces," *J. Opt. Soc. Am. A*, vol. 15, no. 7, pp. 1877-1885, July 1998.
- [6] M. El-Shenawee, V. Jandhyala, E. Michielssen and W. C. Chew, "An enhanced steepest descent fast multipole method for the analysis of scattering from two dimensional multilayered

- rough surfaces,” *Proc. of the IEEE APS/URSI '98 Conf.*, Atlanta, Georgia, p. 182, June 1998.
- [7] G. Zhang, L. Tsang and K. Pak, “Angular correlation function and scattering coefficient of electromagnetic waves scattered by a buried object under a two-dimensional rough surface,” *J. Opt. Soc. Am. A*, vol. 15, no. 12, pp. 2995-3002, December 1998.
 - [8] S. Li, C. H. Chan, L. Tsang, Q. Li, and L. Zhou, “Parallel implementation of the sparse matrix/canonical grid method for the analysis of two-dimensional random rough surfaces (three-dimensional scattering problem) on a Beowulf system,” *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 4, pp. 1600-1608, July 2000.
 - [9] D. Torrungrueng, H. Chou and J. T. Johnson, “A novel acceleration algorithm for the computation of scattering from two-dimensional large scale perfectly conducting random rough surfaces with the forward-backward method,” *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 7, pp. 1656-1666, July 2000.
 - [10] M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, “3-D subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces: Use of the steepest descent fast multipole method (SDFMM),” *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 6, pp. 1174-1182, June 2001.
 - [11] M. El-Shenawee, C. Rappaport and M. Silevitch, “Monte Carlo simulations of electromagnetic wave scattering from random rough surface with 3-D penetrable buried object: Mine detection application using the SDFMM,” the *J. Optical Society of America A*, vol.18, no. 12, pp.3077-3084, December 2001.
 - [12] S. V. Velamparambil, J. E. Schutt-Aine, J. G. Nickel, J. M. Song, and W. C. Chew, “Solving large scale electromagnetic problems using a linux cluster and parallel MLFMA,” *IEEE Antennas Propagat. Symp.*, 1:636-639, July 1999.
 - [13] S. Velamparambil, J. Song, and W. C. Chew, “On the parallelization of dynamic multilevel fast multipole method on distributed memory computers,” *Proc. International Workshop on Innovative Architecture for Future Generation High-Performance Processors and Systems, Maui, Hawaii*, November 1999.
 - [14] J. J. Dongarra and D. W. Walker, “MPI: A Standard Message Passing Interface”, *Supercomputer*, Vol. 12, No. 1, 1996, pp. 56-68.
 - [15] Message Passing Interface Forum. MPI: A Message-Passing Interface standard. *The International Journal of Supercomputer Applications and High Performance Computing*, no. 8, 1994.
 - [16] S. M. Rao, D. R. Wilton, and A. W. Glisson, “Electromagnetic scattering by surfaces of arbitrary shape,” *IEEE Trans. on Anten. and Prop.*, vol. AP 30, no. 3, pp.409-418, May 1982.
 - [17] L. Medgyesi-Mitschang, J. Putnam, and M. Gedera, “Generalized method of moments for three-dimensional penetrable scatterers,” *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1383-1398, April 1994.
 - [18] D. Jiang, W. Meleis, M. El-Shenawee, E. Mizan, M. Ashouei and C. Rappaport, “Parallel Implementation of the Steepest Descent Fast Multipole Method (SDFMM) On a Beowulf Cluster for Subsurface Sensing Applications,” the *IEEE Microwave and Wireless Components Letters (MWCL)*, vol. 12, no. 1, pp. 1-3, January 2002.
 - [19] R. W. Freund, “A transpose-free quasi-minimal residual algorithm for non-hermitian linear systems,” *SIAM J. Sci. Comput.*, vol. 14, no. 2, pp. 470-482, March 1993.
 - [20] G. M. Amdahl, “Validity of the single processor approach to achieving large scale computing capabilities”, *Proc. AFIPS Spring Joint computer Conference* 30, Atlantic City, N. J., pp. 483-485, April 1967.

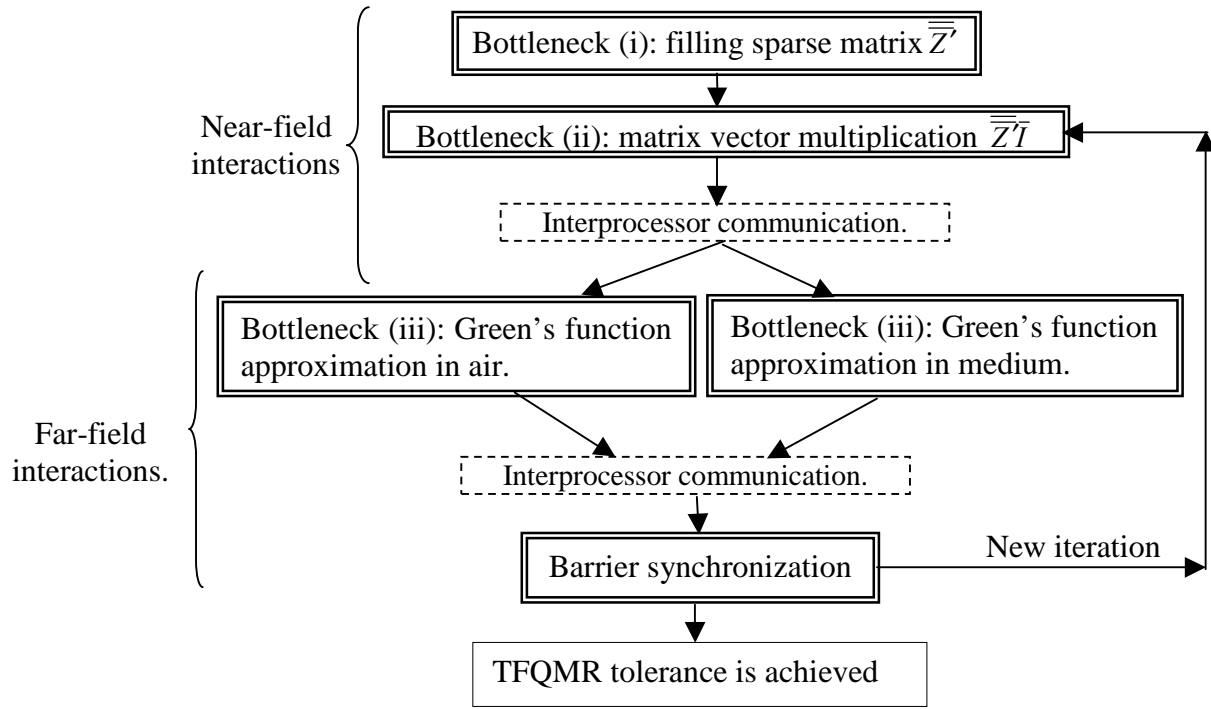


Fig. 1. Structure of parallelized SDFMM showing major computational tasks and their interrelation.

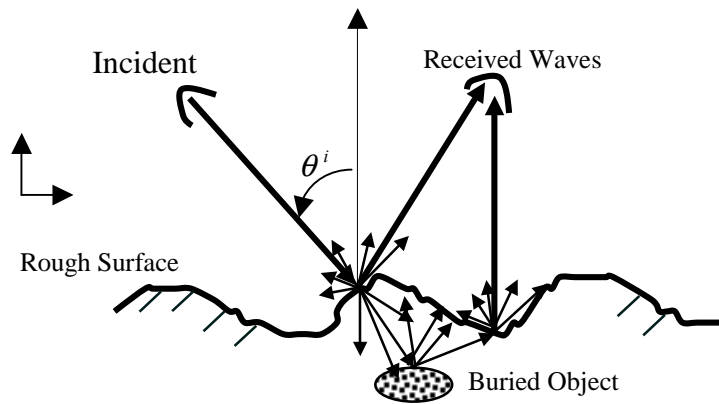
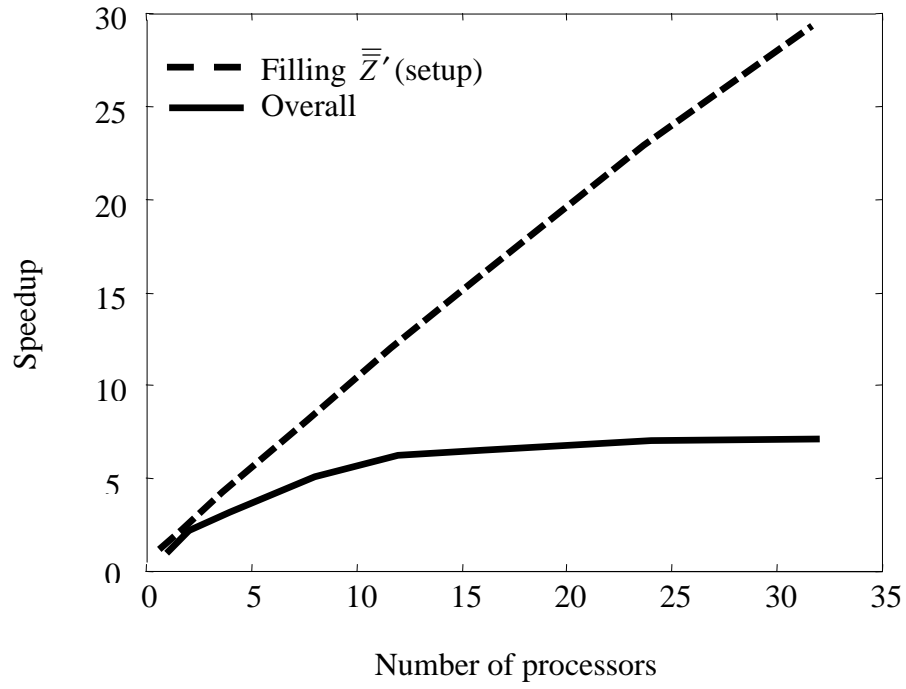
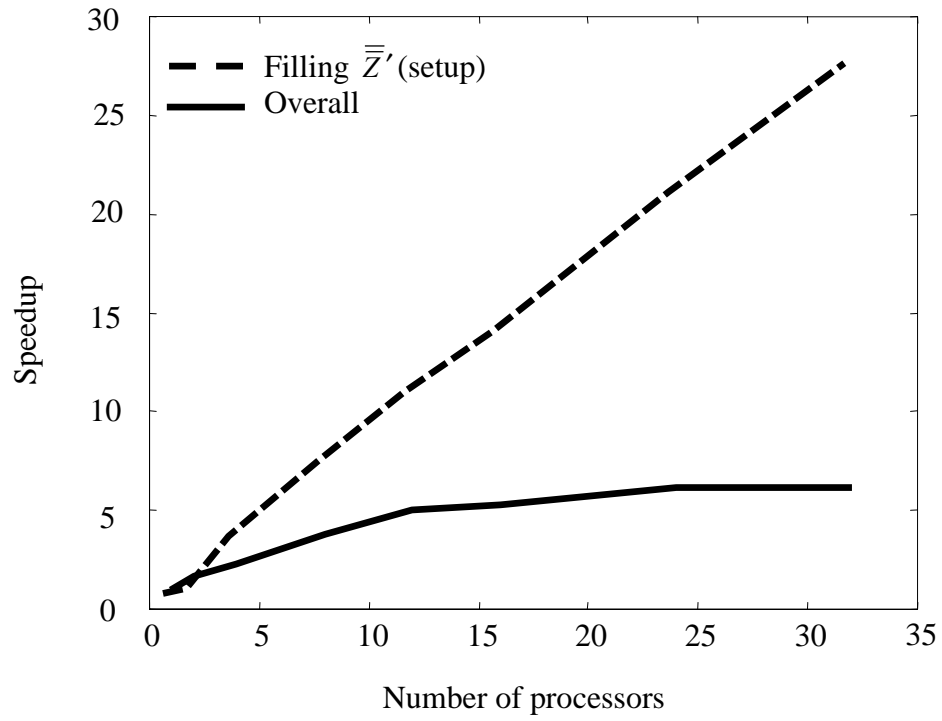


Fig. 2. Cross section of the object buried under the rough ground (3-D problem).

(a)



(b)



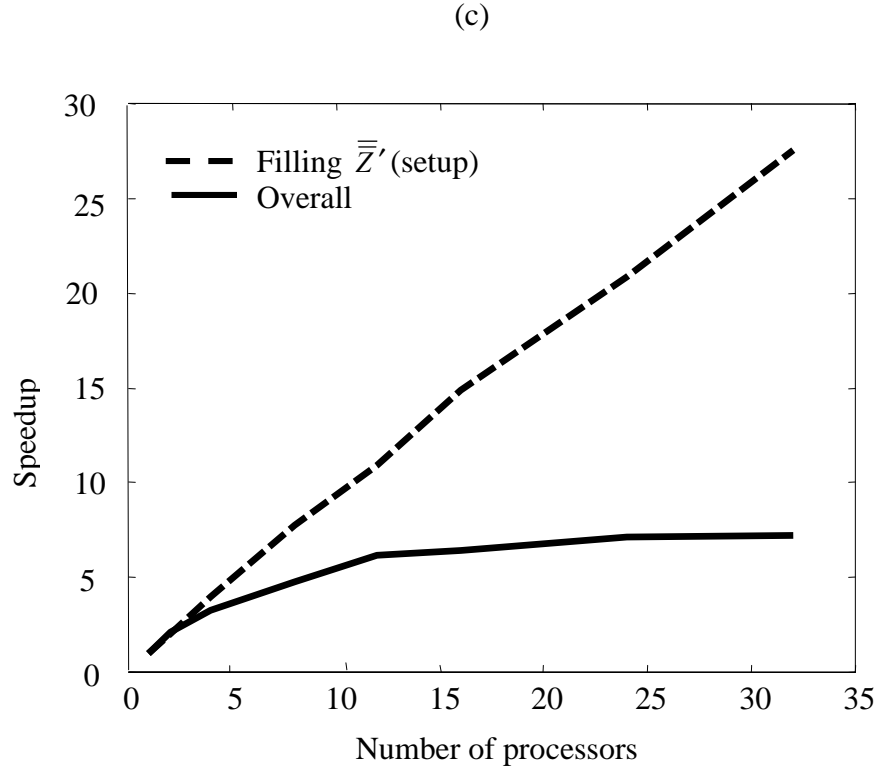


Fig. 3. Speedup of the Beowulf cluster: (a) Case 1, target-free rough ground, (b) Case 2, penetrable sphere buried under moderately rough ground surface, (c) Case 3, penetrable spheroid buried under slightly rough ground surface.

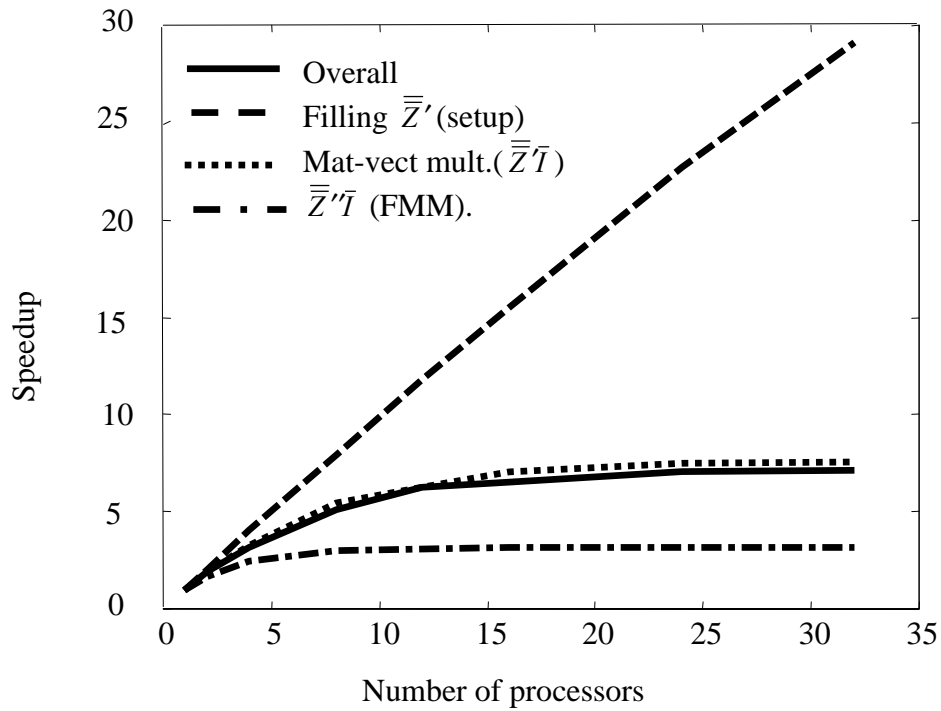


Fig. 4. Performance improvement for each of the separate component tasks and overall speedup of the SDFMM algorithm, as a function of the number processors in the Beowulf cluster.

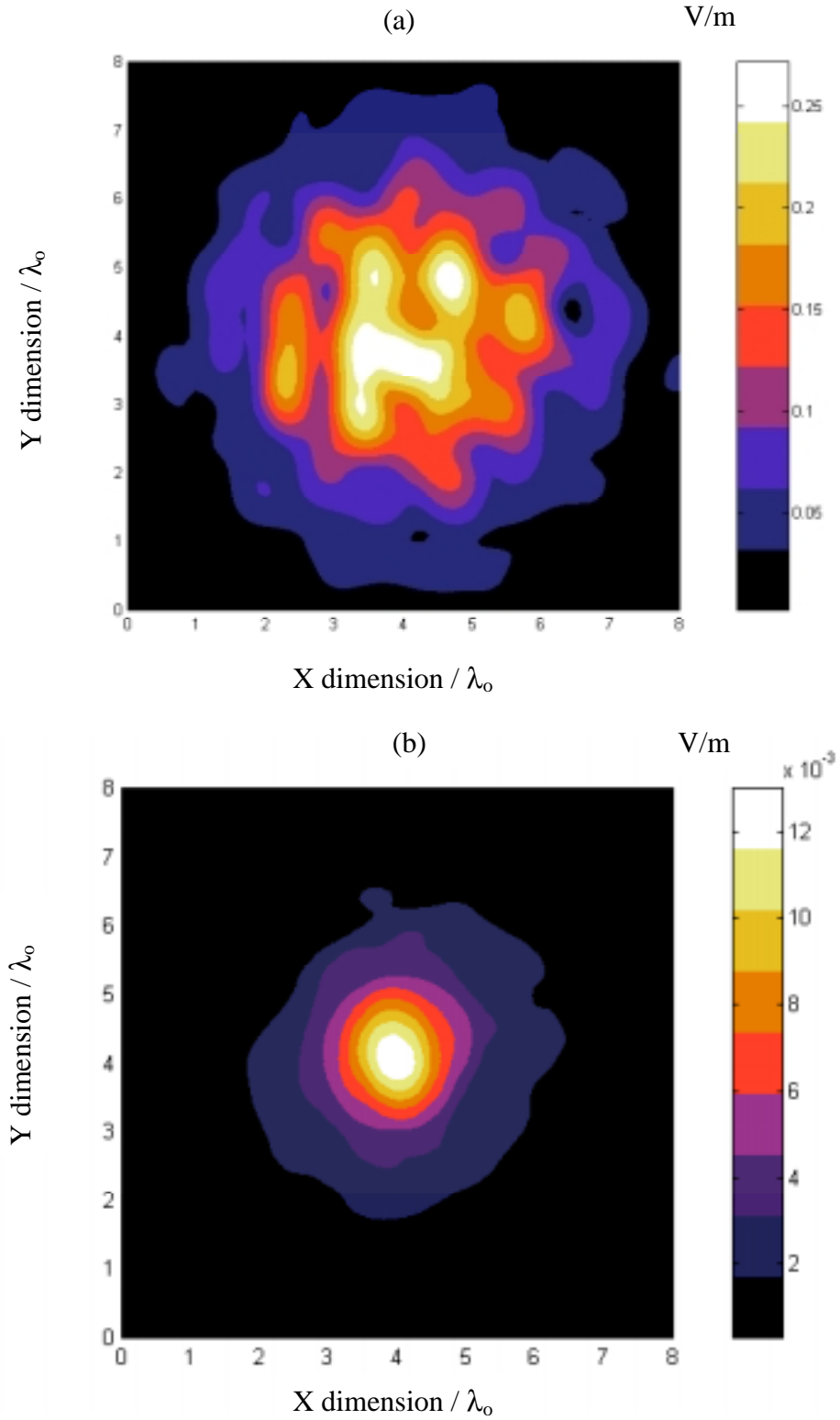
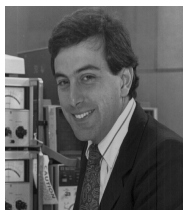


Fig. 5 The near electric field scattered above the rough ground at $z = 0.5\lambda_0$ for the spheroid of Case 4 ($a = 0.3\lambda_0$, $b = 0.15\lambda_0$, buried at $z = -0.3\lambda_0$ in conductive clay loam soil), computed using the 4-processor Alpha Server: (a) the rough ground with the buried spheroid (total field), (b) just the spheroid obtained by subtraction.



Magda El-Shenawee received the Ph.D. degree in electrical engineering from the University of Nebraska-Lincoln in 1991. In 1992, she worked as a Research Associate in the Center for Electro-Optics at the University of Nebraska. In 1997, she worked as Visiting Scholar at the University of

Illinois at Urbana-Champaign and in 1999, she joined the Center for Electromagnetics Research at Northeastern University, Boston. Currently, Dr. El-Shenawee is an assistant professor in the Department of Electrical Engineering at the University of Arkansas, Fayetteville. Her research areas are rough surface scattering, computational electromagnetics, subsurface sensing of buried objects, breast cancer modeling, numerical methods, and micro-strip circuits. Dr. El-Shenawee is a member of Eta Kappa Nu electrical engineering honor society.



Carey M. Rappaport received five degrees from the Massachusetts Institute of Technology: the SB in Mathematics, the SB, SM, and EE in Electrical Engineering in June 1982, and the PhD in Electrical Engineering in June 1987. Prof. Rappaport has

worked as a teaching and research assistant at MIT from 1981 until 1987, and during the summers at COMSAT Labs in Clarksburg, MD, and The Aerospace Corp. in El Segundo, CA. He joined the faculty at Northeastern University in Boston, MA in 1987. During Fall 1995, he was Visiting Professor of Electrical Engineering at the Electromagnetics Institute of the Technical University of Denmark, Lyngby, as part of the W. Fulbright International Scholar Program. He has consulted for Geo-Centers, Inc., PPG, Inc. on wave propagation and modeling, and microwave heating and safety. He is Principal Investigator of the ARO Multidisciplinary University Research Initiative in Humanitarian Demining and Co-Principal Investigator of the Center for Subsurface Sensing and Imaging Systems (CenSSIS) Engineering Research Center. Prof. Rappaport has authored over 180 technical journal and conference papers in the areas of microwave antenna design, electromagnetic scattering computation, and bio-electromagnetics, and has received two reflector antenna patents, two biomedical device patents and three subsurface sensing device patents. He was awarded the IEEE Antenna and Propagation Society's H.A. Wheeler Award for best applications paper of 1985. He is a member of Sigma Xi and Eta Kappa Nu professional honorary societies.



Desheng Jiang received the BS degree in Chemistry from Jiangxi Normal University in 1993, the MS in Oceanography from Texas A&M University in 1999, and the MS in Electrical Engineering from Northeastern University in 2001, respectively. He is currently a

Software Engineer with Openwave Systems Inc.



Waleed M. Meleis received the BSE degree in electrical engineering from Princeton University in 1990 and the MS and PhD degrees in computer science and engineering from the University of Michigan in 1992 and 1996, respectively. He is an assistant professor in the Department of

Electrical and Computer Engineering at Northeastern University. His research interests are in scheduling algorithms and bounds for modern processors and compilers.



David R. Kaeli received his B.S. in Electrical Engineering from Rutgers University, his M.S. in Computer Engineering from Syracuse University, and his PhD in Electrical Engineering from Rutgers University. He is currently an Associate Professor on the faculty

of the Department of Electrical and Computer Engineering at Northeastern University. Prior to 1993, he spent 12 years at IBM, the last 7 at IBM T.J. Watson Research in Yorktown Heights, N.Y.. In 1996 he received an NSF CAREER Award. He currently directs the Northeastern University Computer Architecture Research Laboratory (NUCAR). Dr. Kaeli's research interests include computer architecture and organization, compiler optimization, VLSI design, trace-driven simulation and workload characterization. He is a member of the IEEE and ACM. He is presently an Associate Editor for IEEE Transactions on Computer and IEEE Computer Architecture Letters.

Scattering from Objects Buried in Multilayered Random Rough Surfaces Using the Steepest Descent Fast Multipole Method

Magda El-Shenawee
Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701
Fax: 501-575-7967, Tel: 501-575-6582
magda@uark.edu

ABSTRACT

The scattering of electromagnetic waves from a penetrable shallow object buried beneath 2-D multilayered random rough surface is presented. A rigorous electromagnetic model has been developed for scattering from inhomogeneous rough ground surfaces that is based on the classical equivalence theorem and the method of moments (MoM). The model is significantly accelerated by implementing the Steepest Descent Fast Multipole Method (SDFMM). Four homogeneous regions are involved in this application; air, dielectric object, upper dry-soil layer and lower wet-soil layer. Numerical results for normalized radar cross sections (RCS) are presented for a variety of lossy and lossless multilayered ground, incident angles, thickness of upper-layer, and incident wave polarization. The results clearly demonstrate the considerable effect of both the thickness and the dielectric constant of the upper-layer relative to that of the lower-layer on the scattered waves.

I. INTRODUCTION

The electromagnetic sensing of buried objects in the presence of a random rough interface is a crucial step for subsurface detection problems in general. An observation agreed upon by most theoreticians and experimentalists is that surface roughness constitutes a major source of clutter (i.e., noise) in the received electromagnetic signals. As a result, several researchers have thoroughly investigated and published works on electromagnetic scattering from one- (1-D) or

two-dimensional (2-D) rough surfaces (i.e., x - or y -directions), e.g., [1]-[15]. These publications are for rough surface scattering without any buried objects. In addition, works on modeling electromagnetic waves from objects buried under a 1-D rough surface are published; e.g., [16]-[18]. Moreover, modeling and detecting objects buried beneath flat half-space surfaces were presented [19]-[22]. There are few published works for a single object buried beneath a 2-D random rough surface [23]-[26]. In addition, in an effort to simulate a landmine buried near a clutter object (e.g., tree root), results for sensing two penetrable objects buried beneath a 2-D random rough ground are presented in [27], [28].

However, in all work cited above, ground surface is assumed to be single layered and *not* multilayered; this is not the case in the real environment. There are few works published to model scattering from multilayered half space or multilayered rough surfaces *without* buried objects, [29]-[32]; while no work has yet been published on scattering from objects buried beneath multilayered randomly rough ground. Therefore, the objective of this work is to increase the efficacy of the present subsurface sensing methods by accounting for the multilayer nature of the rough ground. However, the computational complexity of the problem dramatically increases upon inserting penetrable objects under the multilayered rough interface. This complex scenario necessitates the use of fast computational algorithms such as the Fast Multipole Method (FMM) [33]-[35], the SDFMM [13],[36]-[38], the Sparse Matrix/Canonical Grid Method (SMCG) [23],[39] and the Spectral Algorithm combined with the Forward-Backward Method (FB/NSA) [40]. Recently, the SDFMM that has $O(N)$ computational complexity where N is the number of surface current unknowns, has been successfully implemented for a variety of subsurface sensing applications [24],[26]-[28],[32],[41]. The rigorous electromagnetic model presented in [27] for analyzing the scattering from two dielectric objects buried under the 2-D random rough ground is

used in this work. The geometry of the current application is shown in Fig. 1 where a dielectric object is buried between the air-ground rough interface and the underground rough interface with burial depth H that is less than one free space wavelength. It is important to mention that the SDFMM was originally developed for analyzing quasi-planar 3-D structures [36], which implies shallow burial depth H compared with the free space wavelength.

II. FORMULATION

The inhomogeneous geometry considered in this work is shown in Fig. 1c and is composed of four different regions; the air, the ground, the first inner scatterer and the second inner scatterer with relative permittivity and permeability as ϵ_1 and μ_1 , ϵ_2 and μ_2 , ϵ_3 and μ_3 , and ϵ_4 and μ_4 , respectively. The rigorous integral equation-based electromagnetic model derived in [27] is employed to calculate the unknown equivalent surface currents on all scatterers shown in Fig. 1. Upon simulating the incident waves as a carefully tapered Gaussian beam [23], [42], the closed scatterers can be approximated by open scatterers where the excited surface currents on the back of the scatterers S_1 and S_3 can be neglected as depicted in Fig. 1b. All scatterers shown in Fig. 1c; the air-ground rough interface, the buried object, and the underground rough interface are discretized into triangular patches [43],[44]. After some algebraic manipulations, the unknown equivalent electric and magnetic surface currents (in Fig. 1c) are \bar{J}_1 and \bar{M}_1 on the air-ground interface, \bar{J}_3 and \bar{M}_3 on the buried object and \bar{J}_5 and \bar{M}_5 on the underground interface as included in the following six integral equations (details are presented in [27]):

$$\bar{E}^{inc}(\bar{r})\big|_{\text{tang.}} = \left[(L_1 + L_2)\bar{J}_1 - (K_1 + K_2)\bar{M}_1 - L_3\bar{J}_3 + K_3\bar{M}_3 - L_4\bar{J}_5 + K_4\bar{M}_5 \right]_{\text{tang.}} \quad (1a)$$

$$\bar{H}^{inc}(\bar{r})\big|_{\text{tang.}} = \left[(K_1 + K_2)\bar{J}_1 + \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{M}_1 - K_3\bar{J}_3 - \frac{L_3}{\eta_2^2} \bar{M}_3 - K_4\bar{J}_5 - \frac{L_4}{\eta_2^2} \bar{M}_5 \right]_{\text{tang.}} \quad (1b)$$

$$0 = [-L_2 \bar{J}_1 + K_2 \bar{M}_1 + (L_3 + L_5) \bar{J}_3 - (K_3 + K_5) \bar{M}_3 + L_4 \bar{J}_5 - K_4 \bar{M}_5]_{\text{tang.}} \quad (1c)$$

$$0 = \left[-K_2 \bar{J}_1 - \frac{L_2}{\eta_2^2} \bar{M}_1 + (K_3 + K_5) \bar{J}_3 + \left(\frac{L_3}{\eta_2^2} + \frac{L_5}{\eta_3^2} \right) \bar{M}_3 + K_4 \bar{J}_5 + \frac{L_4}{\eta_2^2} \bar{M}_5 \right]_{\text{tang.}} \quad (1d)$$

$$0 = [-L_2 \bar{J}_1 + K_2 \bar{M}_1 + L_3 \bar{J}_3 - K_3 \bar{M}_3 + (L_4 + L_6) \bar{J}_5 - (K_4 + K_6) \bar{M}_5]_{\text{tang.}} \quad (1e)$$

$$0 = \left[-K_2 \bar{J}_1 + \frac{L_2}{\eta_2^2} \bar{M}_1 + K_3 \bar{J}_3 + \frac{L_3}{\eta_2^2} \bar{M}_3 + (K_4 + K_6) \bar{J}_5 + \left(\frac{L_4}{\eta_2^2} + \frac{L_6}{\eta_4^2} \right) \bar{M}_5 \right]_{\text{tang.}} \quad (1f)$$

where the intrinsic impedance is $\eta_i = \sqrt{\mu_i / \epsilon_i}$, with $i=1, 2, \dots, 4$ and L_j and K_j , $j=1, 2, \dots, 6$ are the integro-differential operators derived in [27]. The unknown electric and magnetic surface currents in (1) are approximated using the vector basis functions $\bar{j}(\bar{r})$ [43],[44] as:

$$\bar{J}_i(\bar{r}) = \sum_{n=1}^{N_k} I_{in} \bar{j}_{kn}(\bar{r}), \quad \bar{M}_i(\bar{r}) = \eta_1 \sum_{n=1}^{N_k} I_{(i+1)n} \bar{j}_{kn}(\bar{r}), \quad \bar{r} \in S_k, \quad \text{for } i=1, 3 \text{ and } 5 \quad (2)$$

in which $k=1, 2$ and 3 for $i=1, 3$ and 5 , respectively. Substituting (2) in (1), the linear system of equations is obtained [27]

$$\bar{\bar{Z}} \bar{I} = \bar{V} \quad (3a)$$

The total impedance matrix $\bar{\bar{Z}}$ has order of $2(N_1 + N_2 + N_3) \times 2(N_1 + N_2 + N_3)$, where the number of current unknowns on the air/ground interface, the buried object and the underground interface are $2N_1$, $2N_2$ and $2N_3$, respectively. The vector \bar{V} represents the tested tangential incident electric field \bar{E}^{inc} and the normalized magnetic field $\eta_1 \bar{H}^{inc}$ on the air/ground interface.

The matrix $\bar{\bar{Z}}$ is given by

$$\bar{\bar{Z}} = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22} & Z_{23} \\ Z_{31} & Z_{32} & Z_{33} \end{bmatrix} \quad (3b)$$

in which the submatrices Z_{11} , Z_{22} , and Z_{33} represent the self interactions on the air-ground interface, the buried object and the underground interface, respectively. On the other hand, the submatrices Z_{12} (and Z_{21}) and Z_{13} (and Z_{31}) represent the interactions between the air-ground interface with the buried object and with the underground interface, respectively. The submatrices Z_{32} (and Z_{23}) represent the interactions between the underground interface and the buried object. The SDFMM is implemented to dramatically accelerate solving for the unknown surface current coefficients \bar{I} in (3a) [27].

III. NUMERICAL RESULTS

As discussed in Section II and pictorially described in Fig. 1, the tapered Gaussian beam is used to illuminate the multilayered ground surface [23], [42]. The rough surfaces are characterized with Gaussian statistics for the height and for the autocorrelation function [45]. In this section, the transpose-free quasi-minimal iterative solver (TFQMR) is used to solve (3a) with a tolerance range of 10^{-3} - 10^{-5} [46]. In all the examples presented here, only one rough surface realization is considered.

For the multilayered rough ground with buried object shown in Fig. 1c, it is necessary to validate the SDFMM implemented for this application versus the MoM. The percentage of the relative norm of the error in surface currents is plotted in Fig. 2 where the vector C represents the electric and magnetic currents on surfaces S_1 , S_2 and S_3 shown in Fig. 1b. In Fig. 2a, the norm of the error and the computer memory requirements are plotted versus the SDFMM's finest block size depicted in Fig. 1c. The scale on the left is for the relative error, and on the right is for the computer memory. In this work; flat/flat implies that both the air-ground and the underground interfaces are flat while flat/rough implies the air-ground interface is flat and the underground one is rough, etc. The multilayered ground in this small-scale example consists of a flat/flat with a buried oblate

spheroid. The two flat surfaces have dimensions of $2.96\lambda_0 \times 2.96\lambda_0$, the underground interface is buried at $z = -0.5\lambda_0$ (i.e., $H = 0.5\lambda_0$ in Fig. 1c), and the oblate spheroid's dimensions are $a = 0.3\lambda_0$ and $b = 0.15\lambda_0$ with burial depth measured from its center equal to $z = -0.25\lambda_0$. The total number of the electric and magnetic surface currents unknowns is 16,732 ($2N_1 = 2N_3 = 8066$ and $2N_2 = 600$). The incident angles of the Gaussian beam are $\theta^i = 0$ and $\phi^i = 0$. The relative dielectric constants of the upper layer of the ground, the buried object and the lower layer are $\epsilon_{r2} = 2.5 - j0.18$, $\epsilon_{r3} = 2.9 - j0.0092$ and $\epsilon_{r4} = 6.5 - j0.1$, respectively. The results show the monotonic decrease in the error with increasing the FMM's finest block size. On the other hand, a rapid increase in the computer memory requirements, due to the multilayered nature of the scatterer, is clearly observed in Fig. 2a. In Fig. 2b, the relative norm of the error in $\|C\|$ is plotted versus the height H of the upper layer (see Fig. 1c). The results show the trend of increasing the error with increasing H since the SDFMM is more efficient for quasi-planar structures. For non-quasi-planar structures, the SDFMM multiple interaction model can be used instead of the complete SDFMM model implemented in the current work [47]. In this section, the SDFMM's finest block has dimensions of $0.32\lambda_0 \times 0.32\lambda_0$.

In the second example, the dimensions of the multilayered ground are increased to be $6.48\lambda_0 \times 6.48\lambda_0$ with the same buried spheroid as before. The total number of the electric and magnetic surface current unknowns is 78,684 ($2N_1 = 2N_3 = 39042$ and $2N_2 = 600$). The object is buried at $z = -0.3\lambda_0$ while the underground layer is buried at $z = -0.6\lambda_0$. Both the air-ground and the underground interfaces are random rough surfaces with rms height and correlation length equal to $\sigma_1 = 0.04\lambda_0, l_{c1} = 0.5\lambda_0$ and $\sigma_2 = 0.03\lambda_0, l_{c2} = 0.4\lambda_0$, respectively. The normalized

RCS is defined in this work by $4\pi r^2 |\bar{E}^s|^2 / (2\eta_0 P^i A_z)$, where A_z is the footprint area and P^i is the incident power [31]. All RCS results in this work are presented as RCS/λ_0^2 where λ_0 is the free space wavelength in meters.

In Fig. 3a the three regions are assumed to be lossless with no buried objects with $\epsilon_{r2} = 6.5$ and $\epsilon_{r4} = 2.5$ for Case 1 while $\epsilon_{r2} = 2.5$ and $\epsilon_{r4} = 6.5$ for Case 2. On the other hand, in Fig. 3b there is a spheroid buried at $z = -0.3\lambda_0$ and the three regions are assumed to be lossy with $\epsilon_{r2} = 6.5 - j0.1$ and $\epsilon_{r4} = 2.5 - j0.18$ for Case 1 and $\epsilon_{r2} = 2.5 - j0.18$, and $\epsilon_{r4} = 6.5 - j0.1$ for Case 2. It is assumed that $\epsilon_{r3} = 2.9$ and $\epsilon_{r3} = 2.9 - j0.0092$ for lossless and lossy cases, respectively. In Fig. 3c, the data of Fig. 3b is used but with no spheroid buried in the ground. The results in Figs. 3a-c are for normal incidence while in Fig. 3d they are for $\vartheta^i = 10^\circ$. The results of Fig. 3a show larger magnitudes for the lossless case than those for the lossy cases of Figs. 3b-d. Moreover, in Figs. 3a-c, the results show larger magnitudes for the flat/flat Case 2 than for the flat/flat Case 1. This implies that when $\epsilon_{r4} > \epsilon_{r2}$, the lower interface dominates the scattering mechanism at $H = -0.6\lambda_0$. Similar observation is shown for the lossless case when both interfaces are rough surfaces, however, for the lossy cases in Figs 3b and 3c, slightly larger magnitudes are observed for Case 1 than for Case 2. Moreover, a slight difference is observed between the results of Fig. 3b (no buried spheroid) and 3c (with buried spheroid) due to the small size of the object which agrees with previous work where a single or multiple objects were buried in a single layered ground [24],[26],[27].

The results in Fig. 4 are for a lossless rough/rough multilayered ground with the same buried spheroid, polarization and incident angle as in Case 1 in Fig. 3. In Fig. 4a, the norms $\|C\|$ of the currents on the spheroid surface, the air-ground and the underground interfaces relative to the

norm of all currents $\|C_{\text{total}}\|$ are plotted versus the height H . These quantitative results show that the currents on the underground interface are the largest in this example. Moreover, the currents on the spheroid are changing slightly with increasing height H while the currents on both the air-ground and the underground interfaces show a trend to change only at $H = 0.65\lambda_0$. Even though the results of Fig. 3a showed larger magnitudes for the rough/rough Case 2 than the rough/rough Case 1 when $H = 0.6\lambda_0$, however, the results of Fig. 4b suggest that this observation could be changing with increasing H . The results in this example are for the HH-polarization in the backscatter direction.

In Fig. 5, the ground is assumed to be a lossless rough/rough multilayer with $\epsilon_{r2} = 6.5$, $\epsilon_{r3} = 2.9$ and ϵ_{r4} is real. The relative norm of the electric and magnetic currents $\|C\|/\|C_{\text{total}}\|$ are plotted versus the dielectric constant of the underground layer ϵ_{r4} . The results show that the currents on the spheroid are smaller than those on the upper- and underground interfaces. The currents on the underground interface are decreasing while the currents on the air-ground interface are increasing with increasing ϵ_{r4} as shown in Fig. 5a. Interestingly, when $\epsilon_{r4} = \epsilon_{r2} = 6.5$, i.e., homogeneous ground, the underground surface currents become almost the same as those on the underground interface, which could be due to the lossless nature of the ground and due to the relatively small height ($H = 0.6\lambda_0$). On the other hand, for the lossy case (as will be shown in Fig. 7a), the currents on the underground interface are clearly smaller than those on the air-ground one. In Fig. 5b, the normalized HH-RCS is increasing with ϵ_{r4} where the results are for the backscatter direction at $\vartheta^i = 0$. These results show that for $\epsilon_{r4} < \epsilon_{r2}$ the wave transmission into the lower layer is larger than the waves transmission when $\epsilon_{r4} > \epsilon_{r2}$. This mechanism could explain the observation that the equivalent currents on the underground

interface are larger than those on the air-ground one when $\epsilon_{r4} < \epsilon_{r2}$ and vice-versa when $\epsilon_{r4} > \epsilon_{r2}$ as shown in Fig. 5a.

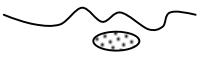

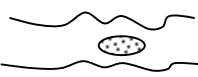
Similar investigations for the lossless case are conducted but versus the dielectric constant of the upper layer as shown in Fig. 6a, where $\|C\|/\|C_{\text{total}}\|$ is plotted versus ϵ_{r2} with $\epsilon_{r3} = 2.9$ and $\epsilon_{r4} = 6.5$. The results quantitatively show, that as ϵ_{r2} increases the currents on the spheroid decreases. In this case, fewer waves are transmitted into the ground causing less excitation to the object. Consistent with the observation of Fig. 5a, the results in Fig. 6a show that when $\epsilon_{r2} < \epsilon_{r4}$, the equivalent currents on the underground interface are smaller than those on the air-ground one and vice versa when $\epsilon_{r2} > \epsilon_{r4}$. Interestingly, the normalized HH-RCS shows initial decrease followed by monotonic increase with increasing ϵ_{r2} as shown in Fig. 6b (the results are in the backscatter at $\vartheta^i = 0$). These results show that the RCS has minimum value when $\epsilon_{r2} \approx 4.2$ which could be due to a destructive multiple interaction mechanism between the air-ground and the underground interfaces.

In Fig. 7, the rough/rough ground is assumed to be lossy with $\epsilon_{r2} = 2.5 - j0.18$, $\epsilon_{r3} = 2.9 - j0.0092$, $\epsilon_{r4} = 6.5 - j0.1$, and the incident waves are assumed to be vertically-polarized. The relative norms $\|C\|/\|C_{\text{total}}\|$ are plotted versus incident angle θ^i where all results are in the backscatter direction ($\theta^i = \theta^s, \phi^i - \phi^s = \pi$). The results in Fig. 7a show that the currents on the spheroid are smaller than those on the air-ground and underground interfaces. Moreover, the currents on the underground interface reflect the lossy nature of the ground in this example compared with the lossless case results in Fig. 6a. In Fig. 7b and 7c, the normalized RCS for the VV- and the HV- polarization are plotted versus the incident angle θ^i , respectively. These results are compared with those of the single layer rough ground and the same buried

spheroid, i.e., $\epsilon_{r2} = \epsilon_{r4} = 2.5 - j0.18$ and $\epsilon_{r3} = 2.9 - j0.0092$. The results in Fig. 7b show the impact of the underground interface on the VV-RCS especially at small elevation angles. However, due to the small ground roughness considered in this work of both the air-ground and the underground interfaces, the HV-RCS is much smaller than the VV- case as shown in Fig. 7b. But the waves scattered from the multilayered rough ground show more depolarization than those scattered from a single layer rough ground as demonstrated in Fig.7c. Similar results are obtained for the HH- and the VH- cases.

The results in Table 1 summarize the computational complexity of three subsurface sensing applications using the SDFMM [24], [27]. These results show the rising computational complexity for the multilayered ground.

Table 1 CPU time and computer memory requirements comparison

Description	Geometry	# Unknowns	CPU time (Hrs)	Memory (MB)
Single object buried beneath single layer rough ground [24].		60,200	2.0	840
Two objects buried beneath single layered rough ground [27].		60,800	3.6	856
Single object buried in multi layered rough ground (current work).		78,684	6.6	1760

IV. CONCLUSIONS

A 3-D analysis of the multilayered rough ground with a buried dielectric object is investigated using a rigorous electromagnetic model significantly accelerated by using the SDFMM. The results show the considerable impact of the multilayered nature of the ground on the scattering process compared with the single layer ground. In both cases, the buried object is clearly the weakest scatterer due to its small size compared with the wavelength. The thickness

and the dielectric constant of the upper-layer of the ground relative to the ground lower-layer significantly contribute to the scattering mechanism in the current application.

ACKNOWLEDGMENTS

This research was sponsored in part by the Arkansas Science and Technology Authority Grant No AR/ASTA/01-B-18 and in part by the Northeastern University's Demining MURI grant # DAA 0-55-97-0013. The SDFMM was originally developed by V. Jandhyala, E. Michielssen and W. Chew at the UIUC.

References

- [1] A. Ishimaru and J. S. Chen, "Scattering from very rough metallic and dielectric surfaces: a theory based on the modified Kirchhoff approximation," *Waves in Random Media*, pp. 21-34, January 1991.
- [2] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.
- [3] M. Nieto-Vesperinas and J. A. Sánchez-Gil, "Intensity angular correlations of light multiply scattered from random rough surfaces," *J. Opt. Soc. Am. A*, vol. 10, no. 1, pp. 150-157, January 1993.
- [4] M. El-Shenawee and E. Bahar, "Numerical method to compute TE and TM multiple scatter from rough surfaces exhibiting backscatter enhancement," *IEEE Trans. on Magnetics*, vol. 30, no. 5, pp. 3140-3143, September 1994.
- [5] K. Fung, M. R. Shah and S. Tjuatja, "Numerical simulations of scattering from three-dimensional randomly rough surfaces," *IEEE Trans. Geosci. & Remote Sensing*, vol. 32, no. 5, pp. 986-994, September 1994.
- [6] L. Tsang, C. H. Chan, K. Pak, H. Sangani, A. Ishimaru and P. Phu, "Monte Carlo simulations of large-scale composite random rough-surface scattering based on the banded-matrix iterative approach," *J. Opt. Soc. Am. A*, vol. 11, no. 2, pp. 691-696, February 1994.
- [7] E. Bahar and M. El-Shenawee, "Enhanced backscatter from one dimensional random rough surfaces-stationary phase approximations to full wave solutions," *J. Opt. Soc. Am. A*, vol. 12, no. 1, pp. 151-161, January 1995.
- [8] F. D. Hastings, J. B. Schneider and S. L. Broschat, "A Monte Carlo FDTD technique for rough surface scattering," *IEEE Trans. Antennas & Prop.*, vol. 43, pp. 1183-1191, November 1995.
- [9] D. A. Kapp and G. S. Brown, "A new numerical method for rough surface scattering calculations," *IEEE Trans. Antennas & Prop.*, vol. 44, no. 5, pp. 711-721, May 1996.
- [10] J. T. Johnson, L. Tsang, R. T. Shin, K. Pak, C. H. Chan, A. Ishimaru and Y. Kuga, "Backscattering enhancement of electromagnetic waves from two-dimensional perfectly

- conducting random rough surfaces: A comparison of Monte Carlo simulations with experimental data,” *IEEE Trans. Antennas & Prop.*, vol. 44, no. 5, pp. 748-756, May 1996.
- [11] R. L. Wagner, J. Song and W. C. Chew, “Monte Carlo simulation of electromagnetic scattering from two-dimensional random rough surfaces,” *IEEE Trans. Antennas & Prop.*, vol. 45, no. 2, pp. 235-245, February 1997.
 - [12] C. H. Chan, L. Tsang and Q. Li, “Monte Carlo simulations of large-scale one dimensional random rough-surface scattering at near grazing incidence: Penetrable case,” *IEEE Trans. Antennas & Prop.*, vol. 46, no. 1, pp. 142-149, January 1998.
 - [13] V. Jandhyala, B. Shanker, E. Michielssen and W. C. Chew, “A fast algorithm for the analysis of scattering by dielectric rough surfaces,” *J. Opt. Soc. Am. A*, vol. 15, no. 7, pp. 1877-1885, July 1998.
 - [14] J. T. Johnson, “Third order small perturbation method for scattering from dielectric rough surfaces,” *J. Opt. Soc. Am. A*, vol. 16, no. 11, pp. 2720-2736, November 1999.
 - [15] M. El-Shenawee and C. Rappaport, “Modeling clutter from Bosnian and Puerto Rican rough ground surfaces for GPR subsurface sensing applications using the SDFMM Technique,” *J. Subsurface Sensing Technologies and Applications (JSSTA)*, in print.
 - [16] K. O’Neill, R. F. Lussky Jr. and K. D. Paulsen, “Scattering from a metallic object embedded near the randomly rough surface of a lossy dielectric,” *IEEE Trans. Geosci. & Remote Sensing*, vol. 34, no. 2, pp. 367-376, March 1996.
 - [17] G. Zhang, L. Tsang and Y. Kuga, “Studies of the angular correlation function of scattering by random rough surfaces with and without a buried object,” *IEEE Trans. Geosci. Remote & Sensing*, vol. 35, no. 2, pp. 444-453, March 1997.
 - [18] A. Madrazo and M. Nieto-Vesperinas, “Scattering of light and other electromagnetic waves from a body buried beneath a highly rough random surface,” *J. Opt. Soc. Am. A*, vol. 14, no. 8, pp. 1859-1866, August 1997.
 - [19] N. Geng, A. Sullivan and L. Carin, “Multilevel fast-multipole algorithm for scattering from conducting targets above or embedded in a lossy half space,” *IEEE Trans. Geosci. & Remote Sensing*, vol. 38, no. 4, pp. 1561-1573, July 2000.
 - [20] A. Sullivan, R. Damarla, N. Geng, Y. Dong and L. Carin, “Ultrawide-band synthetic aperture radar for detection of unexploded ordnance: Modeling and measurements,” *IEEE Trans. Antennas & Prop.*, vol 48, no. 9, pp. 1306-1315, September 2000.
 - [21] J. T. Johnson, “A study of the four-path model for scattering from an object above a halfspace,” *Microwave Opt. Tech. Letters*, July 2001.
 - [22] Ling Li, Jiangqi, Zhijun Liu and Lawrence Carin, “MLFMA analysis of scattering from multiple targets in the presence of a half space,” submitted in 2001 to the *IEEE Trans. Antennas & Prop.*
 - [23] G. Zhang, L. Tsang and K. Pak, “Angular correlation function and scattering coefficient of electromagnetic waves scattered by a buried object under a two-dimensional rough surface,” *J. Opt. Soc. Am. A*, vol. 15, no. 12, pp. 2995-3002, December 1998.
 - [24] M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, “3-D subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces: Use of the steepest descent fast multipole method (SDFMM),” *IEEE Trans. Geosci. & Remote Sensing*, vol. 39, no. 6, pp. 1174-1182, June 2001.
 - [25] J. T. Johnson and R. J. Burkholder, “Coupled canonical grid/discrete dipole approach for computing scattering from objects above or below a rough interface,” *IEEE Trans. Geosci. & Remote Sensing*, vol. 39, no. 6, pp. 1214-1220, June 2001.

- [26] M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo Simulations of Electromagnetic Wave Scattering from Random Rough Surface with 3-D Penetrable Buried Object: Mine Detection Application Using the SDFMM," *J. Optical Society of America A*, vol. 18, no. 12, pp. 3077-3084, December 2001.
- [27] M. El-Shenawee, "The steepest descent fast multipole method for scattering from two penetrable shallow objects buried under two-dimensional randomly rough surface," submitted in June 2001 to the *IEEE Trans. Antennas & Prop.*
- [28] M. El-Shenawee and C. Rappaport, "Monte Carlo Simulations for the Statistics of Clutter in Minefields: Plastic Target Buried Near a Dielectric Object Beneath Two-Dimensional Randomly Rough Ground," submitted in November 2001 to the *IEEE Trans. Geosci. & Rem. Sensing*.
- [29] W. C. Chew, *Waves and fields in inhomogeneous media*, IEEE Press, 1995.
- [30] M. El-Shenawee, V. Jandhyala, E. Michielssen and W.C. Chew, "An Enhanced Steepest Descent Fast Multipole Method for the Analysis of Scattering from Two Dimensional Multilayered Rough Surfaces," *Proc. of the IEEE APS/URSI '98*, Atlanta, Georgia, June 16-21, 1998.
- [31] Y. Zhang and E. Bahar, "Mueller matrix elements that characterize scattering from coated random rough surfaces," *IEEE Trans. Antennas & Prop.*, vol. 47, no. 5, pp. 949-955, May 1999.
- [32] M. El-Shenawee, "The Multiple Interaction Model for Non-Shallow Scatterers Buried Beneath Two-Dimensional Random Rough Surfaces," *IEEE Trans. on Geosci. & Remote Sensing*, in print.
- [33] V. Rokhlin, "Rapid solution of integral equations of scattering theory in two dimensions," *J. Comput. Phys.*, vol. 36, pp. 414-439, 1990.
- [34] R. Coifman, V. Rokhlin and S. Wandzura, "The fast multipole method for the wave equation: A pedestrian description," *IEEE Antennas & Prop. Mag.*, vol. 35, no. 3, pp. 7-12, June 1993.
- [35] C. C. Lu and W. C. Chew, "A multilevel fast-algorithm for solving a boundary integral equation of wave scattering," *Microwave Opt. Tech. Let.*, vol. 7, pp. 466-470, July 1994.
- [36] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [37] V. Jandhyala, E. Michielssen, B. Shanker and W.C. Chew, "A combined steepest descent-fast multipole algorithm for the fast analysis of three-dimensional scattering by rough surfaces," *IEEE Trans. Geosci. & Remote Sensing*, vol. 36, no. 3, pp. 738-748, May 1998.
- [38] M. El-Shenawee, V. Jandhyala, E. Michielssen and W. C. Chew, "The steepest descent fast multipole method (SDFMM) for solving combined field integral equation pertinent to rough surface scattering," *Proc. of the IEEE APS/URSI '99 Conf.*, Orlando, Florida, pp. 534-537, July 1999.
- [39] S. Li, C. H. Chan, L. Tsang, Q. Li and L. Zhou, "Parallel implementation of the sparse matrix/canonical grid method for the analysis of two-dimensional random rough surfaces (three-dimensional scattering problem) on a Beowulf system," *IEEE Trans. Geosci. & Remote Sensing*, vol. 38, no. 4, pp. 1600-1608, July 2000.
- [40] D. Torrungrueng, H. Chou and J. T. Johnson, "A novel acceleration algorithm for the computation of scattering from two-dimensional large scale perfectly conducting random

- rough surfaces with the forward-backward method,” *IEEE Trans. Geosci. & Remote Sensing*, vol. 38, no. 7, pp. 1656-1666, July 2000.
- [41] D. Jiang, W. Meleis, M. El-Shenawee, E. Mizan, M. Ashouei and C. Rappaport, “Parallel implementation of the steepest descent fast multipole method (SDFMM) on a Beowulf cluster for subsurface sensing applications,” *IEEE MWCL*, vol. 18, no. 1, pp. 1-3, January 2002.
 - [42] P. Tran and A. A. Maradudin, “Scattering of a scalar beam from a two-dimensional randomly rough hard wall: Enhanced backscatter,” *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.
 - [43] S. M. Rao, D. R. Wilton and A. W. Glisson, “Electromagnetic scattering by surfaces of arbitrary shape,” *IEEE Trans. Antennas & Prop.*, vol. 30, no. 3, pp. 409-418, May 1982.
 - [44] L. Medgyesi-Mitschang, J. Putnam and M. Gedera, “Generalized method of moments for three-dimensional penetrable scatterers,” *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1383-1398, April 1994.
 - [45] N. Garcia and E. Stoll, “Monte Carlo calculation for electromagnetic-wave scattering from random rough surfaces,” *Phy. Rev. Let.*, vol. 52, no. 20, pp. 1798-1801, May 1984.
 - [46] R. W. Freund, “A transpose-free quasi-minimal residual algorithm for non-hermitian linear systems,” *SIAM J. Sci. Comput.*, vol. 14, no. 2, pp. 470-482, March 1993.
 - [47] M. El-Shenawee, “The multiple interaction model for non-shallow scatterers buried beneath two-dimensional random rough surfaces,” *IEEE Trans. on Geosci. & Remote Sensing*, in print.

List of Figures

Fig. 1. (a) General closed 3-D scatterers R_2 , R_3 and R_4 , (b) modeling closed interfaces S_1 and S_2 by open rough surface interfaces in which negligible currents are excited on the dotted lines, and (c) cross section of the multilayered geometry showing the FMM finest blocks.

Fig. 2. (a) The relative error in surface current vector C and the SDFMM memory requirements versus the finest block size for the multilayered ground (flat/flat) of height $H = 0.5\lambda_0$ with oblate spheroid buried at $z = -0.25\lambda_0$, (b) The relative error in surface current vector $\|C\|$ versus the height H of the multilayered ground (flat/flat) with oblate spheroid buried at $z = -0.3\lambda_0$ (see Fig. 1c).

Fig. 3. The normalized RCS for (a) lossless multilayered ground with no buried objects, (b) lossy multilayered ground with buried spheroid, (c) lossy multilayered ground with no buried objects, all for $\vartheta^i = 0^\circ$, (d) Same data of (b) but at $\vartheta^i = 10^\circ$. Lossless implies $\epsilon_{r3} = 2.9$, Case 1 $\epsilon_{r2} = 6.5, \epsilon_{r4} = 2.5$, Case 2 $\epsilon_{r2} = 2.5, \epsilon_{r4} = 6.5$. Lossy implies $\epsilon_{r3} = 2.9 - j0.0092$, Case 1 $\epsilon_{r2} = 6.5 - j0.1, \epsilon_{r4} = 2.5 - j0.18$, Case 2 $\epsilon_{r2} = 2.5 - j0.18, \epsilon_{r4} = 6.5 - j0.1$.

Fig. 4. (a) The norm of the electric and magnetic currents $\|C\|$ on each surface relative to the norm of the currents on all surfaces $\|C_{\text{total}}\|$ plotted versus the height H , (b) the normalized RCS at backscatter versus the height H . The results are for lossless multi-layered ground with buried spheroid at $z = -0.3\lambda_0$, $\vartheta^i = 0^\circ$, H-polarization, $\epsilon_{r3} = 2.9$, $\epsilon_{r2} = 6.5$ and $\epsilon_{r4} = 2.5$ (see Fig. 1c).

Fig. 5. (a) The norm of the electric and magnetic currents $\|C\|$ on each surface relative to the norm of the currents on all surfaces $\|C_{\text{total}}\|$ plotted versus ϵ_{r4} , (b) the normalized RCS at backscatter versus ϵ_{r4} . The results are for lossless multilayered ground with buried spheroid at $z = -0.3\lambda_0$, $\vartheta^i = 0^\circ$, H-polarization, $\epsilon_{r3} = 2.9$, $\epsilon_{r2} = 6.5$ and $H = 0.6\lambda_0$ (see Fig. 1c).

Fig. 6. (a) The norm of the electric and magnetic currents $\|C\|$ on each surface relative to the norm of the currents on all surfaces $\|C_{\text{total}}\|$ plotted versus ϵ_{r2} , (b) the normalized RCS at backscatter versus ϵ_{r2} . The results are for lossless multilayered ground with buried spheroid at $z = -0.3\lambda_0$, $\vartheta^i = 0^\circ$, H-polarization, $\epsilon_{r3} = 2.9$, $\epsilon_{r4} = 6.5$ and $H = 0.6\lambda_0$.

Fig. 7. (a) The norm of the electric and magnetic currents $\|C\|$ on each surface relative to the norm of the currents on all surfaces $\|C_{\text{total}}\|$ plotted versus incident angle θ^i , (b) the normalized VV-RCS versus the incident angle θ^i , (c) the normalized HV-RCS versus the incident angle θ^i at backscatter direction ($\theta^i = \theta^s, \phi^i - \phi^s = \pi$). The results are for lossy multi-layered ground with buried spheroid at $z = -0.3\lambda_0$, $\epsilon_{r2} = 2.5 - j0.18$, $\epsilon_{r3} = 2.9 - j0.0092$, $\epsilon_{r4} = 6.5 - j0.1$ and $H = 0.6\lambda_0$.

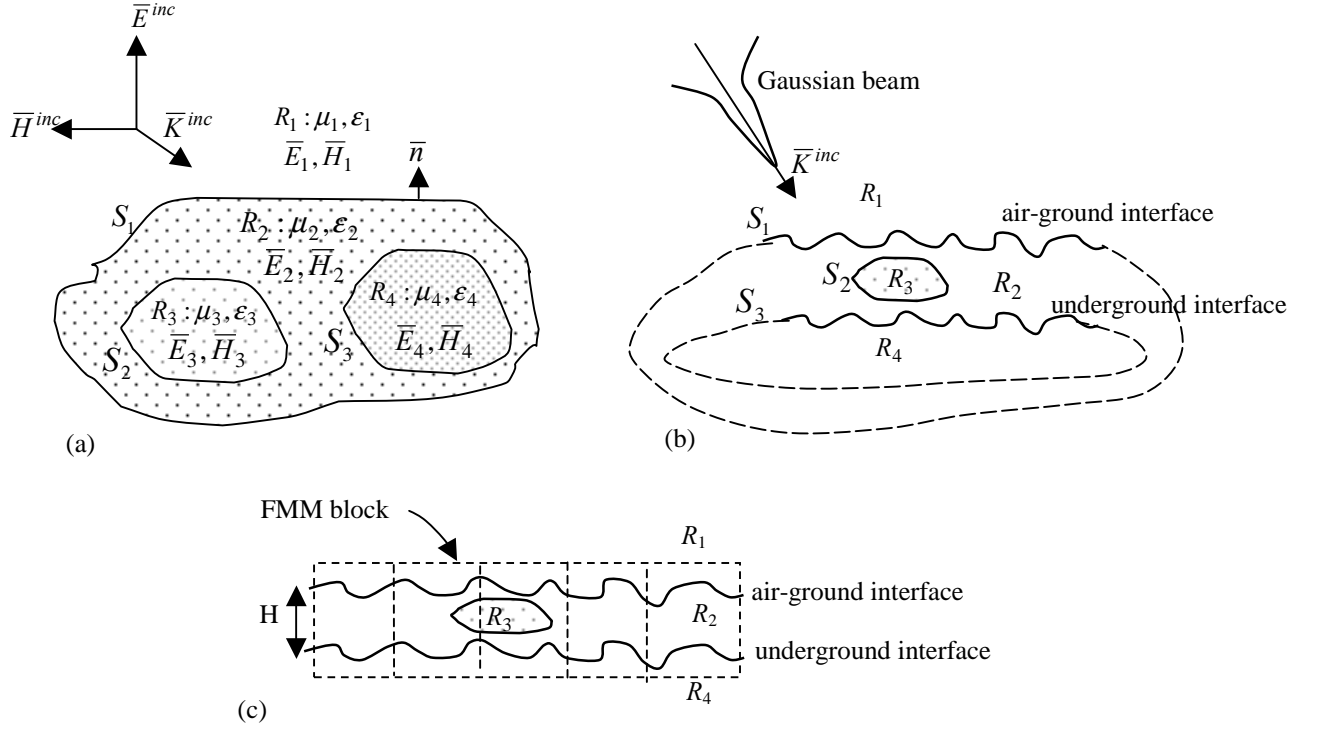


Fig. 1. (a) General closed 3-D scatterers R_2 , R_3 and R_4 , (b) modeling closed interfaces S_1 and S_2 by open rough surface interfaces in which negligible currents are excited on the dotted lines, and (c) cross section of the multilayered geometry showing the FMM finest blocks.

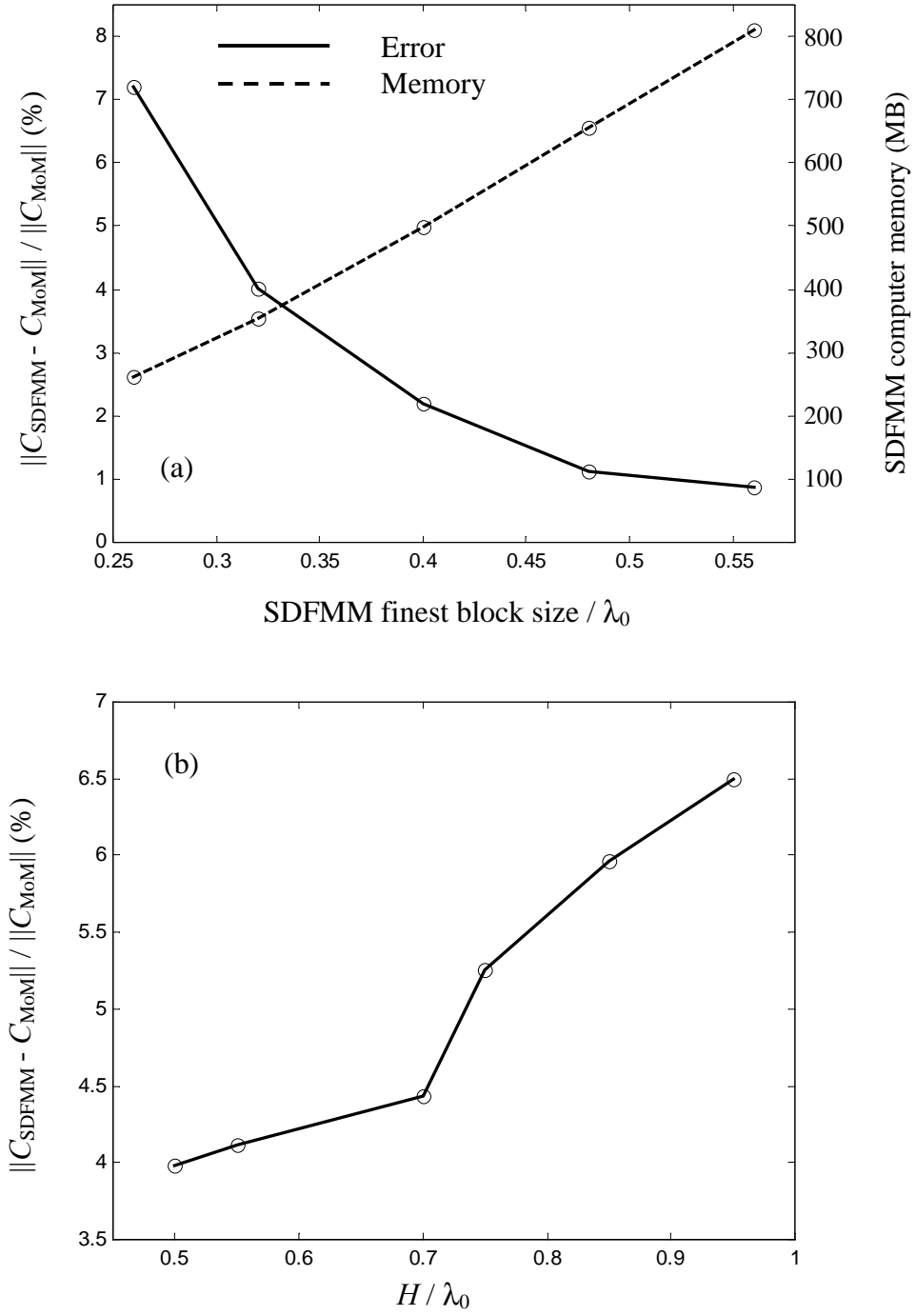
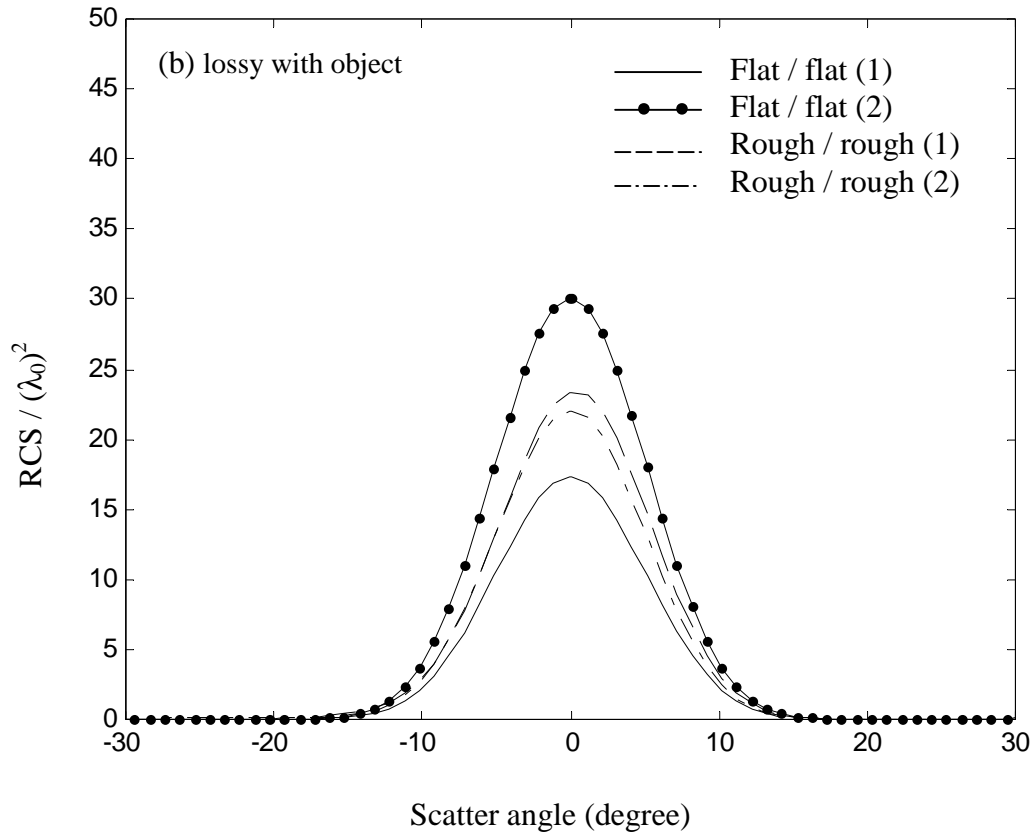
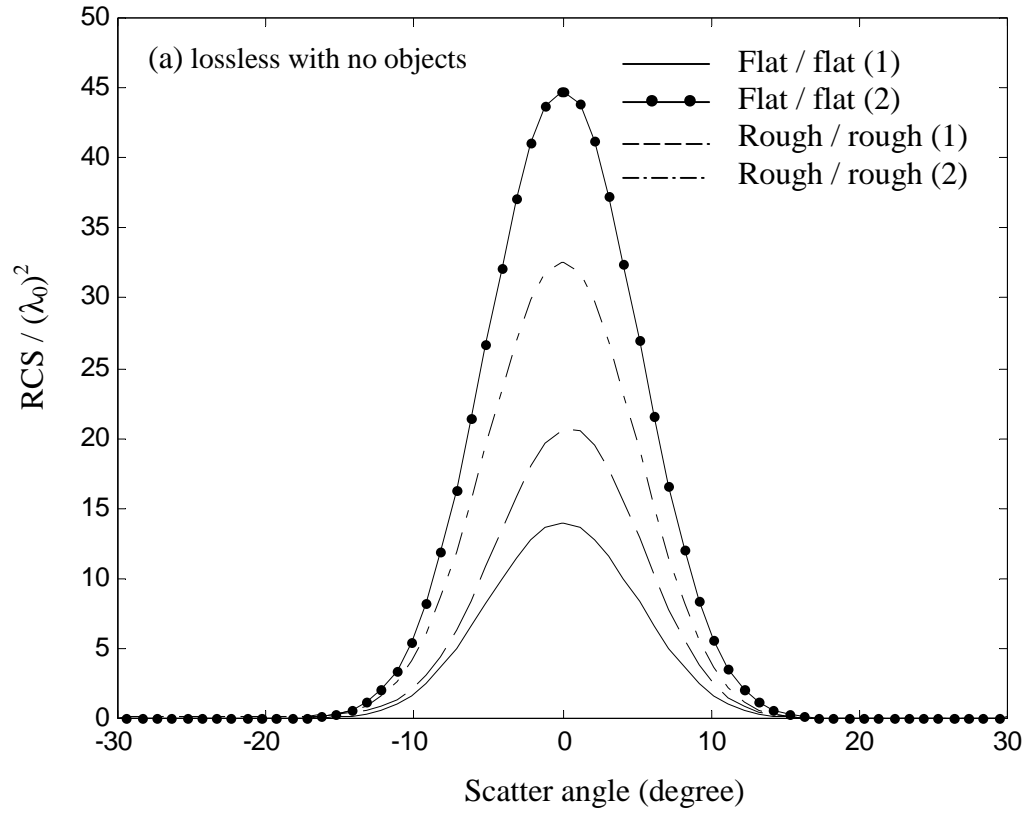


Fig. 2. (a) The relative error in surface current vector $\|C\|$ and the SDFMM memory requirements versus the finest block size for the multilayered ground (flat/flat) of height $H = 0.5\lambda_0$ with oblate spheroid buried at $z = -0.25\lambda_0$, (b) The relative error in surface current vector C versus the height H of the multilayered ground (flat/flat) with oblate spheroid buried at $z = -0.3\lambda_0$ (see Fig. 1c).



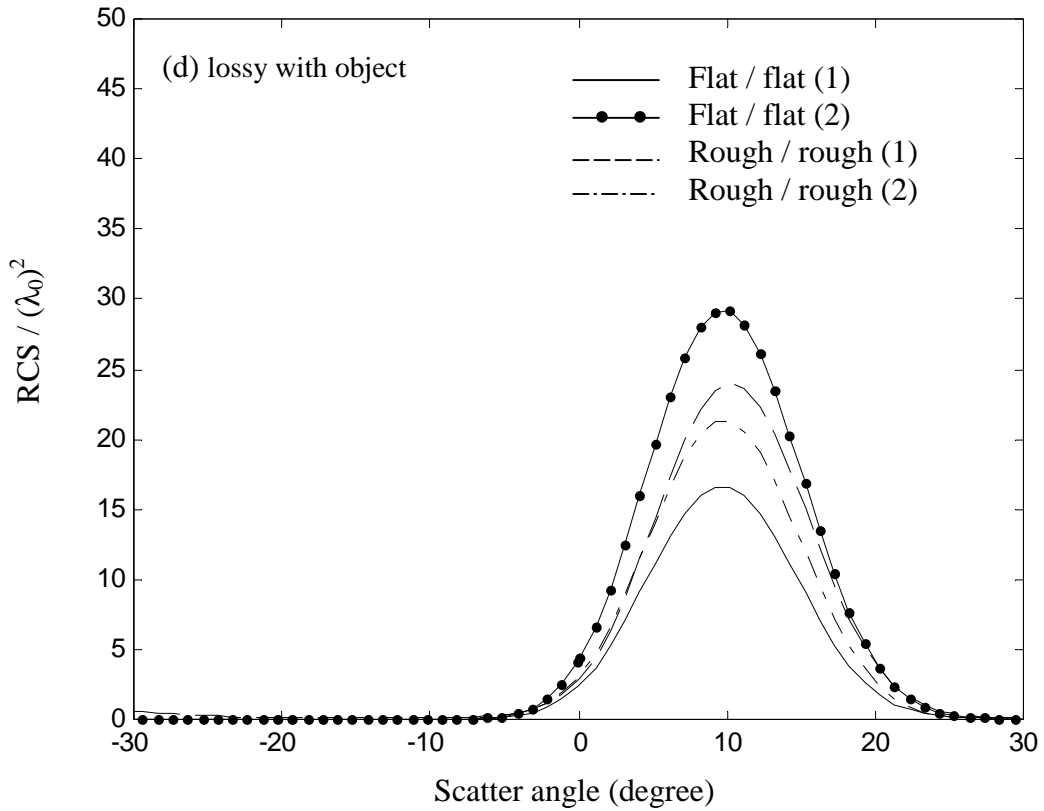
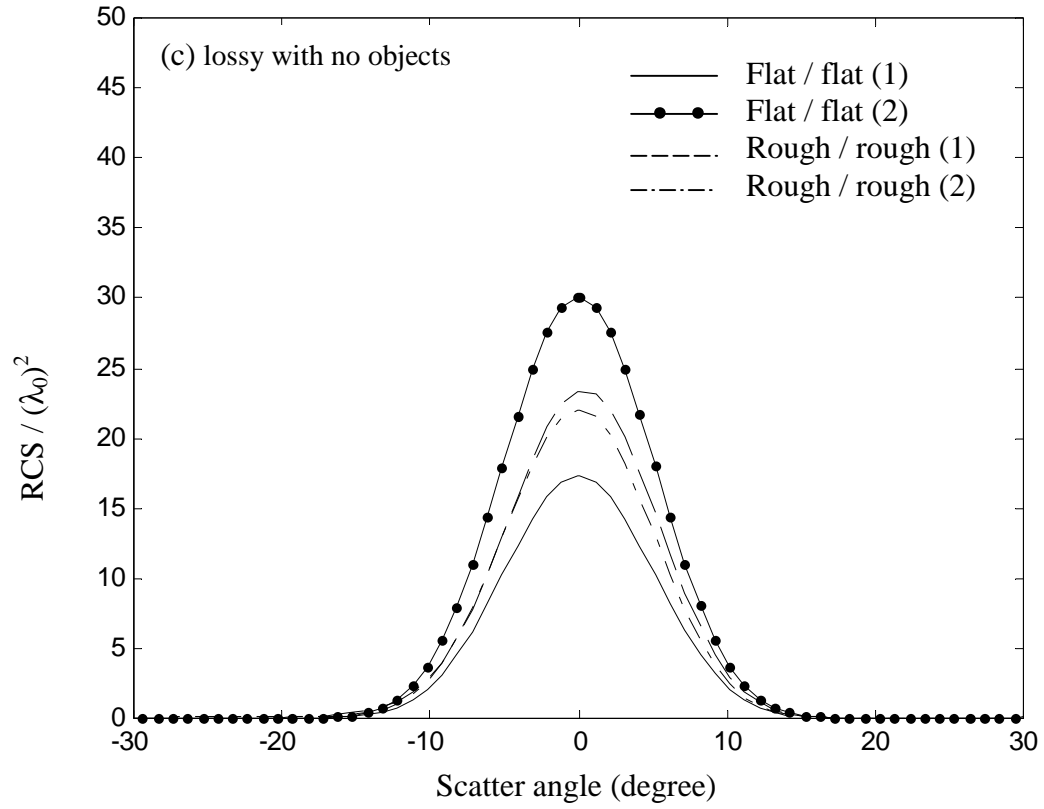


Fig. 3. (a)-(c) for $\vartheta^i = 0^\circ$ and (d) for $\vartheta^i = 10^\circ$.

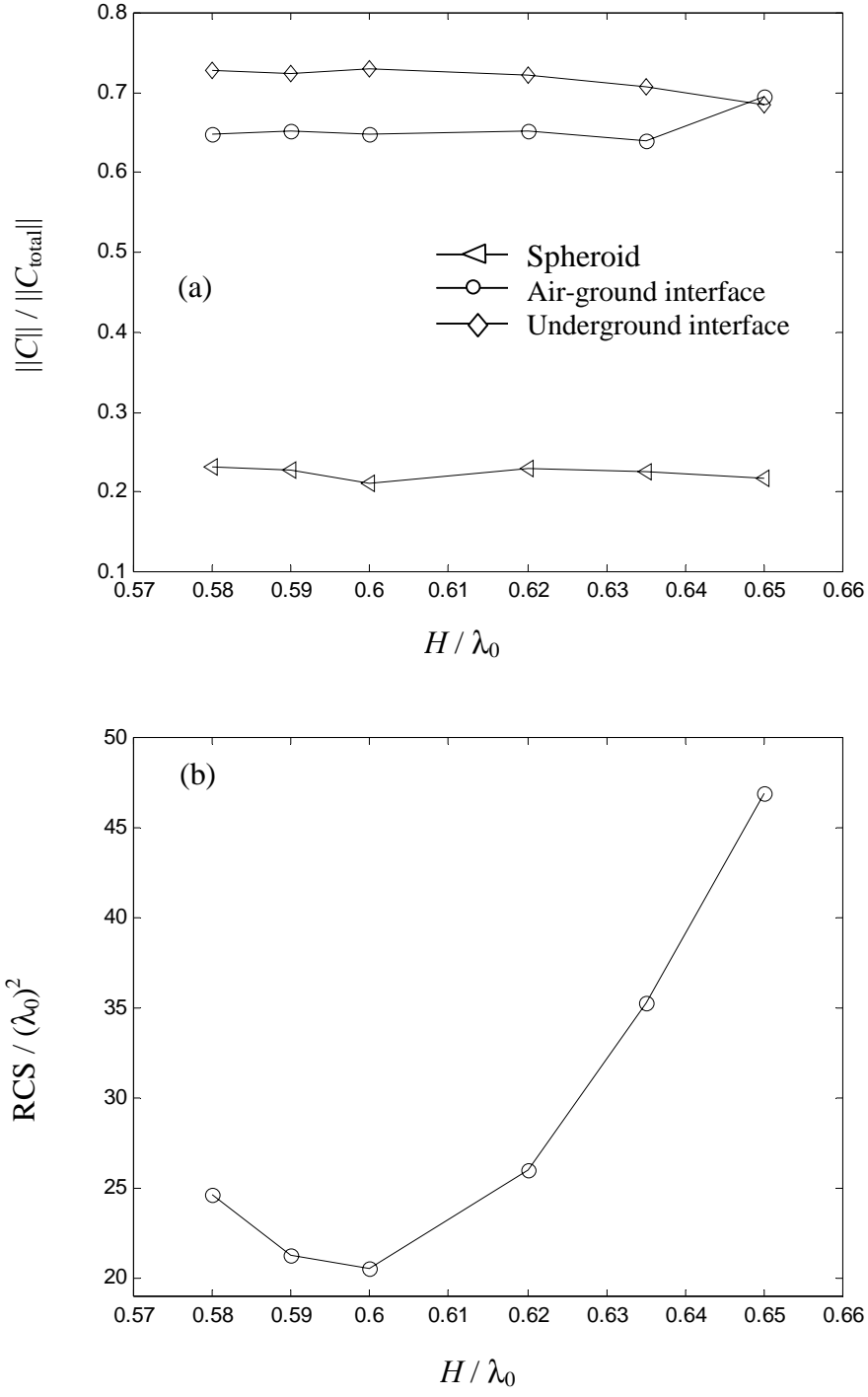


Fig. 4. (a) The norm of the electric and magnetic currents $\|C\|$ on each surface relative to the norm of the currents on all surfaces $\|C_{\text{total}}\|$ plotted versus the height H , (b) the normalized RCS at backscatter versus the height H . The results are for lossless multi-layered ground with buried spheroid at $z = -0.3\lambda_0$, $\vartheta^i = 0^\circ$, H-polarization, $\epsilon_{r3} = 2.9$, $\epsilon_{r2} = 6.5$ and $\epsilon_{r4} = 2.5$ (see Fig. 1c).

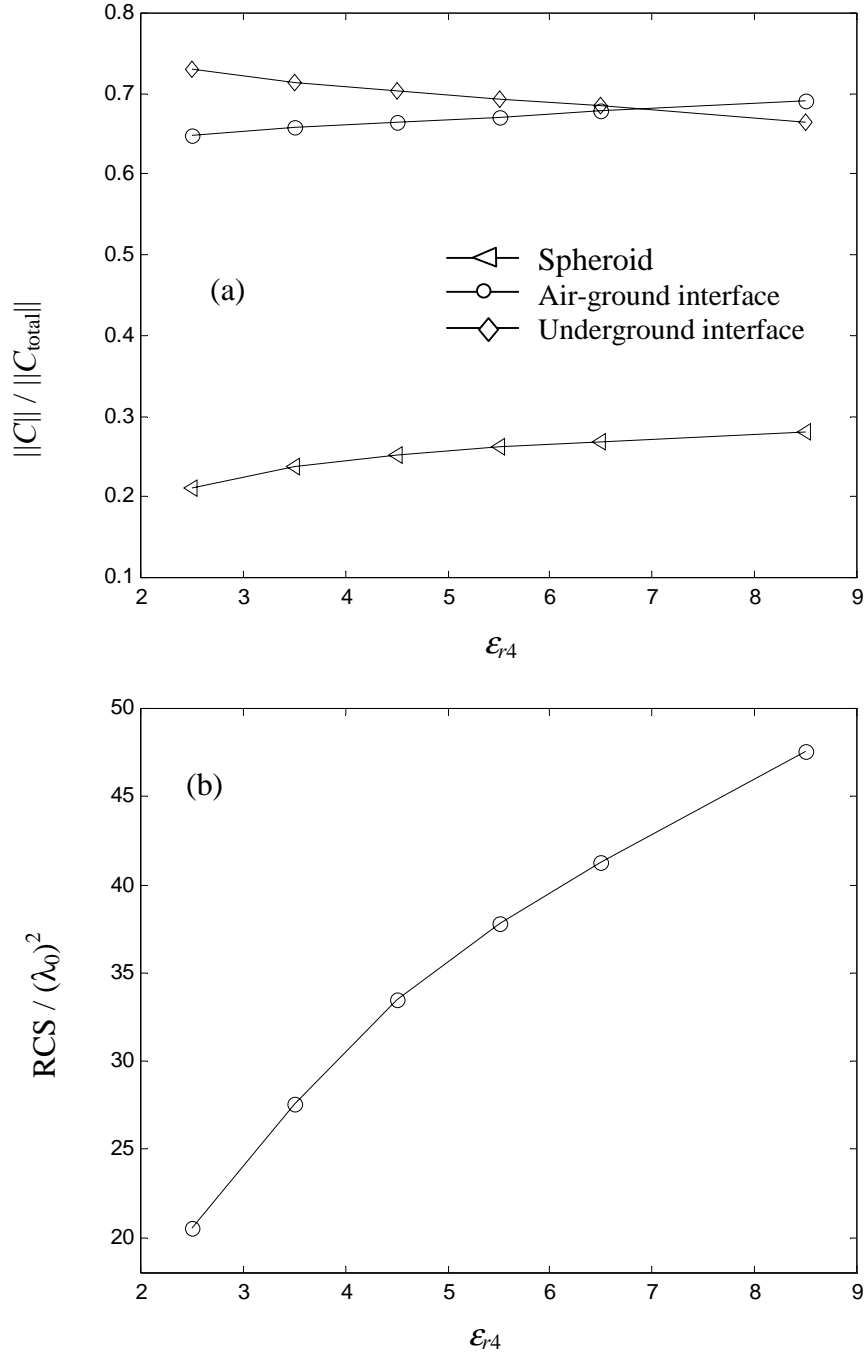


Fig. 5. (a) The norm of the electric and magnetic currents $\|C\|$ on each surface relative to the norm of the currents on all surfaces $\|C_{\text{total}}\|$ plotted versus ϵ_{r4} , (b) the normalized RCS at backscatter versus ϵ_{r4} . The results are for lossless multilayered ground with buried spheroid at $z = -0.3\lambda_0$, $\vartheta^i = 0^\circ$, H-polarization, $\epsilon_{r3} = 2.9$, $\epsilon_{r2} = 6.5$ and $H = 0.6\lambda_0$ (see Fig. 1c).

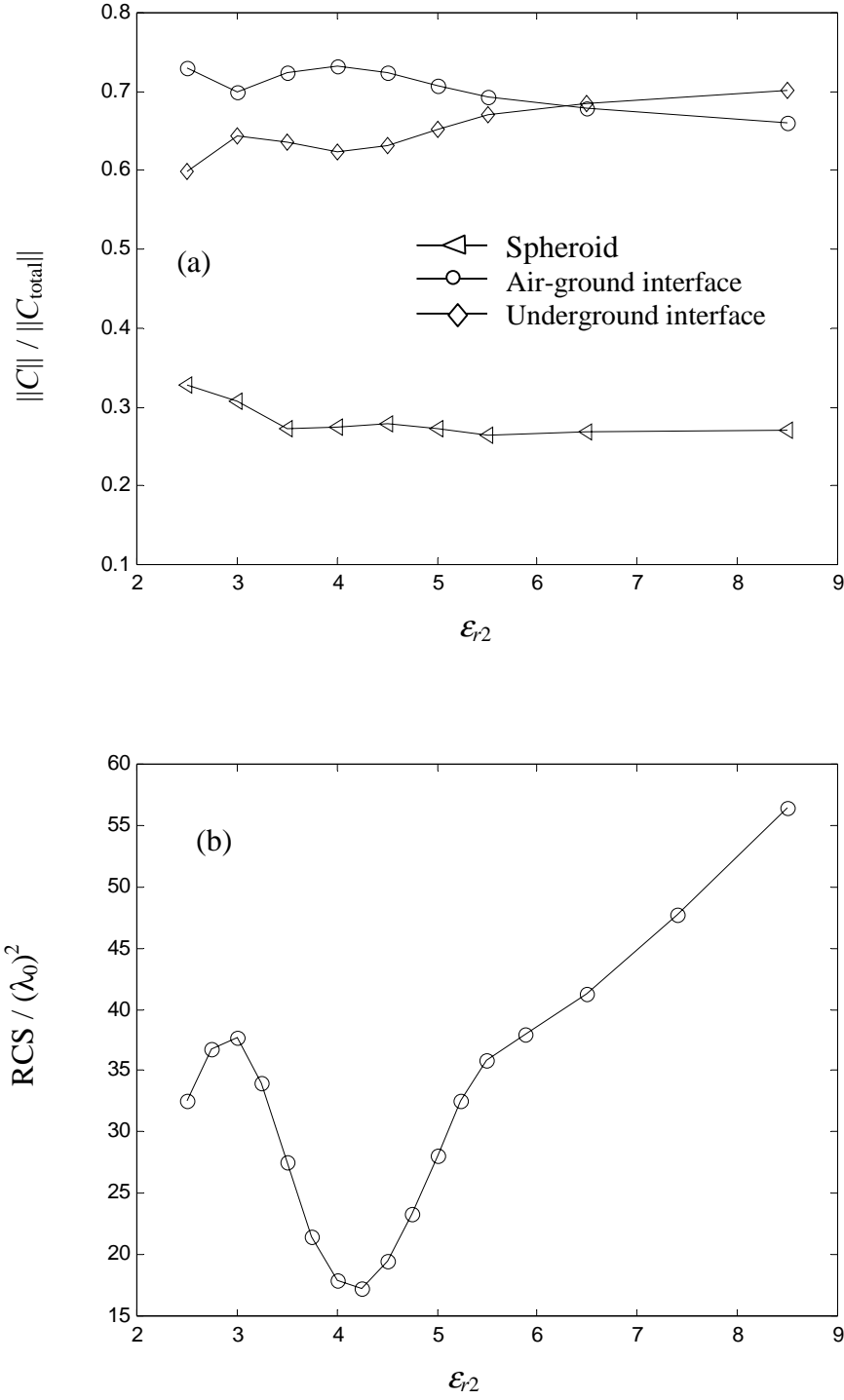
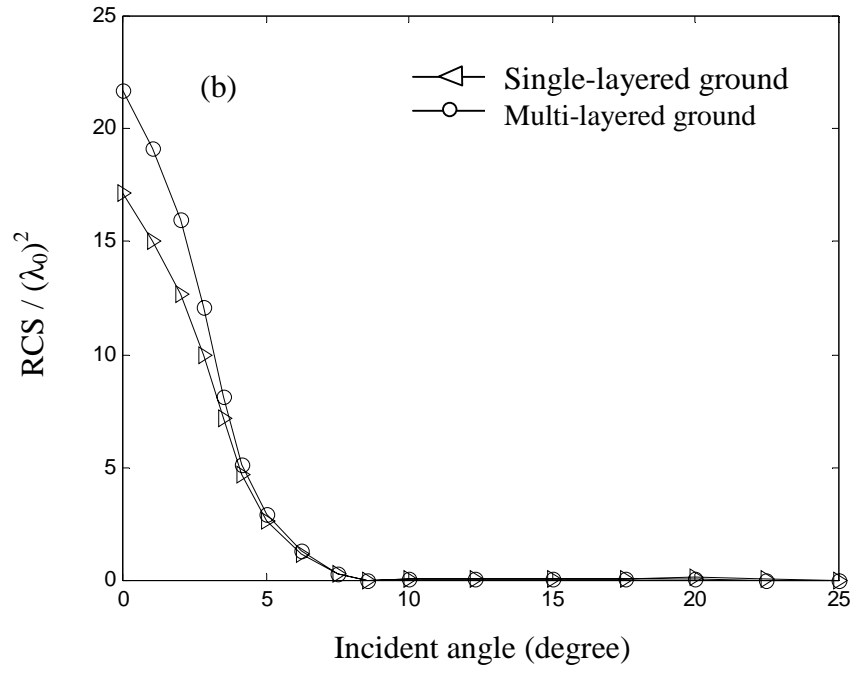
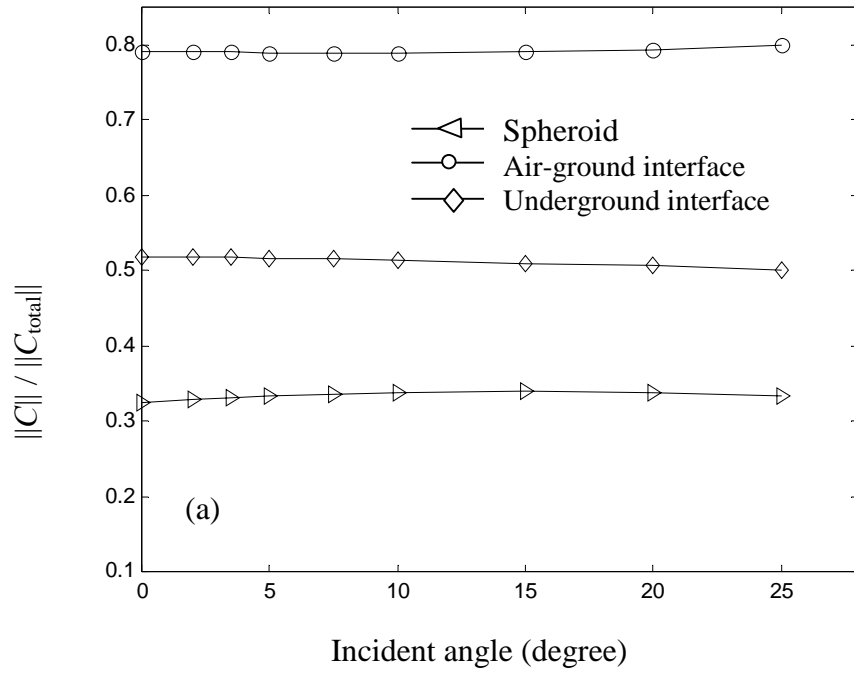


Fig. 6. (a) The norm of the electric and magnetic currents $\|C\|$ on each surface relative to the norm of the currents on all surfaces $\|C_{\text{total}}\|$ plotted versus ϵ_{r2} , (b) the normalized RCS at backscatter versus ϵ_{r2} . The results are for lossless multilayered ground with buried spheroid at $z = -0.3\lambda_0$, $\vartheta^i = 0^\circ$, H-polarization, $\epsilon_{r3} = 2.9$, $\epsilon_{r4} = 6.5$ and $H = 0.6\lambda_0$.



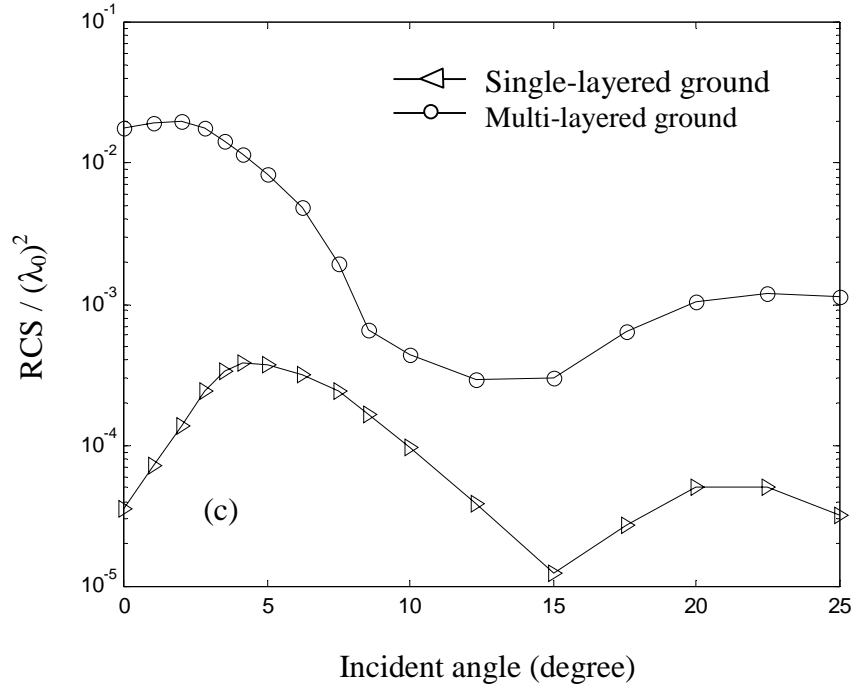


Fig. 7. (a) The norm of the electric and magnetic currents $\|C\|$ on each surface relative to the norm of the currents on all surfaces $\|C_{\text{total}}\|$ plotted versus incident angle θ^i , (b) the normalized VV-RCS versus the incident angle θ^i , (c) the normalized HV-RCS versus the incident angle θ^i at backscatter direction ($\theta^i = \theta^s, \phi^i - \phi^s = \pi$). The results are for lossy multi-layered ground with buried spheroid at $z = -0.3\lambda_0$, $\epsilon_{r2} = 2.5 - j0.18$, $\epsilon_{r3} = 2.9 - j0.0092$, $\epsilon_{r4} = 6.5 - j0.1$ and $H = 0.6\lambda_0$.

Electromagnetic Wave Scattering from Multi-layered Random Rough Surfaces with Buried Dielectric Object

Magda El-Shenawee
Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701
Fax: 501-575-7967, Tel: 501-575-6582
magda@uark.edu

Electromagnetic sensing of buried objects in the presence of a random rough interface is a crucial step for subsurface detection problems in general. Most theoreticians and experimentalists agreed that surface roughness constitutes a major source of clutter (i.e., noise) in the received electromagnetic signals. However, due to the complexity of the problem, the ground surface is often assumed to be single layered and *not* multi-layered. This is not the case in the real environment but no work has yet been published on scattering from objects buried beneath multi-layered randomly rough ground. Therefore, the objective of this work is to investigate the scattering of electromagnetic waves from a penetrable shallow object buried in 2-D multi-layered random rough surfaces. It is known that the computational complexity of the scattering problem dramatically increases for the multi-layered rough interface.

A rigorous electromagnetic model has been developed for scattering from inhomogeneous rough ground surfaces. This model is based on the classical equivalence theorem and the method of moments (MoM) that is dramatically accelerated by implementing the Steepest Descent Fast Multipole Method (SDFMM). Four different homogeneous regions are involved in this application; air, dielectric object, upper soil layer and lower soil layer. A tapered Gaussian beam is used to illuminate the multi-layered ground surface. The rough surfaces are characterized with Gaussian statistics for the height and for the autocorrelation function. Since the SDFMM was originally developed for quasi-planar structures where the whole height of the 3-D scatterer should be in the order of one free space wavelength, it is necessary to validate the SDFMM implemented for this application versus the MoM. The relative norm of the error in surface current is presented. The computer memory requirement is plotted versus the SDFMM finest block size for the multi-layered ground with the buried object. It is necessary to mention that the air-ground interface and the underground layer could be flat and/or could be random rough surfaces with different roughness parameters and lossy or lossless soil. Numerical results representing the RCS of the multi-layered ground with the buried object are shown. The effect of the physical characteristics and the surface roughness of the multi-layers are investigated. The thickness of the multi-layer ground is varied to study its influence on the buried object signature.

The results of this work could increase the efficacy of the present subsurface sensing methods by accounting for the multi-layer nature of the rough ground.

Scattering from Non-Shallow Targets Buried Beneath Two-Dimensional Random Rough Surfaces Using the Multiple Interaction Model

Magda El-Shenawee

Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701
Tel: 501-575-6582, Fax: 501-575-7967
magda@uark.edu

The multiple interaction approach is used with the robust Steepest Descent Fast Multipole Method (SDFMM) to compute the signature of non-shallow penetrable scatterers buried beneath 2-D random rough surfaces. The most attractive feature of the proposed model is removing the quasi-planar structure constraint of the Steepest Descent Fast Multipole Method (SDFMM) when used in analyzing non-quasi-planar scatterers. The basic idea of the multiple interaction model is to decompose certain non-quasi-planar structures into two quasi-planar scatterers where the conventional SDFMM can be applied separately to each one. The interactions between the sub-quasi-planar scatterers are calculated directly using the electromagnetic vector potentials near-field expressions. Significant reductions in the CPU time and computer memory are achieved by using the SDFMM in the model. A variety of geometries are used to test the model and their numerical results are validated with the conventional MoM.

The results show that the buried object's signature is largely due to the first interaction mechanism (i.e. ground-object-ground). However, the contribution of each additional interaction is explicitly calculated using the model. Interestingly, the contributions from repeating this mechanism become insignificant especially for lossy background soil. This conclusion depends on the physical properties of the scatterer.

The multiple interaction model successfully demonstrates the exploitation of the SDFMM robustness when applied to the multilayered rough ground where the burial depth of the underground rough layer is on the order of a wavelength.

**Topics: #21 Rough Surface Scattering or #2 computational electromagnetics,
Oral presentation**

- 1) Scattering from Non-Shallow Targets Buried Beneath Two-Dimensional Random Rough Surfaces Using the Multiple Interaction Model
- 2) Magda El-Shenawee
- 3) University of Arkansas
- 4) 3217 Bell Engineering Center
Fayetteville, AR 72701
Tel.: 501-575-6582
Fax: 501-575-7967
Email: magda@uark.edu

Subsurface Sensing of Targets Buried Beneath 2-D Multilayered Random Rough Surfaces: Use of the Steepest Descent Fast Multipole Method

Magda El-Shenawee

Department of Electrical Engineering
3217 Bell Engineering Center
University of Arkansas
Fayetteville, Arkansas 72701
Tel: (501)-575-6582, Fax: (501)-575-7967
magda@uark.edu

The scattering of electromagnetic waves from a penetrable shallow target buried beneath 2-D multilayered random rough surfaces will be presented. There are several applications for this work, e.g. detection of anti-personnel mines, anti-tank mines, water and/or gas pipes, location of underground water, etc. In reality, these targets are buried under the Earth's surface, which is a randomly rough interface and also is not simply composed of a single layer, but is a multilayered media. The closer the real environment is incorporated into the electromagnetic model, the more accurate and practical inferences can be gained from the numerical results. Without modeling the multiple ground layers, many targets *cannot be detected*. One unobtrusive way these buried targets can be detected is by bombarding the Earth's surface with electromagnetic waves, and comparing the scattered signature of the ground alone with that of the ground with the buried target.

A rigorous electromagnetic model based on the equivalence theorem and the method of moments (MoM) is developed to analyze this 3-D scattering problem. Three layers are considered in this work; air, dry-soil and wet-soil. The penetrable target is buried between the air/dry-soil interface and the dry/wet-soil interface. The Steepest Descent Fast Multipole Method (SDFMM) is implemented to significantly accelerate the computations of the unknown electric and magnetic surface currents. The effect of the lossy underground rough layer (wet soil) on the target signature will be investigated. Moreover, images based on the scattered electric fields for the buried target will be presented.

Invited paper.

Session organizer: Alex Maradudin, UCI.

- 1) Subsurface Sensing of Targets Buried Beneath 2-D Multilayered Random Rough Surfaces: Use of the Steepest Descent Fast Multipole Method.
- 2) Magda El-Shenawee
- 3) University of Arkansas
- 4) 3217 Bell Engineering Center
Fayetteville, AR 72701
Tel.: 501-575-6582
Fax: 501-575-7967
Email: magda@uark.edu

The Steepest Descent Fast Multipole Method for Clutter Statistics in Minefields Using Monte Carlo Simulations

Magda El-Shenawee¹ and Carey Rappaport²

¹Department of Electrical Engineering
University of Arkansas
Fayetteville, Arkansas 72701
Tel: (501)-575-6582, Fax: (501)-575-7967
magda@uark.edu

²Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
Tel: 617-373-2064, Fax: 617-373-8627
rappaport@neu.edu

A rigorous electromagnetic model has been developed to analyze the scattering mechanism of a target buried near a clutter-object under the two-dimensional random rough ground (3-D scattering problem). In realistic landmine fields, the anti-personnel (AP) nonmetallic mine is often buried nearby a rock, tree root, etc. The presence of a second object buried near the nonmetallic mine can easily obscure the target and/or cause a false alarm during the detection process.

The rigorous model is based on the classical electromagnetic equivalence theorem leading to producing six new integral equations. Using the Method of Moment (MoM), the new integral equations are transformed into a linear system of equations to be solved for the unknown electric and magnetic currents on the surface of three scatterers; rough ground, target and clutter-object. The MoM impedance matrix completely represents every interaction between these three scatterers. The Steepest Descent Fast Multipole Method (SDFMM) is used to tremendously accelerate the computations of the unknown MoM surface currents.

In previous work, we thoroughly investigated the effect of ground roughness on the signature of the target when it is buried alone under the ground. In this work, we will present numerical results for parametric investigations of the objects proximity, orientations, materials, and shapes. The results show that in certain situations, the target can be completely obscured due to the presence of the nearby clutter-object (e.g., tree root). In other cases a false indication of presence of a third buried object is observed. When the sources of clutter (e.g. the rough ground and the clutter object) are removed, by subtracting the return from both the rough ground and the clutter-object, the signature of the target can be clearly observed and analyzed. The numerical results show that the ground roughness along with the separation distance, between the target and the clutter-object, play a significant role on the probability of true or false alarm in the detection process.

Topics: #21 Rough Surface Scattering, #2 computational electromagnetics, or # 18 Subsurface Images and detection technology (GPR)

Oral presentation

1) The Steepest Descent Fast Multipole Method for Clutter Statistics in Minefields
Using Monte Carlo Simulations.

2) Presenting Author:

Magda El-Shenawee
University of Arkansas
3217 Bell Engineering Center
Fayetteville, AR 72701
Tel.: 501-575-6582
Fax: 501-575-7967
Email: magda@uark.edu

Co-author:

Carey Rappaport
Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
Tel: 617-373-2064, Fax: 617-373-8627
rappaport@neu.edu

Remote Sensing of Penetrable Objects Buried Beneath 2-D Random Rough Surfaces Using the Mueller Matrix Elements

Magda El-Shenawee
Department of Electrical Engineering
University of Arkansas
Fayetteville, Arkansas 72701
Tel: (501)-575-6582, Fax: (501)-575-7967
magda@uark.edu

ABSTRACT

The modified Mueller matrix elements for electromagnetic scattering from penetrable objects buried under 2-D random rough surfaces are investigated. This matrix relates the incident with the scattered waves and it contains different combinations of the fully polarimetric scattering matrix elements. The statistical average of each Mueller matrix element is computed based on the Monte Carlo simulations with exploiting the speed of the 3-D Steepest Descent Fast Multipole Method (SDFMM). The numerical results clearly show that relying only on the co- and/or the cross-polarized intensities (i.e., vv , hh , vh and hv) is not sufficient for sensing the buried objects. However, examining all the sixteen Mueller matrix elements significantly increases the possibility of detecting these objects. This technique can be used in remote sensing of scatterers buried beneath the rough ground.

Keywords: Subsurface sensing, Mueller matrix elements, multiple buried objects, rough surface scattering, radar cross section.

I. INTRODUCTION

Remote sensing of objects buried under the rough ground has tremendous civil and defense applications, e.g. archeology discovery, mine detection, sensing underground resources, etc. In previous work [1]-[3], the scattered electric near-field was calculated to simulate the ground

penetrating radar (GPR) measurements. A single penetrable and/or perfect electric conductor (PEC) object was buried under the rough ground, where the average of the scattered electric near-fields over the azimuth angle was presented in [1] while the average over the rough surface realizations was presented in [2]. In [3], two dielectric objects were buried under the random rough ground and the electromagnetic near-field interference between the objects was studied. The reported results showed the significant distortion in the near-field signals caused by the presence of the rough air-ground interface. In [1]-[3], the signature of the target was often obtained by removing the background from the received signals (i.e. by subtracting the scattering from the rough ground with no buried objects). In [4], the angular correlation function (ACF) and the radar cross section (RCS) were presented for a PEC sphere buried under the rough ground in which the authors presented the advantage of computing the ACF over the RCS. In [5], the sixteen Mueller matrix elements were analytically obtained for scattering from coated 2-D random rough surface where the configuration represented three-layer geometry (the air and two layered-rough interfaces) with no buried objects. The Mueller matrix elements, which relate the incident with the scattered waves, are defined in terms of the modified Stokes vector [6]-[8]. The results reported in [5] showed a great sensitivity of some Mueller matrix elements (m_{34} and m_{43}) in the backscatter direction to the thickness of the coated layer. In other words, these two elements showed considerable differences due to the underground inhomogeneity, which was not the case for the other Mueller matrix elements (m_{11} , m_{12} , m_{22} , etc.). This finding is the motivation of the current work, however, the configuration here represents two different dielectric objects buried under the 2-D random rough ground as shown in Fig. 1.

It is necessary to differentiate between the results reported in [2] and the current work. In [2], the statistical average of the buried object's signature was computed based on the scattered

electric fields in the near-zone. In the current work, the statistical average of the scattered electric fields from one and/or from two buried objects is calculated in the far-zone and presented in terms of the Mueller matrix elements. More importantly, the subtraction process often used in [1]-[3] is not used in the current work. In other words, the far-fields scattered from the rough ground with the buried objects are directly compared with those scattered from the rough ground with no buried objects.

The formulations of the problem are given in Section II, the numerical results are presented in Section III and the conclusions are stated in Section IV. More details are given in Appendices A and B.

II. FORMULATIONS

The integral equation-based rigorous electromagnetic model was developed in [3] for scattering from multiple objects buried under the rough ground. This technique is employed to calculate the unknown method of moments (MoM) surface currents on the rough ground and on both buried objects (see Fig. 1). The inhomogeneous scatterer is composed of four different regions; the air, the ground, and two different materials for the objects where the relative permittivity and permeability are ϵ_1 and μ_1 for the air, ϵ_2 and μ_2 for the ground, ϵ_3 and μ_3 for the first object, and ϵ_4 and μ_4 for the second object. The unknown equivalent electric and magnetic surface currents are \bar{J}_1 and \bar{M}_1 on the ground, \bar{J}_2 and \bar{M}_2 on the first object and \bar{J}_3 and \bar{M}_3 on the second object. The final set of surface integral equations on the ground (S_1), the first object (S_2) and the second object (S_3) are given by [1], [3],[9]:

$$\bar{E}^{inc}(\bar{r})\Big|_{\text{tang.}} = \left[(L_1 + L_2)\bar{J}_1 - (K_1 + K_2)\bar{M}_1 - L_3\bar{J}_2 + K_3\bar{M}_2 - L_4\bar{J}_3 + K_4\bar{M}_3 \right]_{\text{tang.}}, \bar{r} \in S_1 \quad (1a)$$

$$\bar{H}^{inc}(\bar{r})|_{\text{tang.}} = \left[(K_1 + K_2)\bar{J}_1 + \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{M}_1 - K_3\bar{J}_2 - \frac{L_3}{\eta_2^2} \bar{M}_2 - K_4\bar{J}_3 - \frac{L_4}{\eta_2^2} \bar{M}_3 \right]_{\text{tang.}}, \bar{r} \in S_1 \quad (1b)$$

$$0 = [-L_2\bar{J}_1 + K_2\bar{M}_1 + (L_3 + L_5)\bar{J}_2 - (K_3 + K_5)\bar{M}_2 + L_4\bar{J}_3 - K_4\bar{M}_3]_{\text{tang.}}, \bar{r} \in S_2 \quad (1c)$$

$$0 = \left[-K_2\bar{J}_1 - \frac{L_2}{\eta_2^2} \bar{M}_1 + (K_3 + K_5)\bar{J}_2 + \left(\frac{L_3}{\eta_2^2} + \frac{L_5}{\eta_3^2} \right) \bar{M}_2 + K_4\bar{J}_3 + \frac{L_4}{\eta_2^2} \bar{M}_3 \right]_{\text{tang.}}, \bar{r} \in S_2 \quad (1d)$$

$$0 = [-L_2\bar{J}_1 + K_2\bar{M}_1 + L_3\bar{J}_2 - K_3\bar{M}_2 + (L_4 + L_6)\bar{J}_3 - (K_4 + K_6)\bar{M}_3]_{\text{tang.}}, \bar{r} \in S_3 \quad (1e)$$

$$0 = \left[-K_2\bar{J}_1 + \frac{L_2}{\eta_2^2} \bar{M}_1 + K_3\bar{J}_2 + \frac{L_3}{\eta_2^2} \bar{M}_2 + (K_4 + K_6)\bar{J}_3 + \left(\frac{L_4}{\eta_2^2} + \frac{L_6}{\eta_4^2} \right) \bar{M}_3 \right]_{\text{tang.}}, \bar{r} \in S_3 \quad (1f)$$

in which the intrinsic impedance is $\eta_i = \sqrt{\mu_i / \epsilon_i}$, with $i = 1, 2, \dots, 4$ and L_j and K_j , $j = 1, 2, \dots, 6$ are the integro-differential operators as summarized in Appendix A [1], [3], [9]. The surfaces of the rough ground and the two objects are discretized into triangular patches where the unknown equivalent electric and magnetic currents in (1) are approximated using the Rao, Wilton and Glisson (RWG) vector basis functions $\bar{j}(\bar{r})$ [9], [10] as:

$$\bar{J}_k(\bar{r}) = \sum_{n=1}^{N_k} I_n^{(k)} \bar{j}_n^{(k)}(\bar{r}), \quad \bar{M}_k(\bar{r}) = \eta_1 \sum_{n=1}^{N_k} I_{(n+N_k)}^{(k)} \bar{j}_n^{(k)}(\bar{r}), \quad \bar{r} \in S_k, \text{ for } k = 1, 2, 3 \quad (2)$$

After some algebraic manipulations, the linear system of equations is obtained as $\bar{\bar{Z}} \bar{I} = \bar{V}$ (as summarized in Appendix B) where the total impedance matrix $\bar{\bar{Z}}$ has order $2(N_1 + N_2 + N_3) \times 2(N_1 + N_2 + N_3)$. The number of surface unknowns on the ground, the first object and the second object are $2N_1$, $2N_2$ and $2N_3$, respectively. The vector \bar{V} represents the tested tangential incident electric field \bar{E}^{inc} and normalized magnetic field $\eta_1 \bar{H}^{inc}$ on the exterior surface of the ground. The SDFMM [11] was implemented in [1]-[3] to dramatically

accelerate solving for the unknown surface current coefficients \bar{I} . Once the unknown surface currents are obtained, the scattered electric fields in the far-zone can be computed [12].

For the modified Mueller matrix elements [5]-[8], we are following the notations used in [7] where the normalized modified Stokes vector is given by

$$I = \begin{bmatrix} |E_v|^2 \\ |E_h|^2 \\ 2\text{Re}(E_v E_h^*) \\ 2\text{Im}(E_v E_h^*) \end{bmatrix} / \eta_1 \quad (3a)$$

in which v and h represent the vertical and horizontal polarizations, respectively, thus from [7]

$$I^s = \frac{1}{r^2} M_m I^i \quad (3b)$$

where, the superscripts s and i represent the scattered and incident waves, respectively. The modified Mueller matrix M_m is given by [7]

$$M_m = \begin{bmatrix} |S_{vv}|^2 & |S_{vh}|^2 & \text{Re}(S_{vh}^* S_{vv}) & -\text{Im}(S_{vh}^* S_{vv}) \\ |S_{hv}|^2 & |S_{hh}|^2 & \text{Re}(S_{hv}^* S_{hh}) & -\text{Im}(S_{hv}^* S_{hh}) \\ 2\text{Re}(S_{vv} S_{hv}^*) & 2\text{Re}(S_{vh} S_{hh}^*) & \text{Re}(S_{vv} S_{hh}^* + S_{vh} S_{hv}^*) & -\text{Im}(S_{vv} S_{hh}^* - S_{vh} S_{hv}^*) \\ 2\text{Im}(S_{vv} S_{hv}^*) & 2\text{Im}(S_{vh} S_{hh}^*) & \text{Im}(S_{vv} S_{hh}^* + S_{vh} S_{hv}^*) & \text{Re}(S_{vv} S_{hh}^* - S_{vh} S_{hv}^*) \end{bmatrix} \quad (3c)$$

where S_{pq} is the p -polarized scattered waves due to the q -polarized incident waves ($p, q = v, h$) as described in the fully polarimetric scattering matrix S [7]

$$\begin{bmatrix} E_v^s \\ E_h^s \end{bmatrix} = \frac{e^{jkr}}{r} \begin{bmatrix} S_{vv} & S_{vh} \\ S_{hv} & S_{hh} \end{bmatrix} \begin{bmatrix} E_v^i \\ E_h^i \end{bmatrix} \quad (3d)$$

where $E_{v,h}^i$ and $E_{v,h}^s$ are the vertically and horizontally polarized incident and scattered electric fields in the far-zone, respectively.

III. NUMERICAL RESULTS AND DISCUSSIONS

The random rough surface characterized with Gaussian statistics for simulating the ground is generated 100 times for the configuration shown in Fig. 1. The linear system of equations is solved for each random surface realization for both the vertically- and the horizontally-polarized incident waves. Each set of the Monte Carlo simulations to obtain the average scattered intensity is based on the 100 independent realizations. The sixteen modified Mueller matrix elements in (3c) are calculated for the total scattered waves (i.e. not for the incoherent scattered waves). Using the standard matrix notations in (3c), these elements are $m_{11} = \langle |S_{vv}|^2 \rangle$, $m_{22} = \langle |S_{hh}|^2 \rangle$, $m_{14} = -\text{Im}(\langle S_{vh}^* S_{vv} \rangle)$, etc., where the angular brackets denote the ensemble average. All Mueller matrix elements presented here are normalized by the factor $4\pi/(2\eta_o A_z P^i)$, where A_z , η_o , and P^i are the footprint area on the ground, the intrinsic impedance of the free space and the total incident power. With this normalization, the element m_{11} becomes the normalized vertically co-polarized RCS (vv), the element m_{22} becomes the normalized horizontally co-polarized RCS (hh), the element m_{12} becomes the normalized cross-polarized RCS (vh), and the element m_{21} is the normalized cross-polarized RCS (hv). For the incident tapered Gaussian beam used in this work [5],[13] the total incident power is the sum of the power in each plane wave [14]. Similar to the previous work [1]-[3], the dimensions of the ground are assumed to be $8\lambda_0 \times 8\lambda_0$ with Gaussian half-beam width equal to $1.6\lambda_0$ centered on the ground at $4\lambda_0 \times 4\lambda_0$, where λ_0 is the free space wavelength.

In this work, the first object is modeled as an oblate spheroid ($a = 0.15\lambda_0, b = 0.3\lambda_0$) and the second object is modeled as a circular cylinder ($a = 0.15\lambda_0, h = 0.9\lambda_0$) or a circular disk ($a = 0.3\lambda_0, h = 0.1\lambda_0$) as shown in Figs. 1c-e. The locations of these objects vary in each

example. The rms height and the correlation length of the rough ground are assumed to be $\sigma = 0.1\lambda_0$ and $l_c = 0.5\lambda_0$, respectively. The relative dielectric constants of the ground, the cylinder, the spheroid, and the disk are assumed as $\epsilon_r = 2.5 - j0.18$, $\epsilon_r = 4.0$, $\epsilon_r = 2.9 - j0.072$ and $\epsilon_r = 4.0$, respectively. The total number of surface current unknowns is 61200 for the geometry shown in Fig. 1, where $2N_1 = 60000$ (on the ground surface), $2N_2 = 600$ (on the first object) and $2N_3 = 600$ (on the second object) [3]. These numbers of unknowns are the result of discretizing the ground surface into 10101 nodes and 20000 triangular patches, and discretizing each object into 102 nodes and 200 triangular patches and they are kept the same for all the cases discussed in this Section.

In Example 1, a spheroid and a horizontal cylinder are buried under the ground. The spheroid's center is located at $x = 4.5\lambda_0$, $y = 3.5\lambda_0$ and $z = -0.4\lambda_0$ while the horizontal cylinder is tilted by 30° with the x -axis and is located at $x = 4.01\lambda_0$, $y = 4.375\lambda_0$ and $z = -0.4\lambda_0$ measured from the axis mid-point (see Fig. 1c). For $\theta^i = 0^\circ$, $\phi^i = 0^\circ$, the ensemble average for each modified Mueller matrix element is plotted versus the scatter angle as presented in Fig. 2. As expected, the results show that all the diagonal elements of the Mueller matrix have maximum values approximately in the specular direction ($\theta^i = 0^\circ$). Interestingly, the elements m_{14} , m_{24} , m_{41} and m_{42} show distinguished signature for the rough ground with the two buried objects different from the rough ground alone (i.e. with no buried objects). Moreover, the elements m_{12} , m_{21} , m_{32} , m_{13} , m_{34} and m_{43} show slight differences between these two cases. However, the diagonal elements m_{11} , m_{22} , m_{33} , and m_{44} show almost no differences between the two cases.

It is important to investigate whether the significant differences shown in Fig. 2 are due to the interference between the objects or due to other causes. Therefore, additional results for the cylinder and the spheroid, each is buried alone under the same rough ground, are presented in Figs. 3 and 4, respectively. Each object is buried in its original location under the random rough ground as discussed in the data of Fig. 2. The results in Fig. 3 show the same observation of Fig. 2, i.e., the elements m_{14} , m_{24} , m_{41} , and m_{42} clearly show an indication to the presence of the cylinder under the ground while the elements m_{11} and m_{22} show no such indication.

The results in Fig. 4, for the buried spheroid alone, show that only elements m_{42} and m_{14} indicate to the presence of the spheroid. In other words, the results in Figs. 3 and 4 clearly show the significant sensitivity of some of the Mueller matrix elements to the presence of buried objects. The sensitivity level of these elements clearly depends on the object's orientation, material and size as presented in Figs. 2-4. It is interesting to notice that the expressions of m_{14} , m_{24} , m_{41} , and m_{42} as given in (3c) are the imaginary parts of the product of the co- and cross-polarized scattering elements. This implies that these elements contain some phase information, which is not the case for the other elements, e.g., m_{11} and m_{22} . The results in Figs. 2-4 clearly indicate that if one relies only on the co- and/or the cross-polarized intensities (i.e., m_{11} , m_{22} , m_{12} and m_{21}); it might not be possible to detect the buried objects. However, investigating all the sixteen Mueller matrix elements significantly increases the possibility of detecting these objects.

In Example 2, the cylinder's orientation effect is investigated with keeping the rest of the data of Fig. 2. Therefore, the cylinder is tilted with the x -axis by 90° and is buried alone under the rough interface. The axis's mid-point is located at $x = 3.6\lambda_0$, $y = 4.0\lambda_0$ and $z = -0.4\lambda_0$ (see Fig. 1d). As shown in Fig. 5, almost twelve elements out of the sixteen ones, show considerable differences due to the presence of the cylinder. Comparing the results of Fig. 3 and Fig. 5, where

the same size and material of the horizontal cylinder are used, shows that the cylinder's orientation plays an effective role in the sensitivity level of these elements. The depolarization level of the scattered waves from the rough ground with the buried cylinder vary with the orientation of the cylinder leading to observed variation in some of the Mueller matrix elements.

In Example 3, the cylinder is replaced with a circular disk tilted with the z -axis by 20° and of the same material (i.e., $\epsilon_r = 4.0$). The center of its top circle is located $x = 3.3\lambda_0$, $y = 4.9\lambda_0$, $z = -0.35\lambda_0$ and is buried alone under the rough ground (see Fig. 1e). Again, the results plotted in Fig. 6 show that some of the Mueller matrix elements show the indication to the presence of the disk. In particular, the elements m_{14} , m_{24} , m_{41} , m_{42} , m_{12} and m_{21} show considerable differences between the rough ground only and the rough ground with the buried disk. Consistent with the previous examples, the vv or hh intensities (i.e. m_{11} and m_{22}) show almost no differences between the two cases.

In Example 4, the Mueller elements for the spheroid and the horizontal cylinder, both buried under the ground (same data of Fig. 2), are computed at different incident angles $\theta^i = 30^\circ$, $\phi^i = 120^\circ$ as shown in Fig. 7. None of the elements in the current case show indication to the presence of the buried objects except for the elements m_{34} and m_{43} . In addition, all the elements in this case are significantly different from those presented in Fig. 2, which shows their dependency on the incident angles. In this Example, the far-fields are computed in the same plane of incidence, i.e., at $0^\circ \leq \theta^s \leq 90^\circ$ for $\phi^s = 120^\circ$ and for $\phi^s = 300^\circ$.

As mentioned earlier, in Examples 1-4, all the Mueller matrix elements represent the average values based on the Monte Carlo simulations of 100 rough surface realizations. This indicates that the observed sensitivity of some of these elements is consistent for all rough surface realizations since it survived the averaging process. To investigate this point, as an example, the

elements m_{42} and m_{14} are plotted for the scattered waves from only one rough surface realization as shown in Fig. 8. The numerical results show that these elements for the rough ground alone are different from all the other buried scatterers. This observation indicates that investigating all the Mueller matrix elements can be very useful in detecting the buried objects under the realistic rough ground represented here by one rough surface realization.

The above results support the observation reported in [5] in which the underground inhomogeneity was caused by the presence of the coated irregular layer above the rough ground while here the inhomogeneity is due to the presence of dielectric objects buried beneath the ground. As mentioned in [7], the modified Mueller matrix, in general, is not symmetric which is demonstrated in all figures. However, in all results it is observed that $m_{34} \approx -m_{43}$ which can be attributed to the slight cross-polarization caused by the small roughness parameters considered in this work. For the same reason it is observed that $m_{33} \approx m_{44}$ in all figures. However, in Figs. 2-6, it is observed, as expected, that $m_{11} \approx m_{22}$ (normal incidence case) while in Fig. 7 $m_{22} > m_{11}$ (oblique incidence case). In the later case, the incident electric field for the h -polarization is parallel to the axis of the buried cylinder while for the v -polarization, it is perpendicular to cylinder axis, see Fig. 1-b.

Notice that the elements $m_{12} \neq m_{21}$ because all the Mueller matrix elements presented here are for the bistatic case and not for the backscatter. Even though, it is more practical for remote sensing applications to consider the backscatter case, however, the bistatic results showed the considerable sensitivity of some Mueller matrix elements to the presence of buried objects. In addition, calculating the average of these elements in the backscatter direction will be computationally more expensive.

Examining the Mueller matrix expression in (3c), it is clear that several elements are functions of the amount of wave depolarization. This could be one of the factors that the Mueller elements for the tilted cylinder are the most distinguished ones as shown in the numerical results. The materials, locations of the buried scatterers and their contrast with the surrounding underground medium have impact on these elements as well. More work need to be conducted to investigate how the ground roughness will affect these elements since in this work we assumed small roughness parameters.

IV. CONCLUSIONS

The fast 3-D SDFMM computer code is used to calculate the average of each modified Mueller matrix element for the scattering from dielectric objects buried under the random rough ground. The statistical average for each element is computed using the Monte Carlo simulation. These simulations are obtained by running the 3-D SDFMM computer code hundreds of times, which shows the great advantage of using the Fast Multipole Method (FMM). The numerical results clearly show that if one relies only on the co- and/or the cross-polarized intensities; it is very difficult to sense the buried objects. However, investigating all the sixteen Mueller matrix elements significantly help in detecting these objects.

ACKNOWLEDGMENTS

This research was sponsored in part by the Northeastern University's Demining MURI grant No. DAA 0-55-97-0013 and the Engineering Research Centers Program of the NSF award number EEC-9986821, and in part by the Arkansas Science and Technology Authority Grant No AR/ASTA/01-B-18. The SDFMM was originally developed by V. Jandhyala, E. Michielssen and W. Chew at the UIUC.

Appendix A

With representing the surface electric and magnetic currents \bar{J} and \bar{M} on S_1 , S_2 and S_3 by the vector \bar{X} , the integro-differential operators L_j and K_j , $j=1, 2, \dots, 6$, are [1], [3], [9]:

$$L_{1,2}\bar{X} = \int_{S_1} \left\{ i\omega\mu_{1,2}\Phi_{1,2}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{1,2}}\nabla\nabla' \cdot \bar{X}(\bar{r}')\Phi_{1,2} \right\} ds', \quad K_{1,2}\bar{X} = \int_{S_1} \bar{X}(\bar{r}') \times \nabla\Phi_{1,2} ds' \quad (\text{A1})$$

$$L_{3,5}\bar{X} = \int_{S_2} \left\{ i\omega\mu_{2,3}\Phi_{2,3}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{2,3}}\nabla\nabla' \cdot \bar{X}(\bar{r}')\Phi_{2,3} \right\} ds', \quad K_{3,5}\bar{X} = \int_{S_2} \bar{X}(\bar{r}') \times \nabla\Phi_{2,3} ds' \quad (\text{A2})$$

$$L_{4,6}\bar{X} = \int_{S_3} \left\{ i\omega\mu_{2,4}\Phi_{2,4}\bar{X}(\bar{r}') + \frac{i}{\omega\epsilon_{2,4}}\nabla\nabla' \cdot \bar{X}(\bar{r}')\Phi_{2,4} \right\} ds', \quad K_{4,6}\bar{X} = \int_{S_3} \bar{X}(\bar{r}') \times \nabla\Phi_{2,4} ds' \quad (\text{A3})$$

Appendix B

The linear system of equations, $\bar{\bar{Z}}\bar{I} = \bar{V}$, is given by [1], [3], [9]

$$\begin{pmatrix} \bar{Z}_{11} & \bar{Z}_{12} & \bar{Z}_{13} \\ \bar{Z}_{21} & \bar{Z}_{22} & \bar{Z}_{23} \\ \bar{Z}_{31} & \bar{Z}_{32} & \bar{Z}_{33} \end{pmatrix} \begin{pmatrix} I^1 \\ I^2 \\ I^3 \end{pmatrix} = \begin{pmatrix} \bar{V}_1 \\ 0 \\ 0 \end{pmatrix} \quad (\text{B1})$$

where the submatrix \bar{Z}_{11} is

$$\bar{Z}_{11} = \begin{pmatrix} \langle \bar{j}_1, (L_1 + L_2)\bar{j}_1 \rangle_{S_1} & \langle \bar{j}_1, -\eta_1(K_1 + K_2)\bar{j}_1 \rangle_{S_1} \\ \langle \bar{j}_1, \eta_1(K_1 + K_2)\bar{j}_1 \rangle_{S_1} & \left\langle \bar{j}_1, \eta_1^2 \left(\frac{L_1}{\eta_1^2} + \frac{L_2}{\eta_2^2} \right) \bar{j}_1 \right\rangle_{S_1} \end{pmatrix} \quad (\text{B2})$$

in which $\langle \bar{A}, \bar{B} \rangle_S$ denotes the complex inner product between vector functions \bar{A} and \bar{B} on a

surface S . The submatrices \bar{Z}_{12} , \bar{Z}_{13} , and \bar{Z}_{23} are given by

$$\bar{Z}_{12} = \begin{pmatrix} \langle \bar{j}_1, -L_3 \bar{j}_2 \rangle_{S_1} & \langle \bar{j}_1, \eta_1 K_3 \bar{j}_2 \rangle_{S_1} \\ \langle \bar{j}_1, -\eta_1 K_3 \bar{j}_2 \rangle_{S_1} & \left\langle \bar{j}_1, -\eta_1^2 \left(\frac{L_3}{\eta_2^2} \right) \bar{j}_2 \right\rangle_{S_1} \end{pmatrix}, \bar{Z}_{13} = \begin{pmatrix} \langle \bar{j}_1, -L_4 \bar{j}_3 \rangle_{S_1} & \langle \bar{j}_1, \eta_1 K_4 \bar{j}_3 \rangle_{S_1} \\ \langle \bar{j}_1, -\eta_1 K_4 \bar{j}_3 \rangle_{S_1} & \left\langle \bar{j}_1, -\eta_1^2 \left(\frac{L_4}{\eta_2^2} \right) \bar{j}_3 \right\rangle_{S_1} \end{pmatrix} \quad (B3)$$

$$\bar{Z}_{23} = \begin{pmatrix} \langle \bar{j}_2, L_4 \bar{j}_3 \rangle_{S_2} & \langle \bar{j}_2, -\eta_1 K_4 \bar{j}_3 \rangle_{S_2} \\ \langle \bar{j}_2, \eta_1 K_4 \bar{j}_3 \rangle_{S_2} & \left\langle \bar{j}_2, \eta_1^2 \frac{L_4}{\eta_2^2} \bar{j}_3 \right\rangle_{S_2} \end{pmatrix} \quad (B4)$$

Similar expressions can be obtained for all the other submatrices in B1.

References

- [1] M. El-Shenawee, C. Rappaport, E. Miller and M. Silevitch, "Three-dimensional subsurface analysis of electromagnetic scattering from penetrable/PEC objects buried under rough surfaces: use of the steepest descent fast multipole method (SDFMM)," *IEEE Trans. Geosci. & Rem. Sensing*, vol. 39, no. 6, pp. 1174-1182, June 2001.
- [2] M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo Simulations of Electromagnetic Wave Scattering from Random Rough Surface with 3-D Penetrable Buried Object: Mine Detection Application Using the SDFMM," *J. Optical Society of America A*, vol.18, no. 12, pp.3077-3084, December 2001.
- [3] M. El-Shenawee, "Scattering from Multiple Objects Buried Under Two-Dimensional Randomly Rough Surface Using the Steepest Descent Fast Multipole Method," *IEEE Trans. Anten. & Propag.*, accepted for publication, to appear.
- [4] G. Zhang, L. Tsang and K. Pak, "Angular correlation function and scattering coefficient of electromagnetic waves scattered by a buried object under a two-dimensional rough surface," *J. Opt. Soc. Am. A*, vol. 15, no. 12, pp. 2995-3002, December 1998.
- [5] Yuzhi Zhang and Ezekiel Bahar, "Mueller matrix elements that characterize scattering from coated random rough surfaces," *IEEE Trans. Anten. & Propag.*, vol. 47, no. 5, pp. 949-955, 1999.
- [6] A. Ishimaru, *Wave Propagation and Scattering in Random Media*, New York: Academic Press, 1978.
- [7] F. T. Ulaby and C. Elachi, *Radar Polarimetry for Geoscience Applications*, Artech House, Inc., 1990.
- [8] G. G. Stokes, "On the composition and resolution of streams of polarized light from different sources," *Trans. Cambridge Phil. Soc.*, vol. 9, pp. 399-416, 1852 (reprinted in *Mathematical and Physical papers*. London, U.K.: Cambridge Univ. press, 1901, vol. 3, pp. 233-250.
- [9] L. Medgyesi-Mitschang, J. Putnam, and M. Gedera, "Generalized method of moments for three-dimensional penetrable scatterers," *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1383-1398, April 1994.
- [10] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Anten. & Prop.*, vol. AP-30, no. 3, pp.409-418, May 1982.

- [11] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [12] C. A. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons Inc, 1989
- [13] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.
- [14] R. L. Wagner, J. Song and W. C. Chew, "Monte Carlo simulation of electromagnetic scattering from two-dimensional random rough surfaces," *IEEE Trans. Antennas Propag.*, vol. 45, no. 2, pp. 235-245, February 1997.

List of figures

Figure 1. (a) Cross section of a general 2-D rough ground with two buried objects, (b) top view of the geometry, (c) 3-D geometry for the buried spheroid and the 30°-tilted horizontal cylinder, (d) 3-D geometry for the buried 90°-tilted horizontal cylinder, and (e) 3-D geometry for the buried 20°-tilted disk. All figures in 1c-e show exact locations of the objects.

Figure 2. The normalized bistatic modified Mueller matrix elements (total intensity); $m_{11}, m_{12}, \dots, m_{44}$; the solid line is for the rough ground only and the cross-symbol is for the rough ground with the two buried objects (the spheroid and the horizontal 30°-tilted cylinder, Fig. 1c) for incident angles $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$, and Monte Carlo for 100 rough surface realizations.

Figure 3. The normalized bistatic modified Mueller matrix elements (total intensity); $m_{11}, m_{12}, \dots, m_{44}$; the solid line is for the rough ground only and the cross-symbol is for the rough ground with only the second object (horizontal 30°-tilted cylinder, Fig. 1c) for the incident angles $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$, and Monte Carlo for 100 rough surface realizations.

Figure 4. The normalized bistatic modified Mueller matrix elements (total intensity); $m_{11}, m_{12}, \dots, m_{44}$; the solid line is for the rough ground only and the cross-symbol is for the rough ground with only the first object (the spheroid, Fig. 1c) for the incident angles $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$, and Monte Carlo for 100 rough surface realizations.

Figure 5. The normalized bistatic modified Mueller matrix elements (total intensity); $m_{11}, m_{12}, \dots, m_{44}$; the solid line is for the rough ground only and the cross-symbol is for the rough ground with only the second object (horizontal 90°-tilted cylinder, Fig. 1d) for the incident angles $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$, and Monte Carlo for 100 rough surface realizations.

Figure 6. The normalized bistatic modified Mueller matrix elements (total intensity); $m_{11}, m_{12}, \dots, m_{44}$; the solid line is for the rough ground only and the cross-symbol is for the rough ground with only the second object (20°-tilted disk, Fig. 1e) for the incident angles $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$, and Monte Carlo for 100 rough surface realizations.

Figure 7. The normalized bistatic modified Mueller matrix elements (total intensity); $m_{11}, m_{12}, \dots, m_{44}$; the solid line is for the rough ground only and the cross-symbol is for the rough ground with the two buried objects (the spheroid and the horizontal 30°-tilted cylinder, Fig. 1c) for the incident angles $\theta^i = 30^\circ$ and $\phi^i = 120^\circ$, and Monte Carlo for 100 rough surface realizations.

Figure 8. The normalized bistatic modified Mueller matrix (a) element m_{42} and (b) element m_{14} versus the scatter angle for $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$ from only one rough surface realization (No averaging).

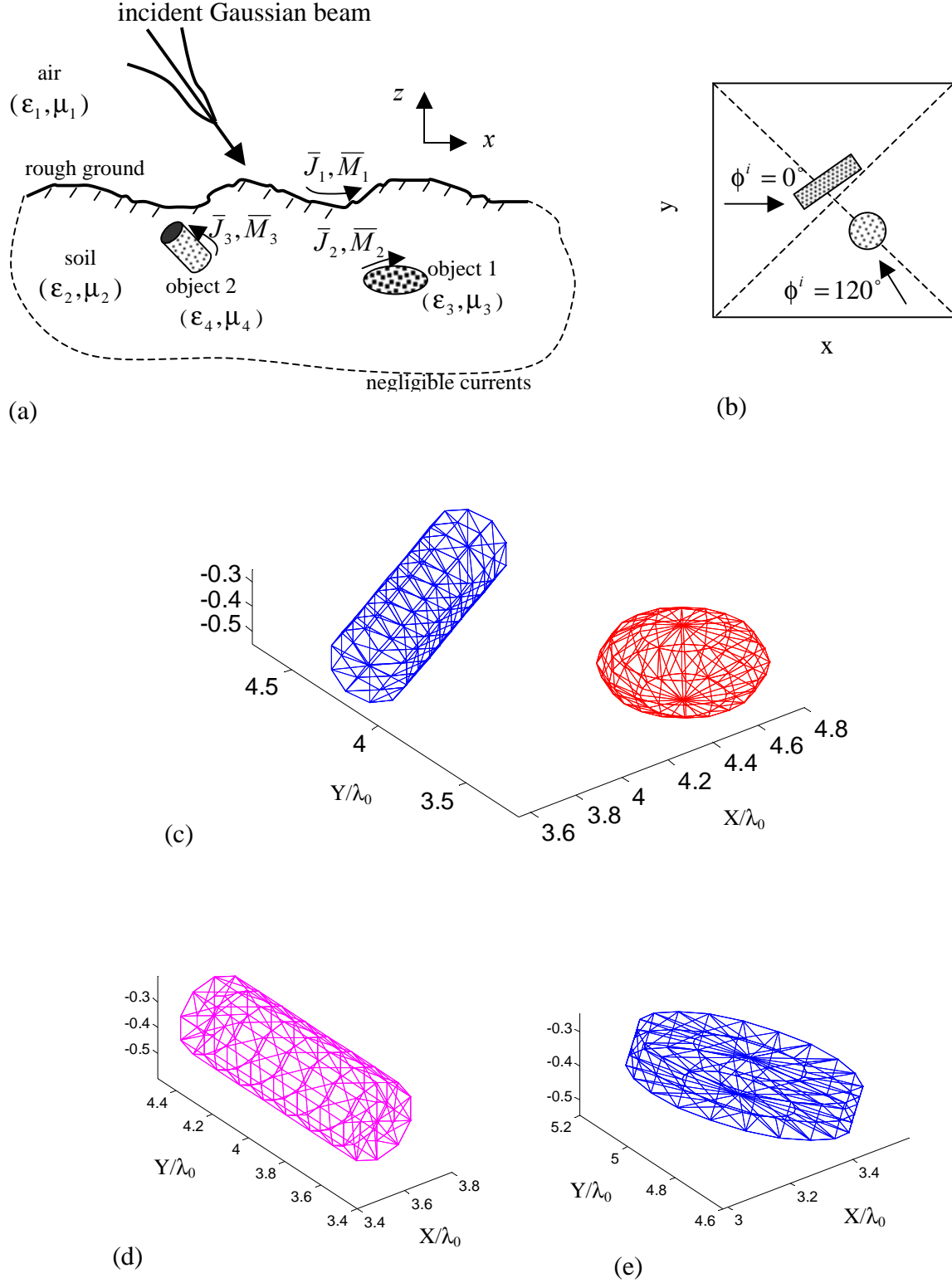


Figure 1. (a) Cross section of a general 2-D rough ground with two buried objects, (b) top view of the geometry, (c) 3-D geometry for the buried spheroid and the 30°-tilted horizontal cylinder, (d) 3-D geometry for the buried 90°-tilted horizontal cylinder, and (e) 3-D geometry for the buried 20°-tilted disk. All figures in 1c-e show exact locations of the objects.

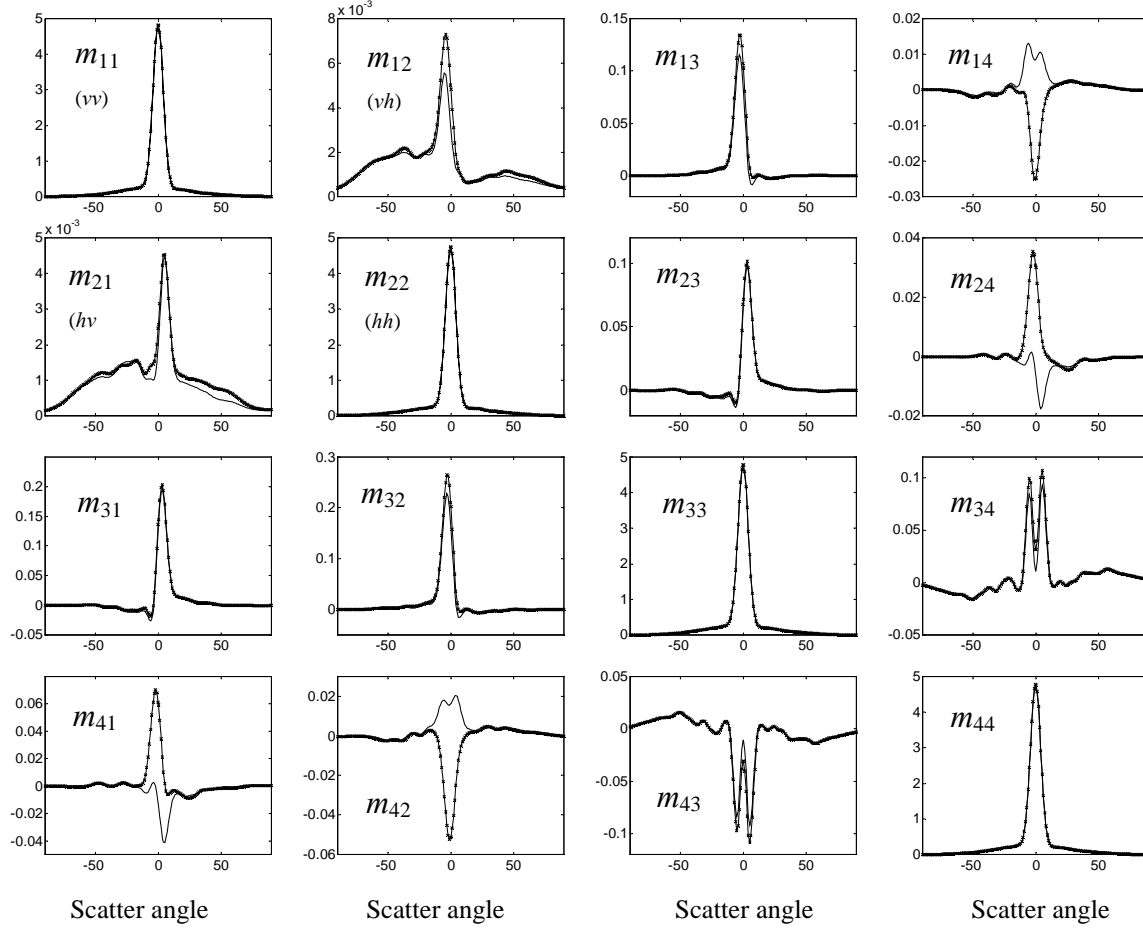


Figure 2. The normalized bistatic modified Mueller matrix elements (total intensity); $m_{11}, m_{12}, \dots, m_{44}$; the solid line is for the rough ground only and the cross-symbol is for the rough ground with the two buried objects (the spheroid and the horizontal 30° -tilted cylinder, Fig. 1c) for incident angles $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$, and Monte Carlo for 100 rough surface realizations.

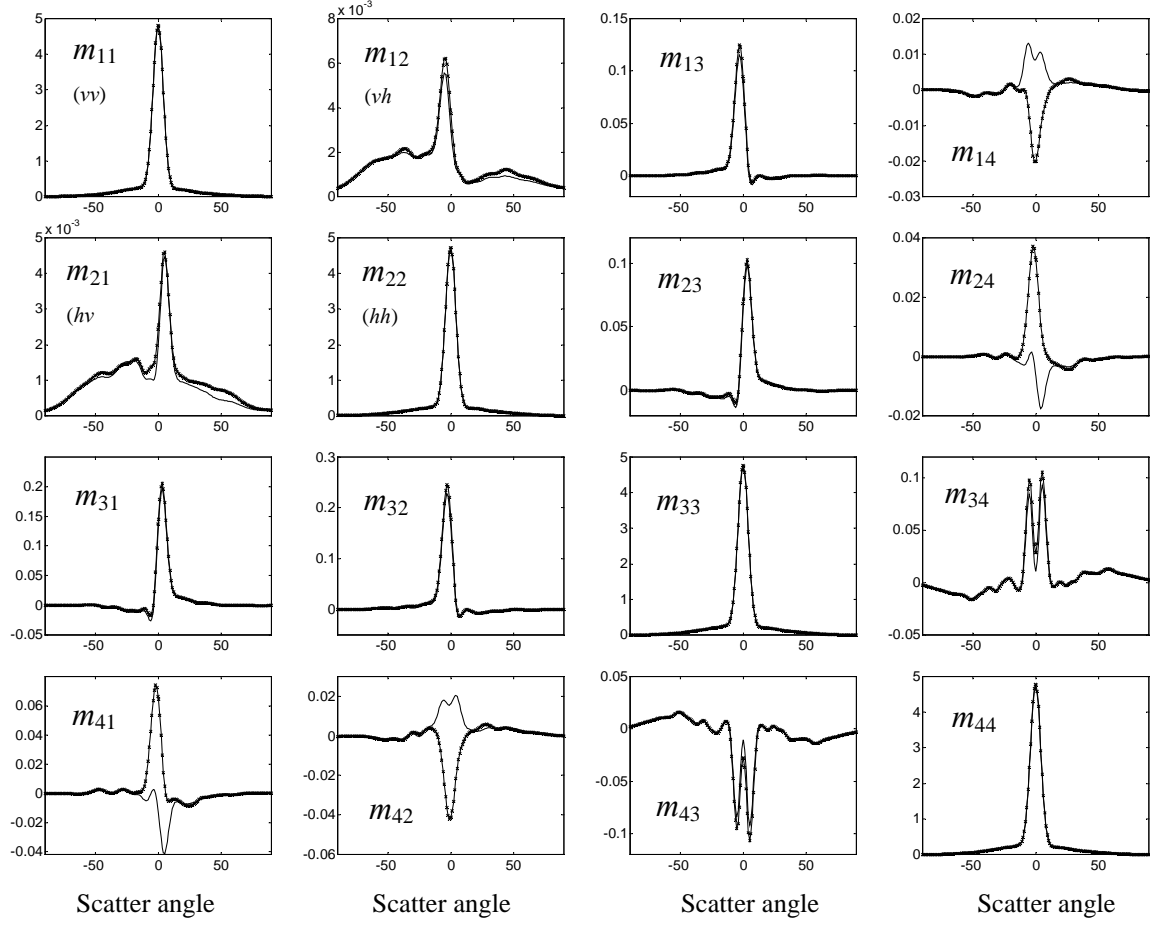


Figure 3. The normalized bistatic modified Mueller matrix elements (total intensity); m_{11} , m_{12} , ..., m_{44} ; the solid line is for the rough ground only and the cross-symbol is for the rough ground with only the second object (horizontal 30°-tilted cylinder, Fig. 1c) for the incident angles $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$, and Monte Carlo for 100 rough surface realizations.

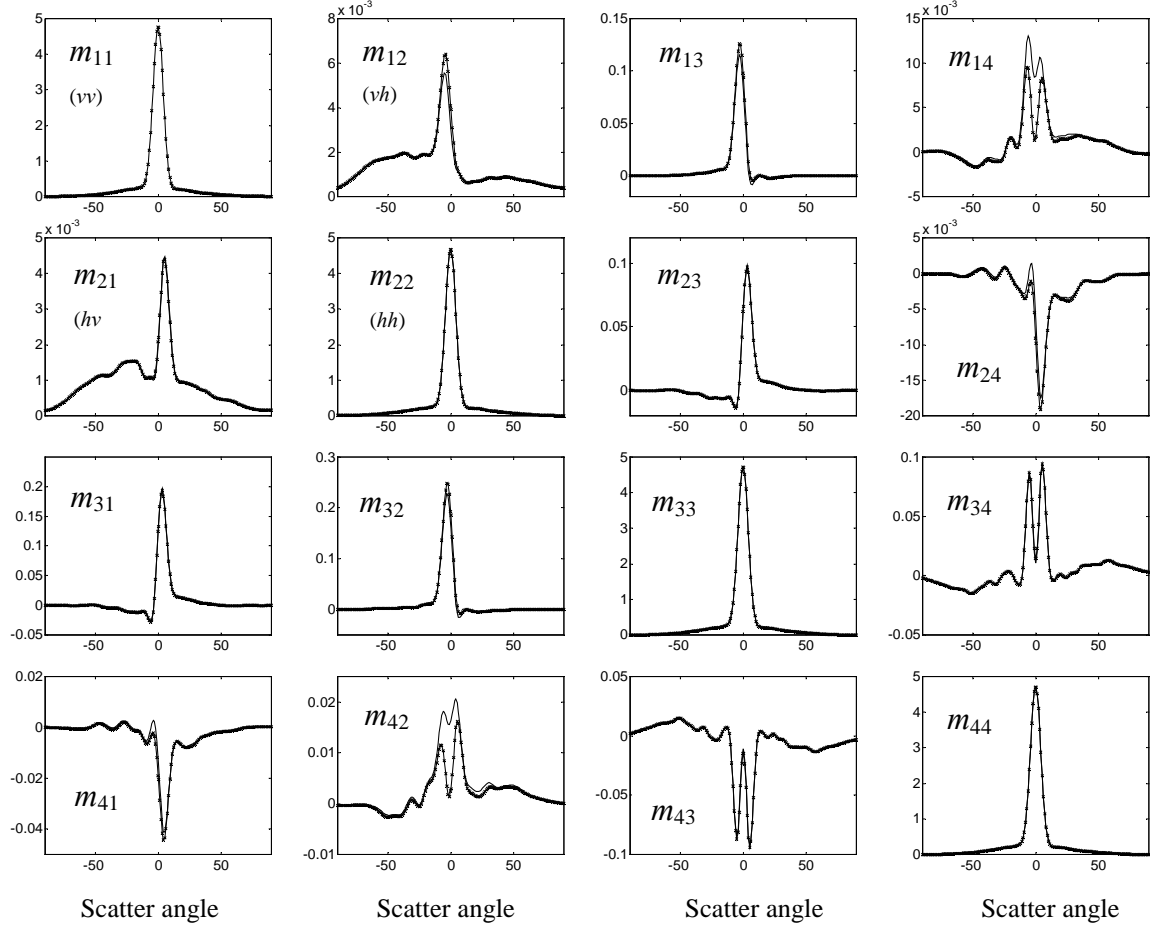


Figure 4. The normalized bistatic modified Mueller matrix elements (total intensity); m_{11} , m_{12} , ..., m_{44} ; the solid line is for the rough ground only and the cross-symbol is for the rough ground with only the first object (the spheroid, Fig. 1c) for the incident angles $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$, and Monte Carlo for 100 rough surface realizations.

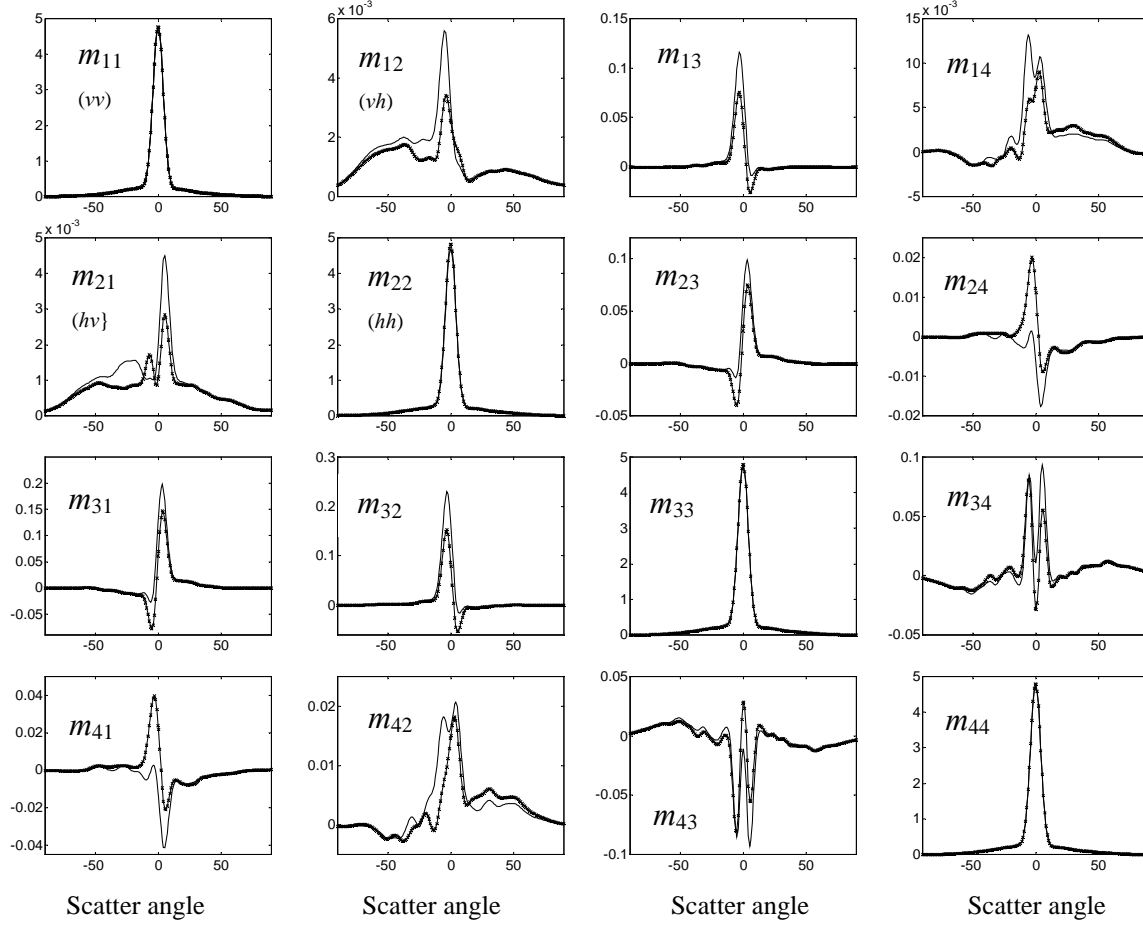


Figure 5. The normalized bistatic modified Mueller matrix elements (total intensity); m_{11} , m_{12} , ..., m_{44} ; the solid line is for the rough ground only and the cross-symbol is for the rough ground with only the second object (horizontal 90° -tilted cylinder, Fig. 1d) for the incident angles $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$, and Monte Carlo for 100 rough surface realizations.

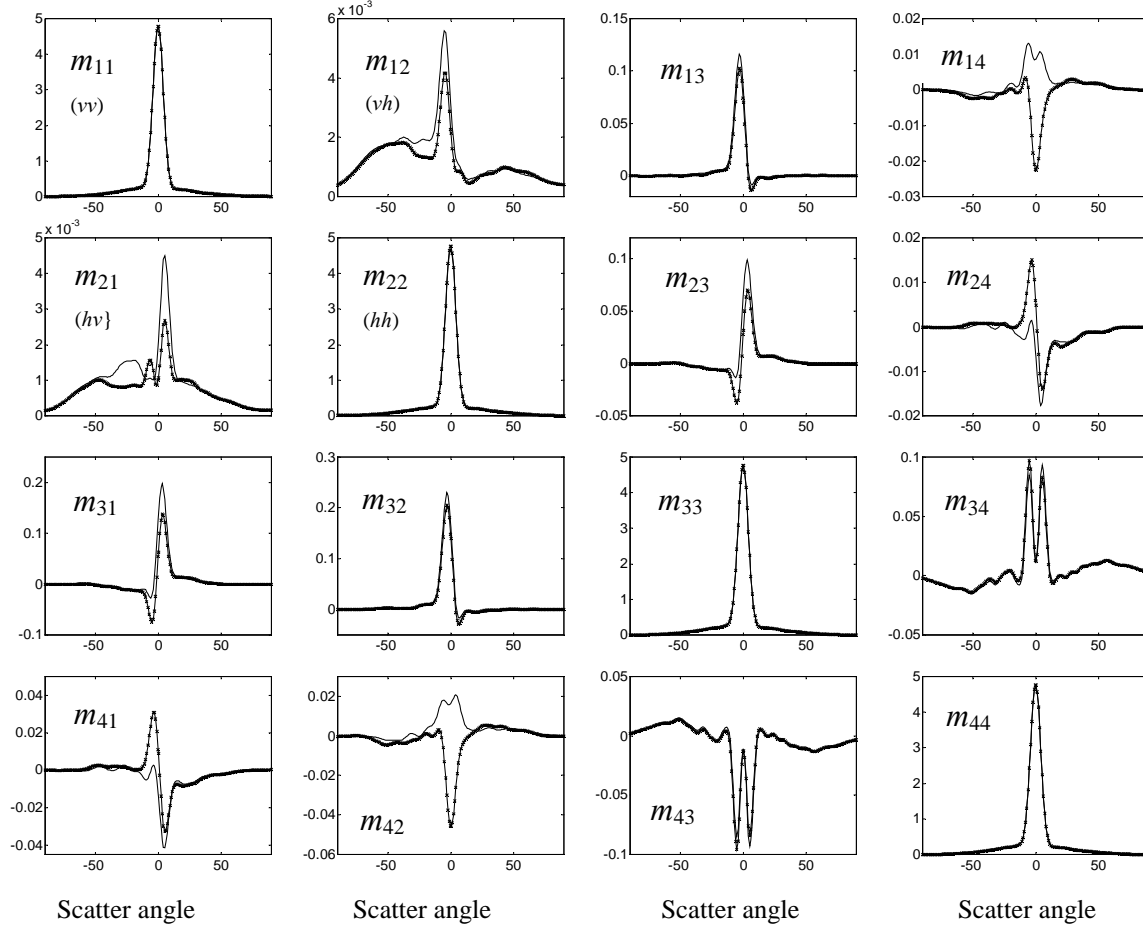


Figure 6. The normalized bistatic modified Mueller matrix elements (total intensity); m_{11} , m_{12} , ..., m_{44} ; the solid line is for the rough ground only and the cross-symbol is for the rough ground with only the second object (20° -tilted disk, Fig. 1e) for the incident angles $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$, and Monte Carlo for 100 rough surface realizations.

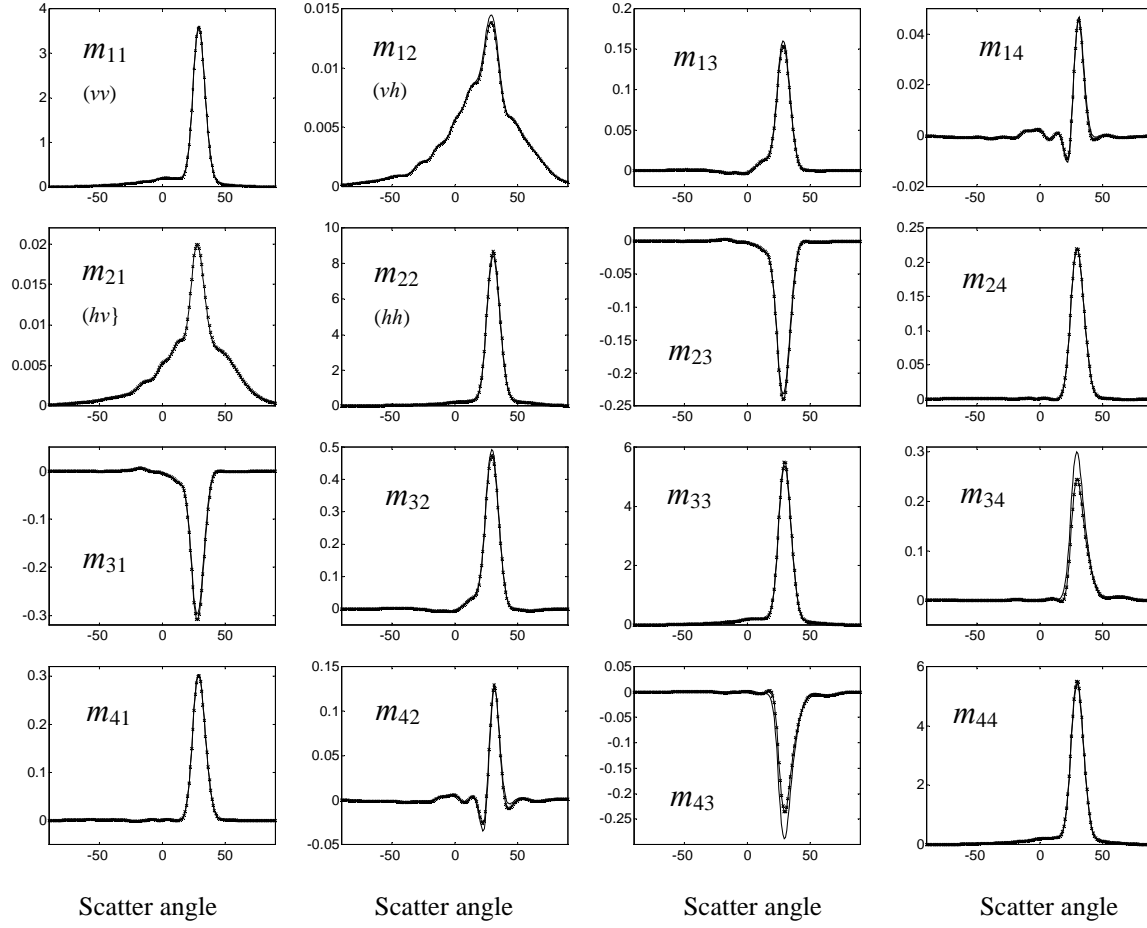


Figure 7. The normalized bistatic modified Mueller matrix elements (total intensity); m_{11} , m_{12} , ..., m_{44} ; the solid line is for the rough ground only and the cross-symbol is for the rough ground with the two buried objects (the spheroid and the horizontal 30°-tilted cylinder, Fig. 1c) for the incident angles $\theta^i = 30^\circ$ and $\phi^i = 120^\circ$, and Monte Carlo for 100 rough surface realizations.

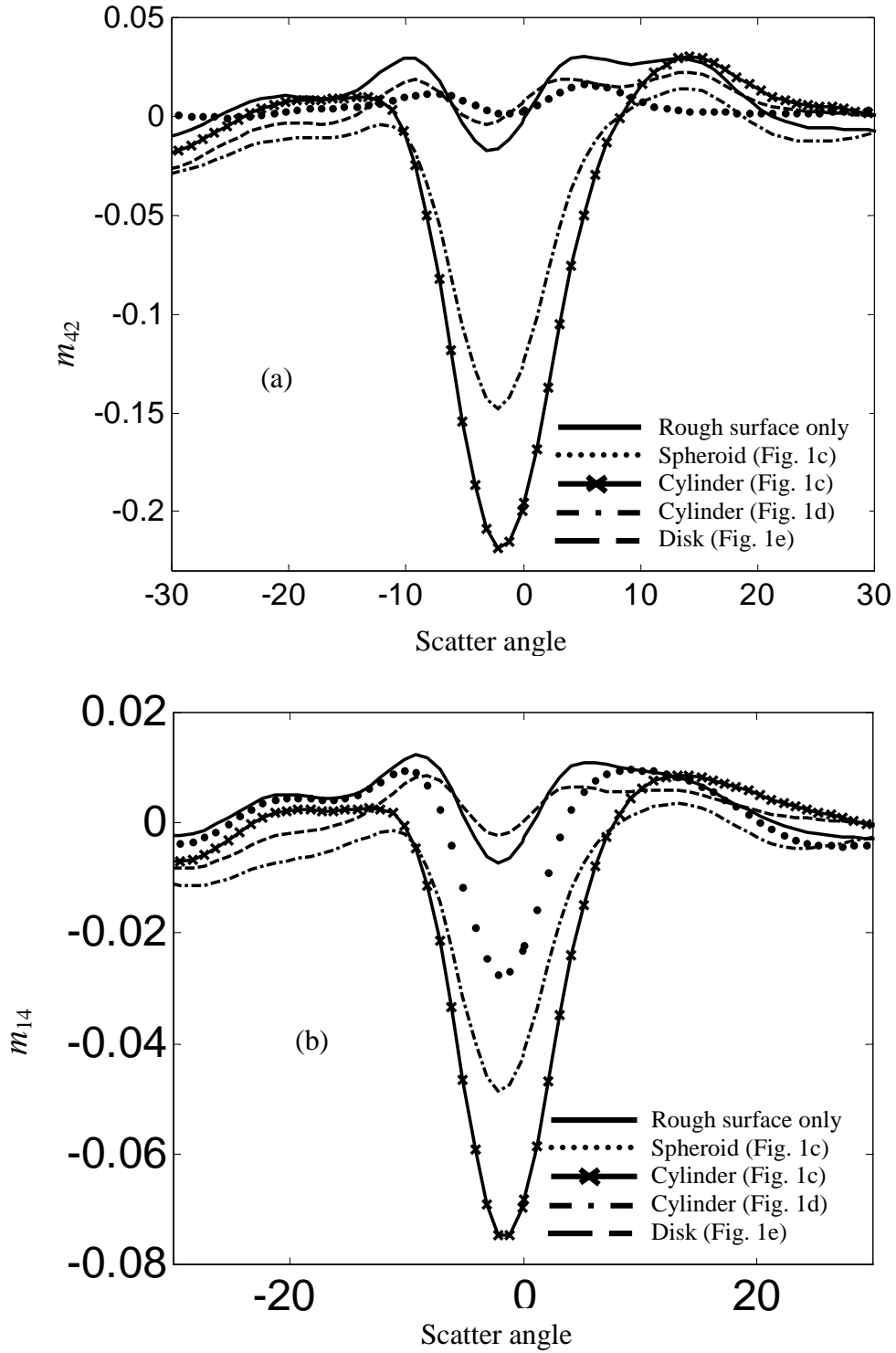


Figure 8. The normalized bistatic modified Mueller matrix (a) element m_{42} and (b) element m_{14} versus the scatter angle for $\theta^i = 0^\circ$ and $\phi^i = 0^\circ$ from only one rough surface realization (No averaging).

Electromagnetic Scattering Interference Between Two Shallow Objects Buried Under 2-D Random Rough Surfaces

Magda El-Shenawee¹ and Carey Rappaport²

¹Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701
magda@uark.edu

²Department of Electrical and Computer Engineering
Northeastern University
Boston, MA
rappaport@neu.edu

Abstract

A rigorous electromagnetic model has been developed to analyze the scattering from two dielectric shallow objects buried under the 2-D random rough ground as a means of predicting false alarms due to a clutter object near a target object. The Method of Moments (MoM) accelerated by the Steepest Descent Fast Multipole Method (SDFMM) is used to compute the unknown electric and magnetic surface currents on both the rough ground surface and the two buried objects. The roughness parameters strongly influence the scattered interference of the two objects, especially when their separation distance is several correlation lengths. This could increase the probability of false alarms.

I. INTRODUCTION

In realistic minefields, buried anti-personnel (AP) nonmetallic mines are often closely accompanied by underground clutter-objects. The presence of this object considerably obscures the targets causing a false alarm during the detection process. The separation distance between the AP-mine and the clutter-object plays a primary role on the probability of false alarms. A rigorous electromagnetic model has been developed to analyze the scattering mechanism of two dielectric objects buried beneath a rough ground surface as reported in [1]. Using the $O(N)$ fast algorithm, the Steepest Descent Fast Multipole Method (SDFMM) [2-3] tremendously accelerates the computations of the N unknown surface currents [1,4]. When the two objects were located close to one another under flat ground, the strong scattering interference generates a false response appearing to be a third buried object [1]. The dependency of the observed scattering interference on the ground roughness parameters is investigated in this work.

II. FORMULATION

The rigorous electromagnetic model derived in [1] is employed in this work where six integral equations are used to obtain the equivalent surface currents on dielectric scatterers shown in Fig. 1 (3-D scattering problem). Four different regions are involved in this scattering problem; air, soil, first object and second object. The unknown electric and magnetic currents on the ground surface, on the target surface, and on the clutter object surface are approximated using the well-known RWG vector basis functions [5]. After some algebraic manipulations, the linear system of equations is obtained as: $\bar{\bar{Z}} \bar{I} = \bar{V}$, where the total impedance matrix $\bar{\bar{Z}}$ has order of $2(N_1 + N_2 + N_3) \times 2(N_1 + N_2 + N_3)$. The number of electric and magnetic current unknowns (edges) on the ground, on the target and on the second object are $2N_1$, $2N_2$ and $2N_3$, respectively. The tested tangential incident electric field \bar{E}^{inc} and the tested normalized magnetic field $\eta_1 \bar{H}^{inc}$ on the exterior of the ground surface are expressed in \bar{V} . The SDFMM is implemented to significantly accelerate solving the linear system of equations for the unknown current coefficients [1-4].

II. NUMERICAL RESULTS

In this Section, we investigate the scattering interference between the two buried objects as a function of their separation distances and the ground roughness. Several values for the root mean square height σ and the correlation length l_c are considered with emphasis on small roughness parameters for the AP-mine detection application. The incident wave is assumed to be a Gaussian beam at normal incidence that is carefully tapered to minimize edge effects [6]. The two objects are oblate spheroids with dimensions $a = 0.3\lambda_0$, $b = 0.15\lambda_0$ and at depth $z = -0.4\lambda_0$ measured from the center. The relative dielectric constant of the ground soil is assumed to be $\epsilon_r = 2.5 - j0.18$ and for both objects is assumed to be $\epsilon_r = 2.9 - j0.072$. The $8\lambda_0 \times 8\lambda_0$ ground surface is discretized into 60,000 electric and magnetic surface current unknowns. Each object is discretized into 600 electric and magnetic surface current unknowns. In order to analyze the object signatures, the electric fields scattered from each rough ground are removed by subtraction similar to our work in [4].

In Fig. 2, the object signatures are plotted across the diagonal as shown in Fig.1. When the two objects are separated by $S = 1.4\lambda_0$, the results show three peaks of almost equal magnitudes; the first peak is above the first object, the second peak is above the second object and the third peak is at mid-point between them. This third peak is due to the strong constructive interference between the two objects. This phenomenon could easily cause a false alarm during the detection process. However, when the separation distance increases, the mid-point peak is dissolved into several secondary peaks as observed when $S = 2.1\lambda_0$ and $2.8\lambda_0$. These secondary peaks become insignificant when S is increased to $4.2\lambda_0$. As noticed in this figure, the asymmetry around the mid-point ($x = 4.0\lambda_0$) is clearly caused by the random roughness of the ground. Moreover, the magnitudes of all peaks decrease with increasing separation distance because the objects are further from the beam footprint center (ground center in this work). In Figs. 3 and 4, the separation distance is $S = 1.4\lambda_0$, the range for σ is $0.04\lambda_0$ - $0.1\lambda_0$ and the range for l_c is $0.4\lambda_0$ - $1.0\lambda_0$. The results clearly show the influence of the roughness parameters, however, the observed strong interference (mid-point peak) is slightly affected by the ground roughness in this case. This study is repeated for $S = 2.1\lambda_0$ as shown in Figs. 5 and 6, and for $S = 2.8\lambda_0$ as shown in Figs. 7 and 8. It is interesting to notice that when the separation distance becomes several correlation lengths, as shown in Figs. 6 and 8, stronger scattered interference is observed. This is demonstrated by the increase in the number of secondary peaks from three in the flat ground case to four and five in the rough ground case as shown in Figs. 6 and 8, respectively. This mechanism could increase the possibility of false alarms.

III. CONCLUSIONS

False alarms could easily occur because of the interference mechanism between the target and a second nearby object buried under the ground. The roughness parameters strongly influence the scattering interference mechanism between the two objects, which could increase the probability of false alarms.

ACKNOWLEDGMENTS

This research was sponsored by the Northeastern University's Demining MURI grant # DAA 0-55-97-0013 and in part by the College of Engineering at the University of Arkansas. The SDFMM was originally developed at the UIUC by V. Jandhyala, E. Michielssen and W. Chew.

References

- [1] M. El-Shenawee, "The Steepest Descent Fast Multipole Method for Scattering from Two Penetrable Shallow Objects Buried Under Two-Dimensional Randomly Rough Surface," submitted to the *IEEE Trans. Anten. & Propag.* under review.
- [2] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [3] V. Jandhyala, E. Michielssen, B. Shanker and W.C. Chew, "A combined steepest descent-fast multipole algorithm for the fast analysis of three-dimensional scattering by rough surfaces," *IEEE Trans. Geosci. & Remote Sensing*, vol. 36, no. 3, pp. 738-748, May 1998.

- [4] M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo simulations of electromagnetic wave scattering from random rough surface with 3-D penetrable buried object: mine detection application using the SDFMM," *J. Optical Society of America A*, to appear in December 2001
- [5] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Anten. & Prop.*, vol. AP-30, no. 3, pp.409-418, May 1982.
- [6] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.

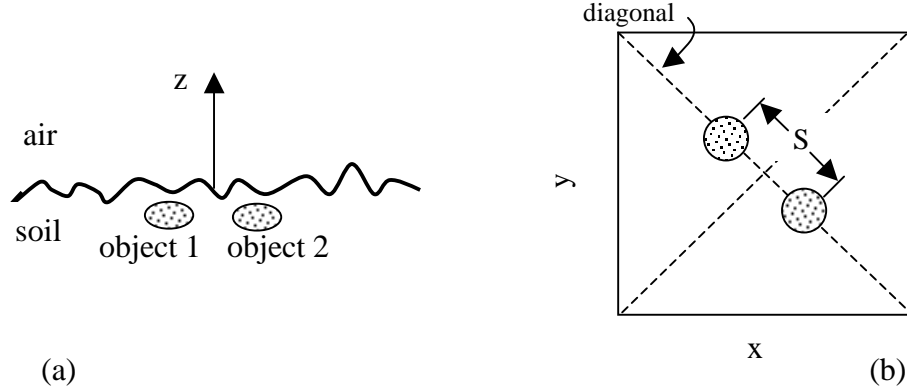


Fig. 1. (a) Cross section along the diagonal direction for two objects buried under rough ground, (b) Top view.

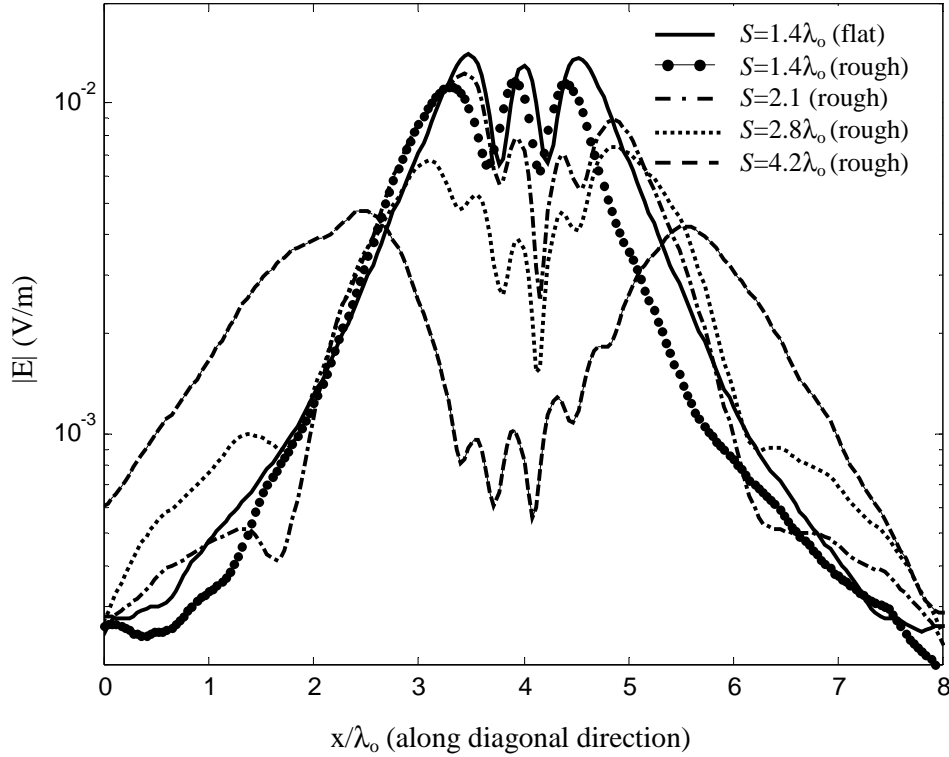


Fig. 2. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0 = 1.4-4.2$ and ground roughness parameters are $\sigma/\lambda_0 = 0.1$ and $l_c/\lambda_0 = 0.5$. The separation distance S is shown in Fig. 1b.

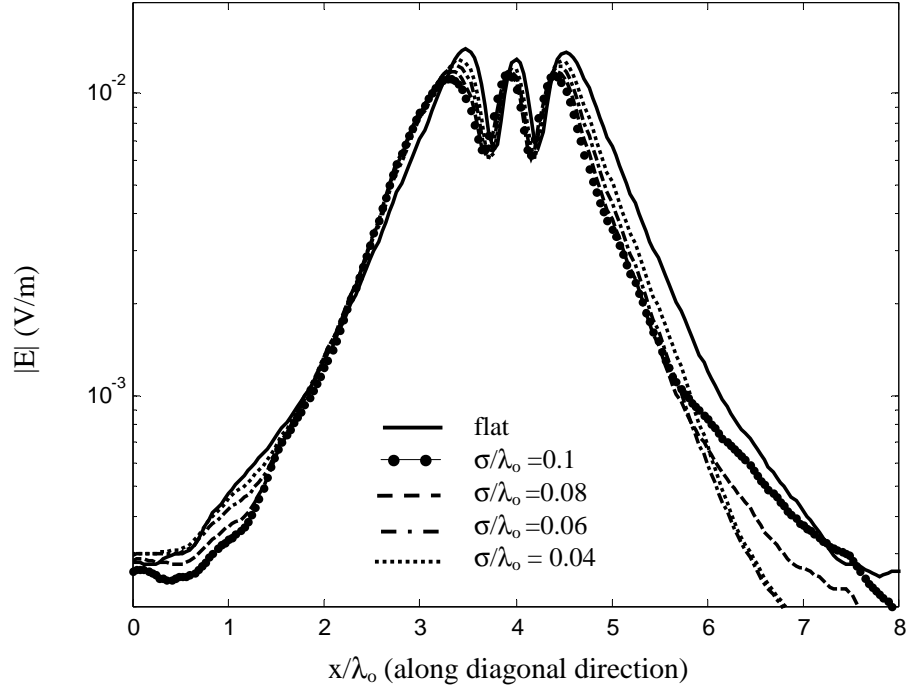


Fig. 3. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=1.4$ and ground roughness parameters are $l_c/\lambda_0=0.5$ and $\sigma/\lambda_0=0.04-0.1$.

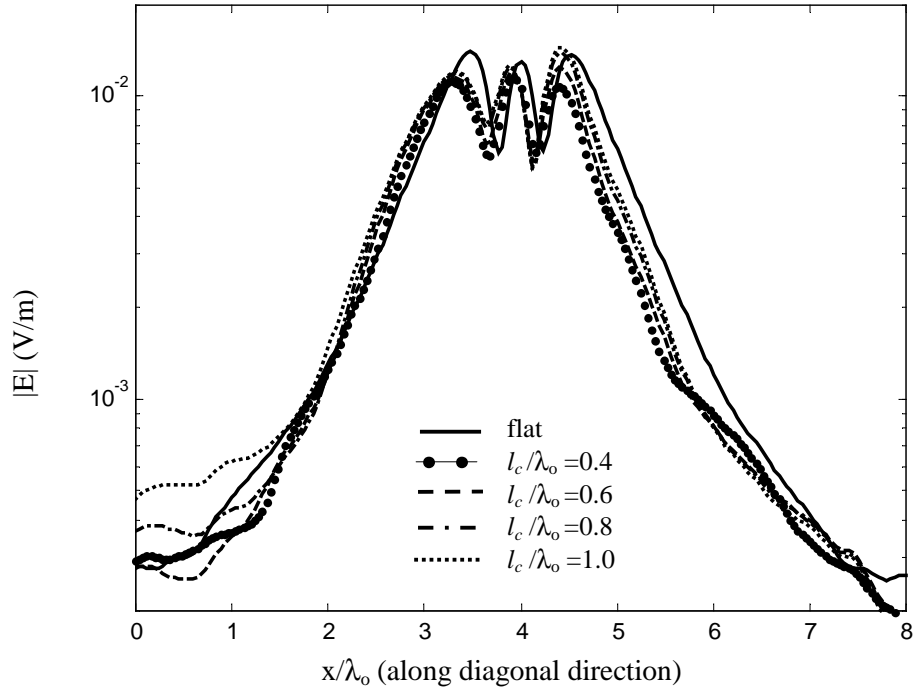


Fig. 4. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=1.4$ and ground roughness parameters are $\sigma/\lambda_0=0.1$ and $l_c/\lambda_0=0.4-1.0$.

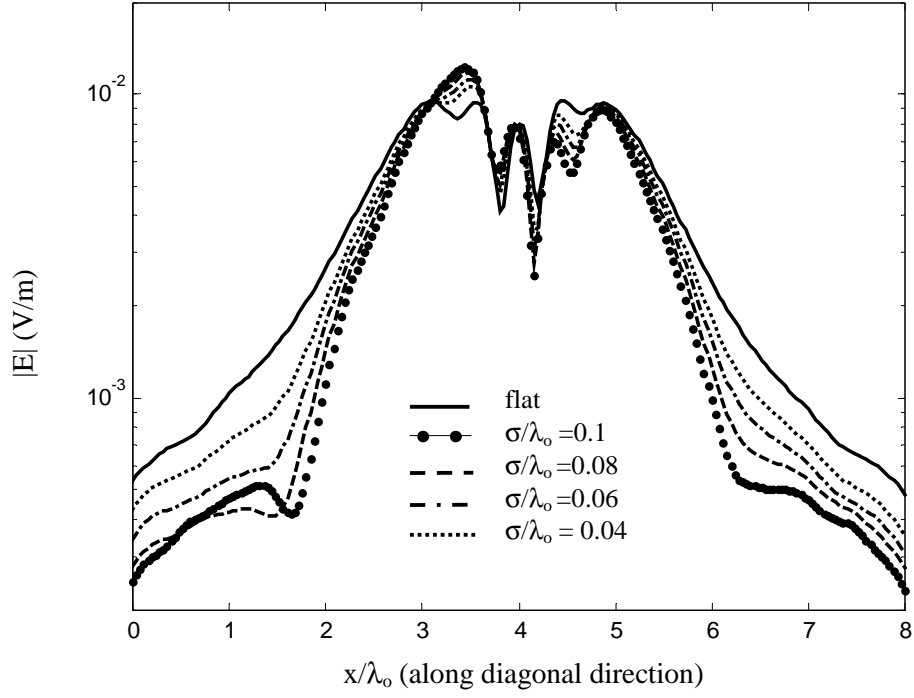


Fig. 5. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=2.1$ and ground roughness parameters are $l_c/\lambda_0=0.5$ and $\sigma/\lambda_0=0.04-0.1$

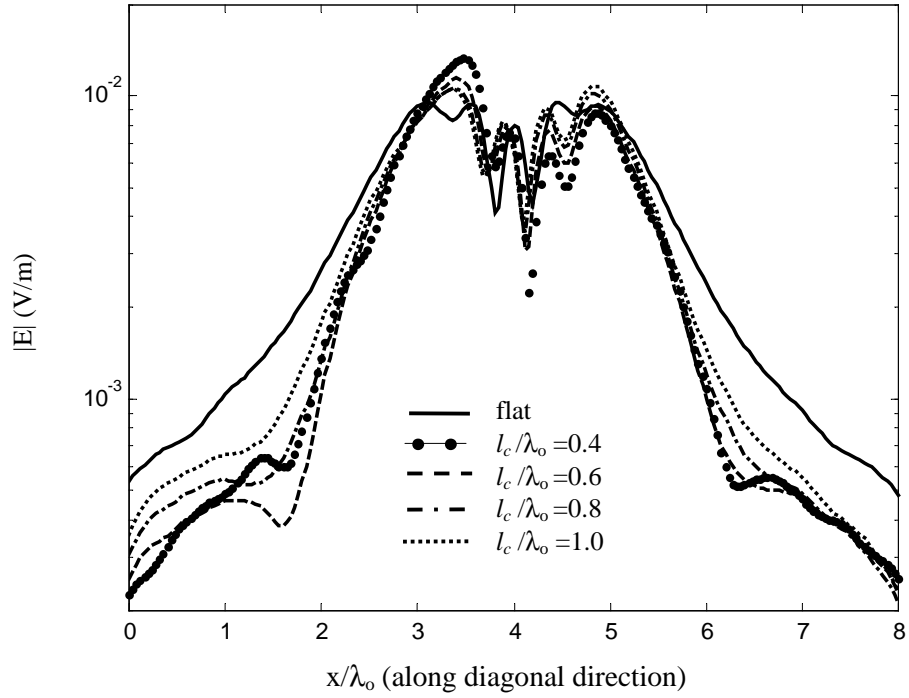


Fig. 6. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=2.1$ and ground roughness parameters are $\sigma/\lambda_0=0.1$ and $l_c/\lambda_0=0.4-1.0$.

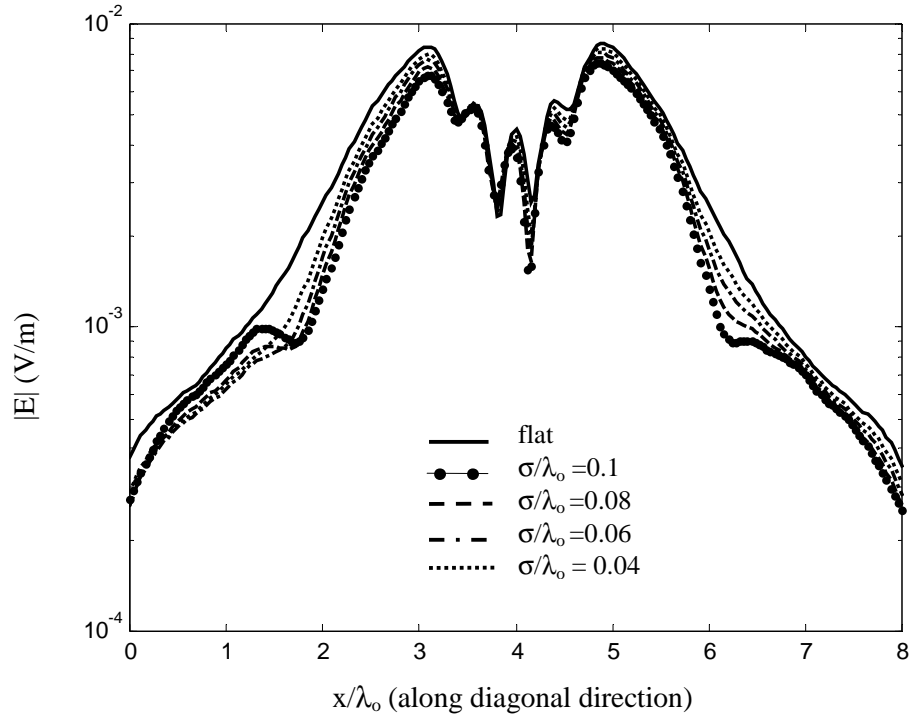


Fig. 7. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=2.8$ and ground roughness parameters are $l_c/\lambda_0=0.5$ and $\sigma/\lambda_0=0.04-0.1$

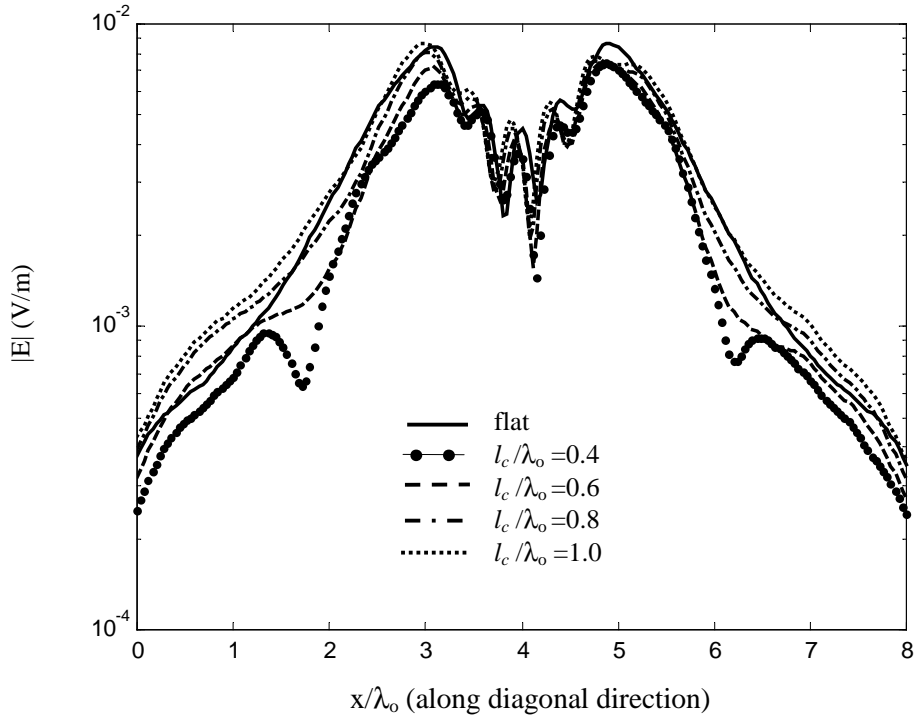


Fig. 8. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=2.8$ and ground roughness parameters are $\sigma/\lambda_0=0.1$ and $l_c/\lambda_0=0.4-1.0$.

Parametric Investigation of Ground Roughness on the Interference Between the AP-Mine and a Clutter-Object Buried Under Two-Dimensional Random Rough Surfaces

Magda El-Shenawee¹ and Carey Rappaport²

¹Department of Electrical Engineering

3217 Bell Engineering Center

University of Arkansas

Fayetteville, Arkansas 72701

Tel: (501)-575-6582, Fax: (501)-575-7967

magda@uark.edu

²Center for Subsurface Sensing and Imaging Systems

Department of Electrical and Computer Engineering

Northeastern University, 235 Forsyth BLD

Boston, MA 02115

Tel: (617)-373-2043, Fax: (617)-373-8627

rappaport@ece.neu.edu

Abstract

In realistic landmine fields, the anti-personnel plastic mine is often buried nearby a clutter-object under the ground. The presence of a second object buried near the mine under a two-dimensional (2-D) rough ground can easily obscure the target and/or cause a false alarm. The separation distance between the AP mine and clutter-object plays a significant role on the probability of true or false alarm in this situation. A rigorous electromagnetic model has been developed to analyze the scattering mechanism between the target and the clutter-object, between the target and the rough ground, between the clutter-object and the rough ground and the multiple scattering between different spots on the rough ground itself. The new rigorous model is based on the classical electromagnetic equivalence theorem leading to producing six new integral equations. Using the Method of Moment (MoM), the new integral equations are transformed into a linear system of equations to be solved for the unknown electric and magnetic currents on the surface of three scatterers; (1) rough ground, (2) target and (3) clutter-object. The MoM impedance matrix completely represents every interaction between these three scatterers. The superior Steepest Descent Fast Multipole Method (SDFMM) is used to tremendously speed up the computations of the unknown MoM surface currents.

I. INTRODUCTION

In realistic minefields, buried anti-personnel (AP) nonmetallic mines are often closely accompanied by underground clutter-objects. The presence of this object considerably obscures the targets causing a false alarm during the detection process. The separation distance between the AP-mine and the clutter-object plays a primary role on the probability of false alarms. A rigorous electromagnetic model has been developed to analyze the scattering mechanism of two dielectric objects buried beneath a rough ground surface as reported in [1]. Using the $O(N)$ fast algorithm, the Steepest Descent Fast Multipole Method (SDFMM) [2-3] tremendously accelerates the computations of the N unknown surface currents [1,4]. When the two objects were located close to one another under flat ground, the strong scattering interference generates a false response appearing to be a third buried object [1]. The dependency of the observed scattering interference on the ground roughness parameters is investigated in this work.

II. FORMULATION

The rigorous electromagnetic model derived in [1] is employed in this work where six integral equations are used to obtain the equivalent surface currents on dielectric scatterers shown in Fig. 1 (3-D scattering problem). Four different regions are involved in this scattering problem; air, soil, first object

and second object. The unknown electric and magnetic currents on the ground surface, on the target surface, and on the clutter object surface are approximated using the well-known RWG vector basis functions [5]. After some algebraic manipulations, the linear system of equations is obtained as: $\bar{\bar{Z}} \bar{I} = \bar{V}$, where the total impedance matrix $\bar{\bar{Z}}$ has order of $2(N_1 + N_2 + N_3) \times 2(N_1 + N_2 + N_3)$. The number of electric and magnetic current unknowns (edges) on the ground, on the target and on the second object are $2N_1$, $2N_2$ and $2N_3$, respectively. The tested tangential incident electric field \bar{E}^{inc} and the tested normalized magnetic field $\eta_1 \bar{H}^{inc}$ on the exterior of the ground surface are expressed in \bar{V} . The SDFMM is implemented to significantly accelerate solving the linear system of equations for the unknown current coefficients [1-4].

III. NUMERICAL RESULTS

In this Section, we investigate the scattering interference between the two buried objects as a function of their separation distances and the ground roughness. Several values for the root mean square height σ and the correlation length l_c are considered with emphasis on small roughness parameters for the AP-mine detection application. The incident wave is assumed to be a Gaussian beam at normal incidence that is carefully tapered to minimize edge effects [6]. The two objects are oblate spheroids with dimensions $a = 0.3\lambda_0$, $b = 0.15\lambda_0$ and at depth $z = -0.4\lambda_0$ measured from the center. The relative dielectric constant of the ground soil is assumed to be $\epsilon_r = 2.5 - j0.18$ and for both objects is assumed to be $\epsilon_r = 2.9 - j0.072$. The $8\lambda_0 \times 8\lambda_0$ ground surface is discretized into 60,000 electric and magnetic surface current unknowns. Each object is discretized into 600 electric and magnetic surface current unknowns. In order to analyze the object signatures, the electric fields scattered from each rough ground are removed by subtraction similar to our work in [4].

In Fig. 2, the object signatures are plotted across the diagonal as shown in Fig.1. When the two objects are separated by $S = 1.4\lambda_0$, the results show three peaks of almost equal magnitudes; the first peak is above the first object, the second peak is above the second object and the third peak is at mid-point between them. This third peak is due to the strong constructive interference between the two objects. This phenomenon could easily cause a false alarm during the detection process. However, when the separation distance increases, the mid-point peak is dissolved into several secondary peaks as observed when $S = 2.1\lambda_0$ and $2.8\lambda_0$. These secondary peaks become insignificant when S is increased to $4.2\lambda_0$. As noticed in this figure, the asymmetry around the mid-point ($x = 4.0\lambda_0$) is clearly caused by the random roughness of the ground. Moreover, the magnitudes of all peaks decrease with increasing separation distance because the objects are further from the beam footprint center (ground center in this work). In Figs. 3 and 4, the separation distance is $S = 1.4\lambda_0$, the range for σ is $0.04\lambda_0$ - $0.1\lambda_0$ and the range for l_c is $0.4\lambda_0$ - $1.0\lambda_0$. The results clearly show the influence of the roughness parameters, however, the observed strong interference (mid-point peak) is slightly affected by the ground roughness in this case. This study is repeated for $S = 2.1\lambda_0$ as shown in Figs. 5 and 6, and for $S = 2.8\lambda_0$ as shown in Figs. 7 and 8. It is interesting to notice that when the separation distance becomes several correlation lengths, as shown in Figs. 6 and 8, stronger scattered interference is observed. This is demonstrated by the increase in the number of secondary peaks from three in the flat ground case to four and five in the rough ground case as shown in Figs. 6 and 8, respectively. This mechanism could increase the possibility of false alarms.

IV. CONCLUSIONS

False alarms could easily occur because of the interference mechanism between the target and a second nearby object buried under the ground. The roughness parameters strongly influence the scattering interference mechanism between the two objects, which could increase the probability of false alarms.

ACKNOWLEDGMENTS

This research was sponsored by the Northeastern University's Demining MURI grant # DAA 0-55-97-0013 and in part by the College of Engineering at the University of Arkansas. The SDFMM was originally developed at the UIUC by V. Jandhyala, E. Michielssen and W. Chew.

References

- [1] M. El-Shenawee, "The Steepest Descent Fast Multipole Method for Scattering from Two Penetrable Shallow Objects Buried Under Two-Dimensional Randomly Rough Surface," submitted to the *IEEE Trans. Anten. & Propag.* under review.
- [2] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [3] V. Jandhyala, E. Michielssen, B. Shanker and W.C. Chew, "A combined steepest descent-fast multipole algorithm for the fast analysis of three-dimensional scattering by rough surfaces," *IEEE Trans. Geosci. & Remote Sensing*, vol. 36, no. 3, pp. 738-748, May 1998.
- [4] M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo simulations of electromagnetic wave scattering from random rough surface with 3-D penetrable buried object: mine detection application using the SDFMM," *J. Optical Society of America A*, to appear in December 2001
- [5] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Anten. & Prop.*, vol. AP-30, no. 3, pp.409-418, May 1982.
- [6] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: enhanced backscatter," *Phy. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.

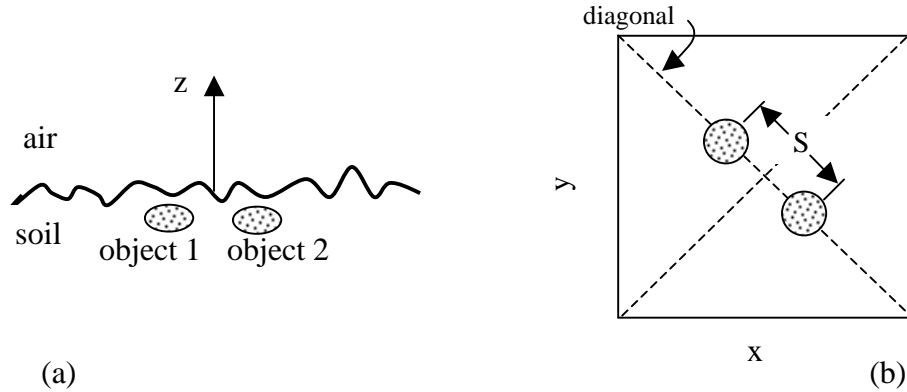


Fig. 1. (a) Cross section along the diagonal direction for two objects buried under rough ground, (b) Top view.

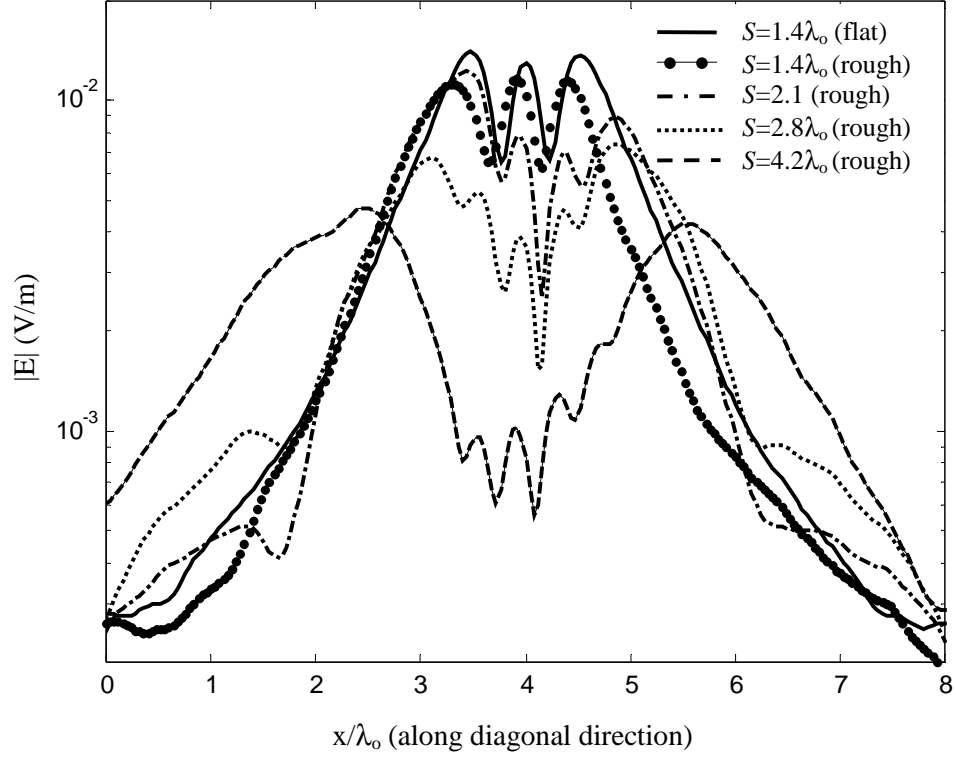


Fig. 2. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=1.4-4.2$ and ground roughness parameters are $\sigma/\lambda_0=0.1$ and $l_c/\lambda_0=0.5$. The separation distance S is shown in Fig. 1b.

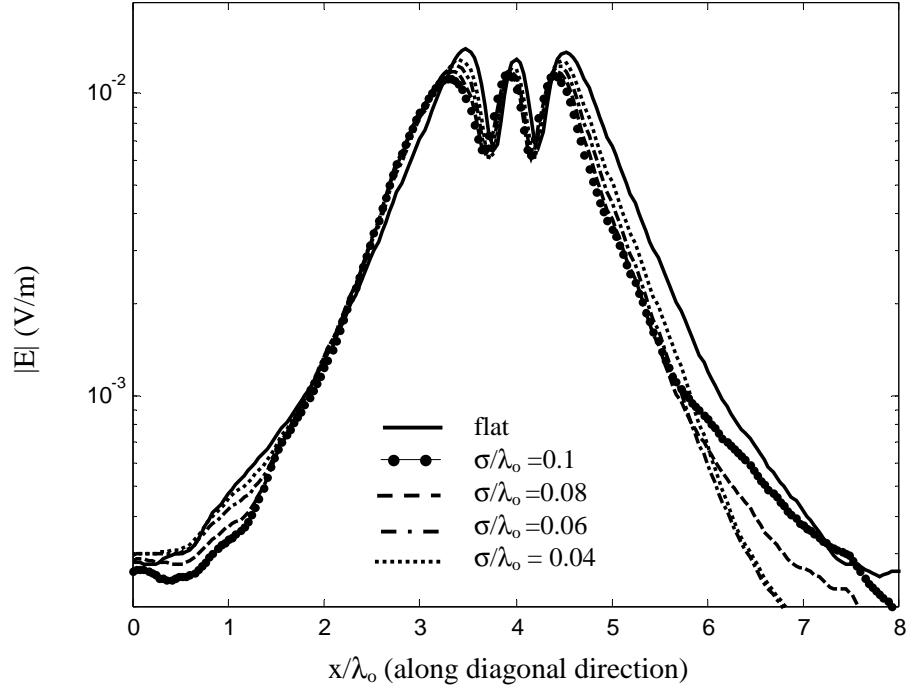


Fig. 3. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=1.4$ and ground roughness parameters are $l_c/\lambda_0=0.5$ and $\sigma/\lambda_0=0.04-0.1$.

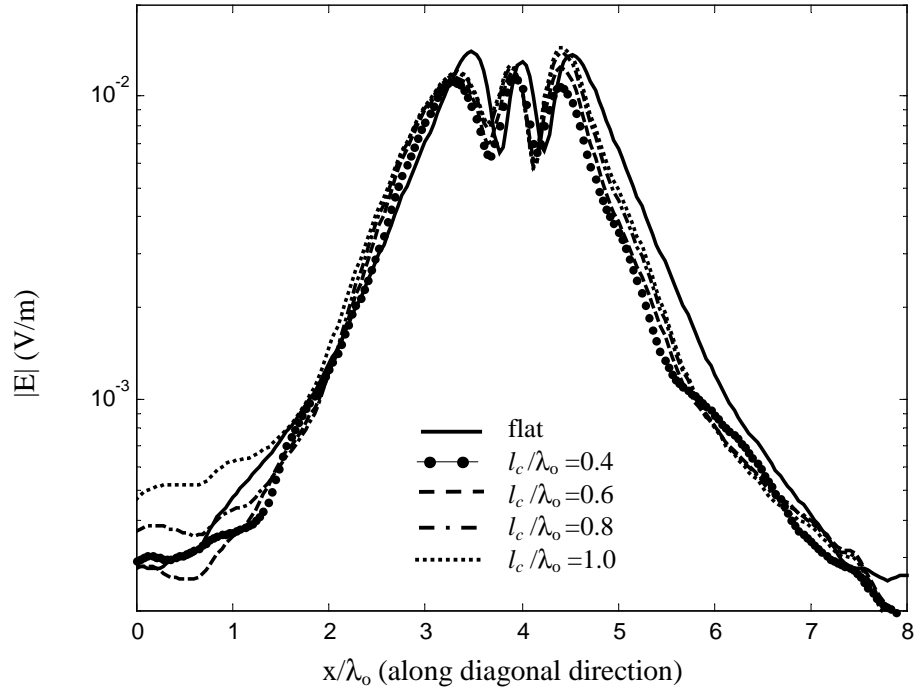


Fig. 4. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=1.4$ and ground roughness parameters are $\sigma/\lambda_0=0.1$ and $l_c/\lambda_0=0.4-1.0$.

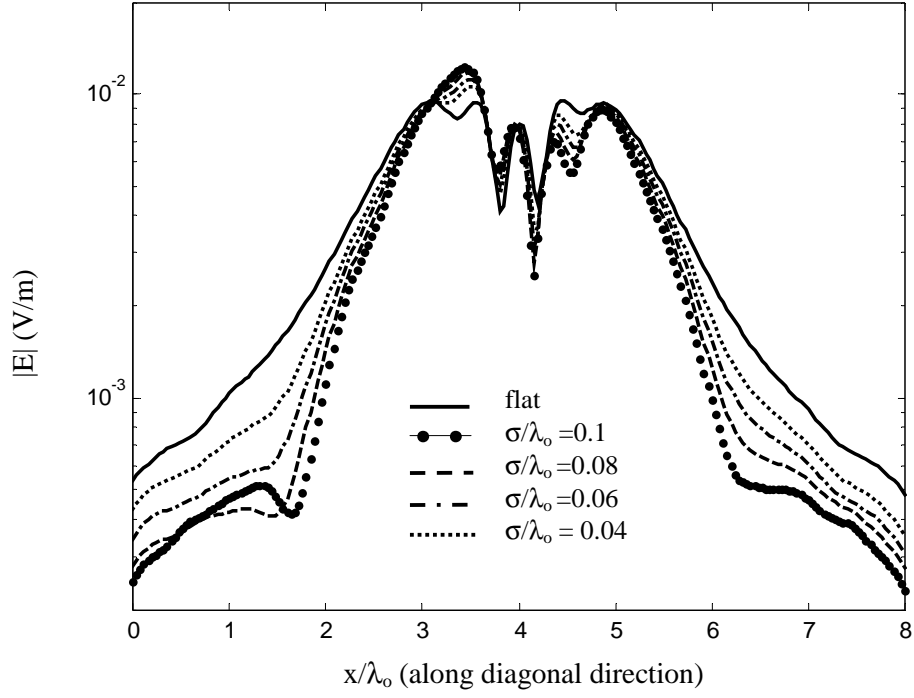


Fig. 5. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=2.1$ and ground roughness parameters are $l_c/\lambda_0=0.5$ and $\sigma/\lambda_0=0.04-0.1$

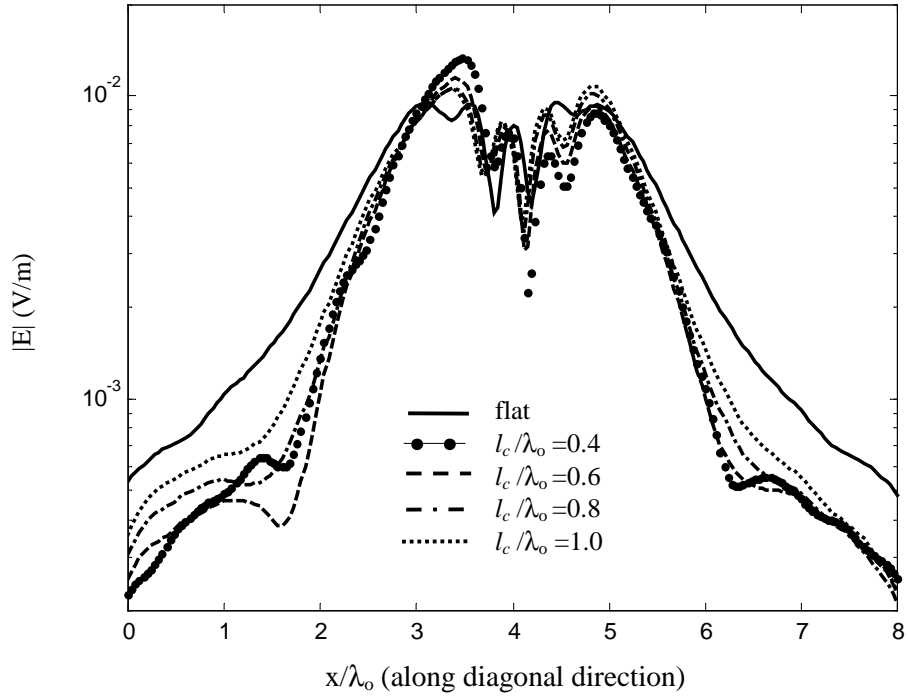


Fig. 6. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=2.1$ and ground roughness parameters are $\sigma/\lambda_0=0.1$ and $l_c/\lambda_0=0.4-1.0$.

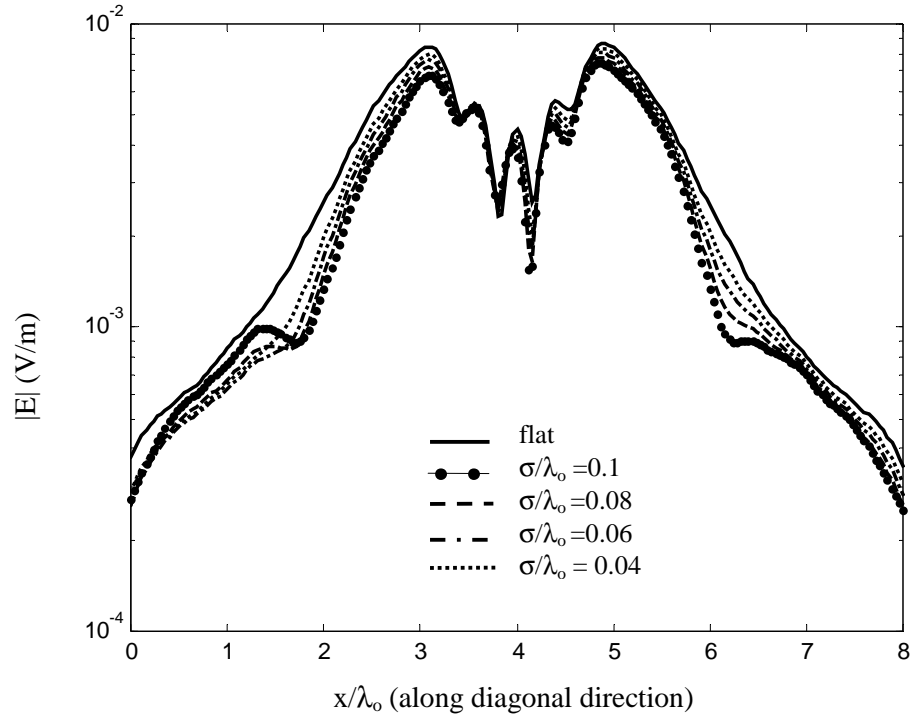


Fig. 7. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=2.8$ and ground roughness parameters are $l_c/\lambda_0=0.5$ and $\sigma/\lambda_0=0.04-0.1$

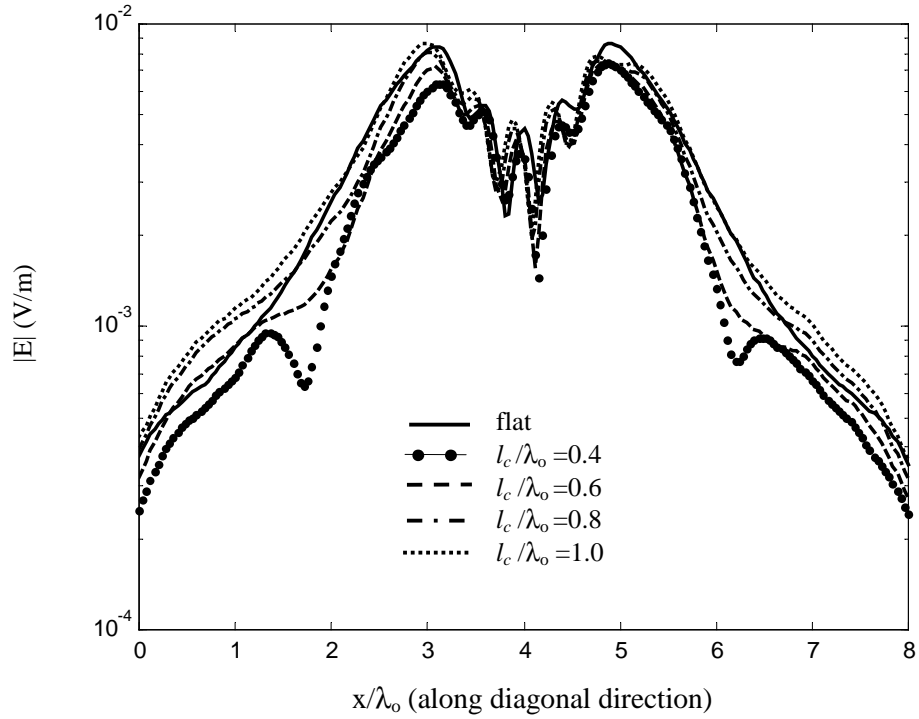


Fig. 8. Scattered near-electric field of just the two objects at normal incidence when $S/\lambda_0=2.8$ and ground roughness parameters are $\sigma/\lambda_0=0.1$ and $l_c/\lambda_0=0.4-1.0$.

Ground Penetrating Antenna Design for Land Mine Detection

Brent Poliquin and Carey Rappaport

Center for Subsurface Sensing and Imaging Systems
Department of Electrical and Computer Engineering
Northeastern University, 302 Stearns Building
Boston, MA 02115
Tel: (617) 373-2043
Fax: (617) 373-8627
rappaport@neu.edu
bpoliqui@coe.neu.edu

ABSTRACT

This paper analyzes various ground-contact microwave antenna designs for land mine detection. Designs include monopole probes with differing tip lengths, and small loop probes. Factors affecting the performance of the antenna include position of the antenna in the soil, angle of insertion into the ground, type of soil (loam, sand, etc.), and the moisture of the soil. It is important to find an antenna whose frequency response varies minimally with depth. Using a metal rod and a deactivated plastic mine as targets, a soil tank, and a network analyzer to observe the coupling of the antenna to the soil and to the target, we experimentally show that for the best antenna tested — a 2 cm tip monopole probe — both targets could be unambiguously detected and discriminated.

INTRODUCTION

Although there are many technologically advanced means of detecting mines in the field, the problem is far from being solved. It is important to consider detector cost, practicality, and adaptability to currently used systems. While a scientifically sophisticated sensor might be able to perfectly image a buried mine shape anomaly or detect the presence of high explosive, these systems often require laboratory-like background conditions, lengthy acquisition times, or a large truck full of support equipment. To address the current high-priority need for improving the effectiveness of current mine detection equipment, this study examines a relatively simple adaptation of the confirmatory probe.

The vast majority of handheld mine detection is performed by first alarming at the presence of buried metal with a electromagnetic induction sensor, and then poking the ground in the vicinity of the suspected target with a narrow nonmetallic probe. This is tedious, dangerous, and often frustrating work, with the most likely outcome being finding nothing: no mine, no metal, not even a rock. Speeding up this probing process by lowering the false alarm rate, and reducing the danger by providing advanced warning would be of great advantage to the person sweeping the field.

The mine detection problem is even more difficult when the target is nonmetallic. Although an EMI metal detector can often detect the small metal mass in firing pins and striker plates and cups in plastic mines, its sensitivity must be increased so much that any scrap bit of metal would register an alarm. In addition, plastic mines share similar dielectric characteristics to soil, thus making it harder to locate with ground penetrating radar [1, 2].

With rough ground surfaces, the random clutter intensity caused by the ground reflection is often greater than the scattering by the nonmetallic target [3, 4]. One advantage of a probe-mounted ground penetrating antenna is that the interrogating signal starts under the ground surface. The clutter from the random rough ground surface still has a measurable effect, but since the reflection is from below the surface, it is much less than for conventional ground penetrating radar.

When designing antennas for this purpose, many factors need to be taken into consideration. The antenna radiation pattern in soil is important. If the antenna generates a symmetrical field pattern, it can be inserted into the ground at any orientation relative to the target. The best frequency range that is both relatively independent of antenna depth in soil and most sensitive to foreign objects must be determined. And it goes without saying that the antenna should be able to pierce the soil with minimal probability of disturbing the land mine.

This experimental study considered two types of antennas which could be easily incorporated into a soil probe: a monopole, constructed by simply removing the outer conductor of a semi-rigid coaxial cable; and a small loop, constructed by bending the extended coaxial center conductor into a loop and joining it with the outer conductor.

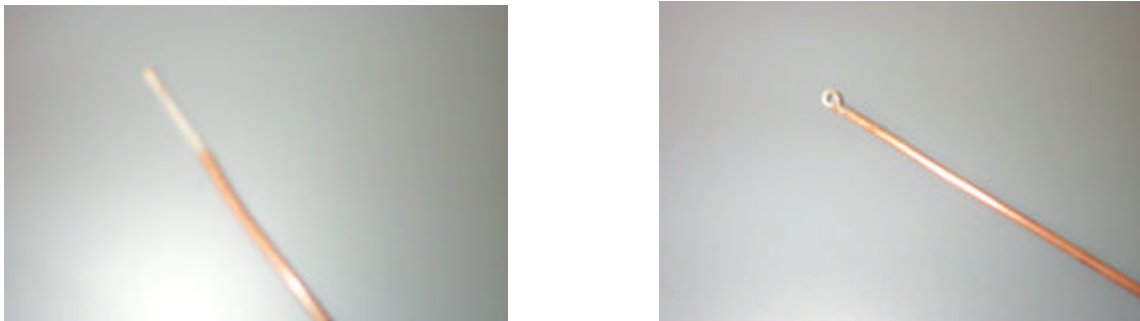


Figure 1: Monopole and loop antennas suitable for ground penetrating probes

By analyzing the reflected signal (S_{11}) from these antennas with a network analyzer it is possible to determine some of the characteristics of the surrounding medium. Measurements were made in the frequency range of 0.5 GHz to 8 GHz to observe any wideband resonance effects and ensure that there was adequate sensitivity to short distances from the probe to the target.

LOOP ANTENNA DESIGN

Constructing a 5 mm loop antenna requires stripping approximately 1.5 cm of the copped outer conductor from a semi-rigid cable, and bending the inner wire around into a loop and soldering it to the remaining outer copper conductor. This exposed loop can be modeled as a small magnetic dipole, couple microwave power primarily though the magnetic field passing through the loop.

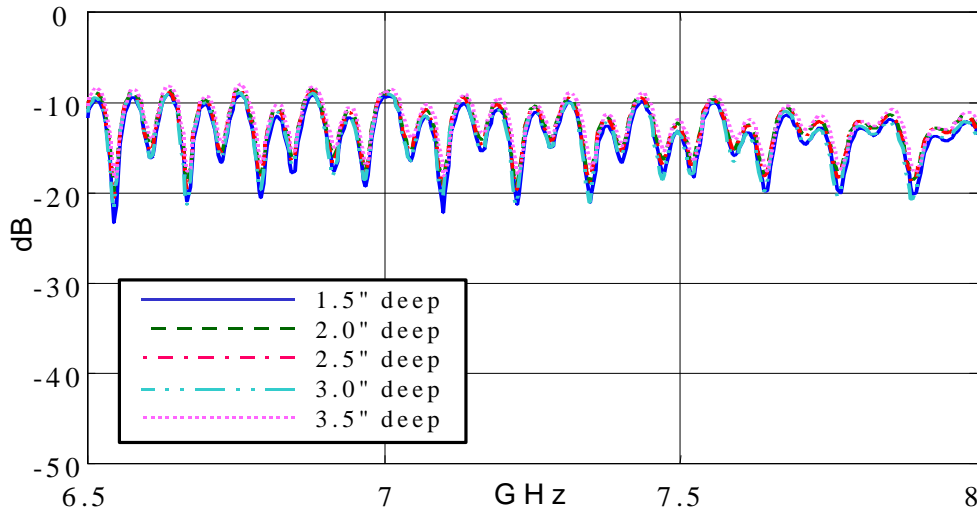


Figure 2: Measured S_{11} as a function of loop antenna depth.

Our results showed that the loop coupled well with the soil, especially from 6.5 GHz to 8 GHz. Its depth independence within this frequency range is excellent, as shown in Figure 2, with almost no variation in S_{11} for 1.5 to 3 inch loop burial depth. However, the sensitivity of the loop to other objects in the soil was poor, as indicated in Fig. 3. It is difficult to distinguish the differences caused by nearby targets.

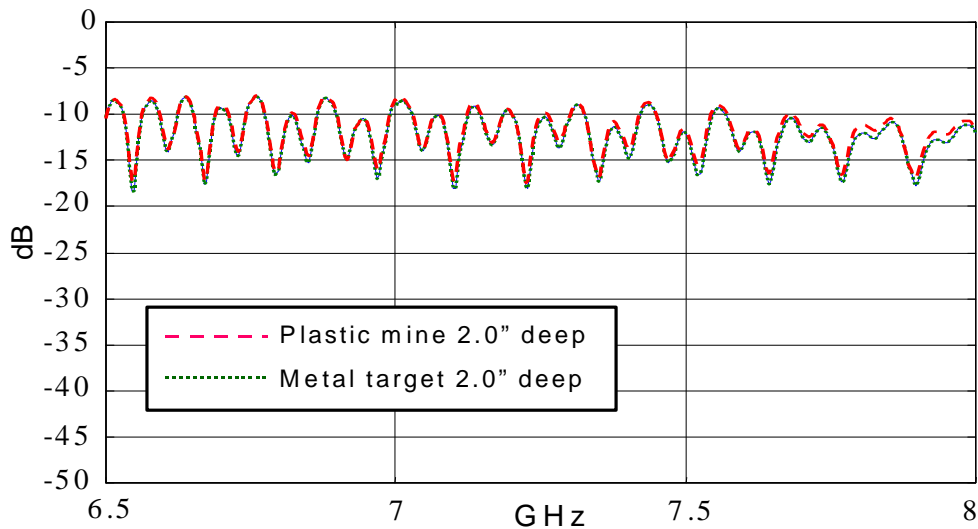


Figure 3: Measured S_{11} of loop antenna near buried targets.

MONOPOLE PROBE ANTENNA DESIGN

The nearfield pattern created by the simple monopole probe of Figure 1 is as uniform as possible, which is advantageous since it allows the observer to place the antenna at almost any orientation in the ground and expect similar results. The only design consideration with this type of probe antenna is the amount of outer copper conductor stripped off. The plastic coating on the inner wire was kept to increase durability for our testing. Probes with different tip lengths were tested (1 cm to 3 cm).

The various length monopole probe antenna were tested for its frequency dependent electromagnetic coupling to the surrounding soil medium, and its dependence with depth of insertion and the extent of perturbation in the presence of nearby buried plastic and metal targets. An ideal antenna will transmit the all power supplied to it across a range of frequencies no matter the depth of the soil in which it is buried. The 2 cm tip probe gave the best results. Figures 4 and 5 show its depth-independent coupling in moist clay soil.

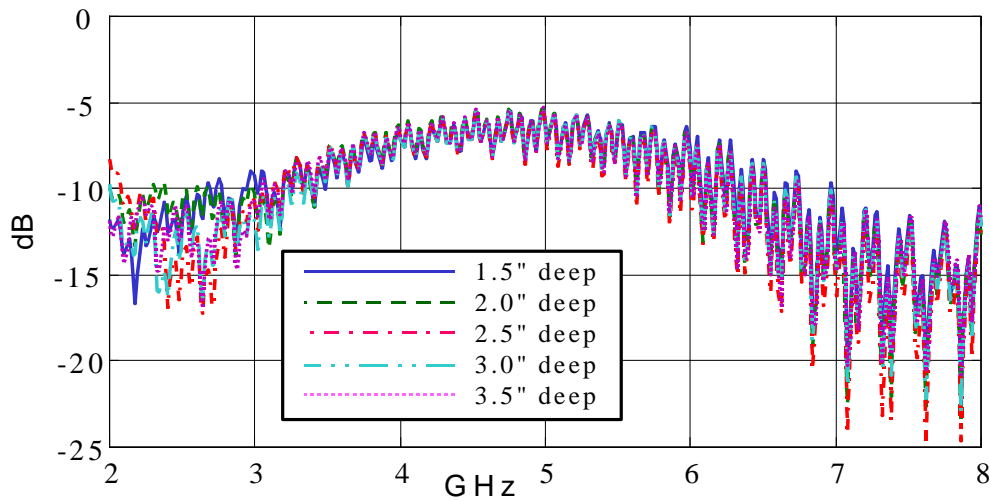


Figure 4: Measured S_{11} of 2 cm tip monopole antenna at 60 degrees insertion angle for a variety of burial depths

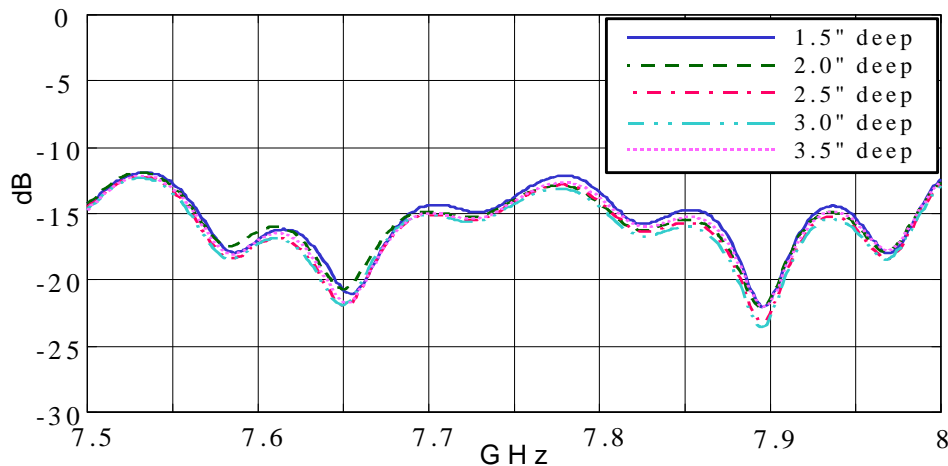


Figure 5: Detail of Figure 4.

Measurements were conducted for both the normal inserted probe, and the more realistic case of 60° insertion angle. We found better results for the latter, with best discrimination of targets in the 7.5 GHz to 8 GHz a range. The 2 cm tip monopole could differentiate between no target, a thin metal rod, and a plastic antipersonnel mine in close proximity for both 2.0 cm (Figure 5) and 2.5 cm (Figure 6). insertion depths in both moist soil and dry sand (Figure 7).

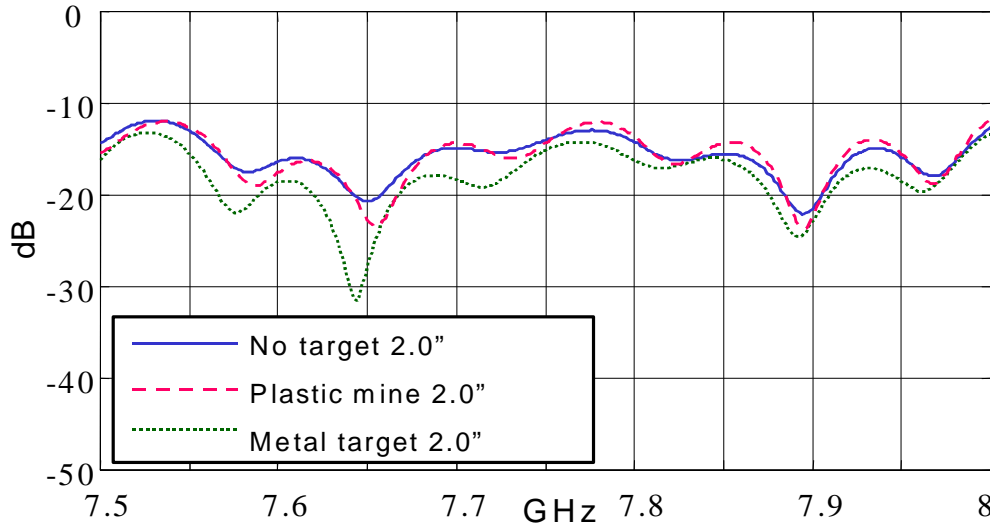


Figure 5: Measured S_{11} of 2 cm tip monopole antenna at 60 degrees insertion angle at 2.0 cm below surface near mine targets in moist clay soil

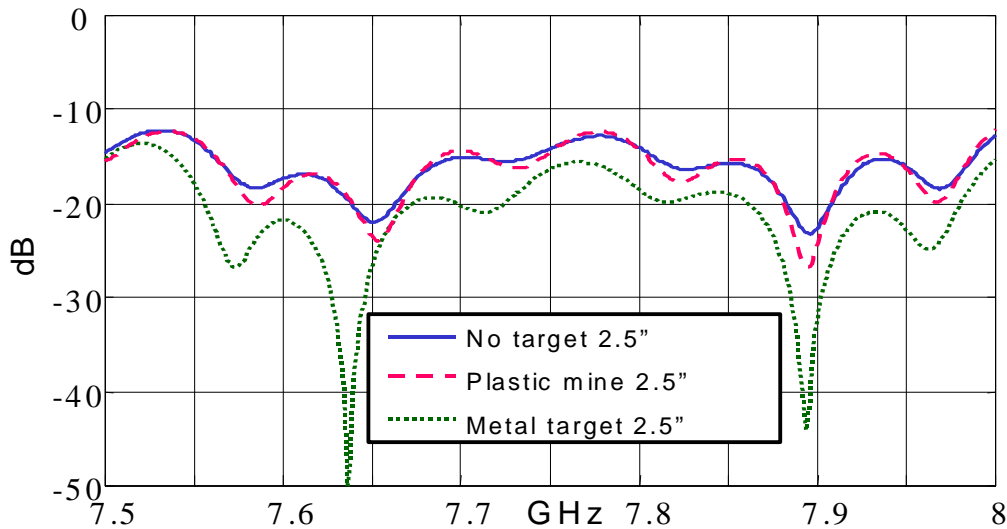


Figure 6: Measured S_{11} of 2 cm tip monopole antenna at 60 degrees insertion angle at 2.5 cm below surface near mine targets in moist clay soil

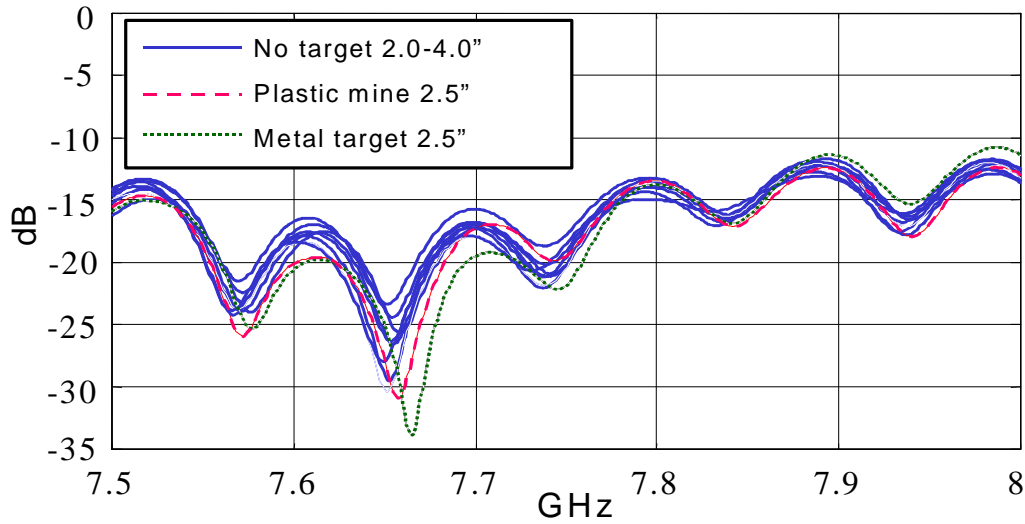


Figure 7: Measured S_{11} of 2 cm tip monopole antenna at 60 degrees insertion angle at 2.0 cm below surface near mine targets in dry sand

CONCLUSIONS

Depth independent soil coupling across a wide frequency range with a simple symmetric field pattern are important characteristics of a good probe-based ground penetrating antenna. We examined several types of small antennas that could be readily incorporated into existing mine confirmation probes, and found that the simple monopole antenna satisfied these requirements. The 2 cm monopole probe was best able to couple to common soils yet detect impedance differences of small buried nearby plastic and metal objects. Using a monopole antenna to sense the impedance variations in soil has the potential of aiding the mine detection process by alarming before actually contacting the object.

ACKNOWLEDGEMENTS

The author gratefully acknowledges support for this work from the Army Research Office, MURI grant # DAA 0-55-97-0013, 0013 and by CenSSIS, the Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation, award number EEC-9986821. Ann Morgenthaler's assistance in preparing this paper is appreciated.

REFERENCES

1. J. Hipp, "Soil Electromagnetic Parameters as Functions of Frequency, soil Density and Soil Moisture," *Proceedings of the IEEE*, vol. 62, no. 1, pp. 98-103, Jan.1974
2. A. von Hippel, *Dielectric Materials and Applications*, Wiley, New York, 1953
3. Young, J., Poirier, M., and Peters, L., "A review of current ground penetrating radar concepts," *IEEE Antennas and Propagation Symposium Digest*, pp.1250-1253, June 1992.
4. Peters, L., and Young, J., "Applications of subsurface transient radars," I *Time Domain Measurements in Electromagnetics*, Miller, E. Ed., New York: Van Nostrand Reinhold, 1986.
5. O'Neill, K., Laskey, R., and Paulsen, K., "Scattering from a metallic object near the randomly rough surface of a lossy dielectric," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 24, no. 2, pp. 367—376, 1996.

Modeling Radar Detection of Floating Sea Mines

Carey Rappaport
Northeastern University
Boston, MA 02115
rappaport@neu.edu

Abstract

To detect and classify mines floating in seawater using radar, it is essential to understand and quantify the electromagnetic scattering by the target immersed in its realistic cluttered background. While the metallic surface of the mine presents a strong scattering target, the water surrounding it will strongly scatter incident waves as well. In addition, seawater is electromagnetically lossy, readily absorbing incident waves. While deeply submerged mines are undetectable, a floating mine might be partially exposed, scattering waves that could be distinguished from those scattered by the background. The degree to which the target scattered waves can be discriminated from the waves scattered by the random water surface height variations determines whether the mine can be detected in realistic situations.

Computational modeling shows that even when a mine is mostly submerged, the scattered fields are still quite characteristic of a large metallic scatterer. Close inspection shows that the scattered waves are scattered most strongly in the vertical direction. This study shows that only a small amount of the target need be exposed to generate significant scattered field that is distinguishable from a very rough water surface.

Introduction

To determine the expected scattering features of the mine in its actual environment, a full phenomenological study of the target in a variety of backgrounds must be conducted. While it might be possible to experimentally measure a portion of the fields scattered by a scale model mine in a test tank, it would be extremely challenging to capture the entire wave behavior throughout all space near the target without perturbing the field one is measuring with measurement antennas. In addition, controlling the state of the water surface is problematic. Thus, the only practical way to capture the scattered field is to model it numerically.

Current electromagnetic computational capabilities allow for sophisticated modeling of wave scattering on desktop computers. The most applicable modeling method for this problem for standard radar sources is the finite difference frequency domain (FDFD) method. For impulse radars, the time domain version, FDTD is best. Both the FDFD and FDTD methods discretize all space in the vicinity of the target, and approximate the differential form of Maxwell's equations with second-order finite differences of nearest neighbors. Both methods are flexible, with easy specification of rough surfaces, corners, and fins, and both can model metal, conductive dielectrics, and frequency-dependent material.

One important detail, which must be carefully addressed for exterior, unbounded geometries, is simulating the wave's radiation to infinity. Although the computational domain terminates at the end of the array lattice, the wave must act as if it were continuing outward. In other words, the wave must not reflect off the artificial boundary. The boundary must absorb all incident waves; hence it is referred to as an Absorbing Boundary Condition (ABC). In principle, no numerical condition can absorb waves from every incident angle, but there have been several ABC's proposed which absorb those near normal incidence making use lossy matched layers and pseudo-differential annihilation operators.

The FDFD method requires the solution of the set of simultaneous equations that self-consistently describe the electric and/or magnetic fields everywhere in space for a given frequency [1, 2]. As such, the resulting matrix equation involves a very large, very sparse square matrix. Depending on the ABC, as few as a dozen entries in any given row of a 2-D or 3-D FDFD matrix are non-zero. The ABC used in this analysis is the perfectly matched layer (PML) of Berenger [3-6]

Prototypical Problem Geometry

As an example of the 2-D modeled fields near a typical spherical sea mine in rough seas is shown in Fig. 1. This geometry considers a mine floating substantially out of the water. The water surface varies with Gaussian height and correlation length statistics, which might be typical of ocean waves. The various identified regions represent those perturbations to a nominal air/water half-space that give rise to the non-ideal scattered waves. The nominal plane wave scattering from a smooth, target-free problem geometry is determined analytically, and can be added to the perturbation-scattered field to give the total field.

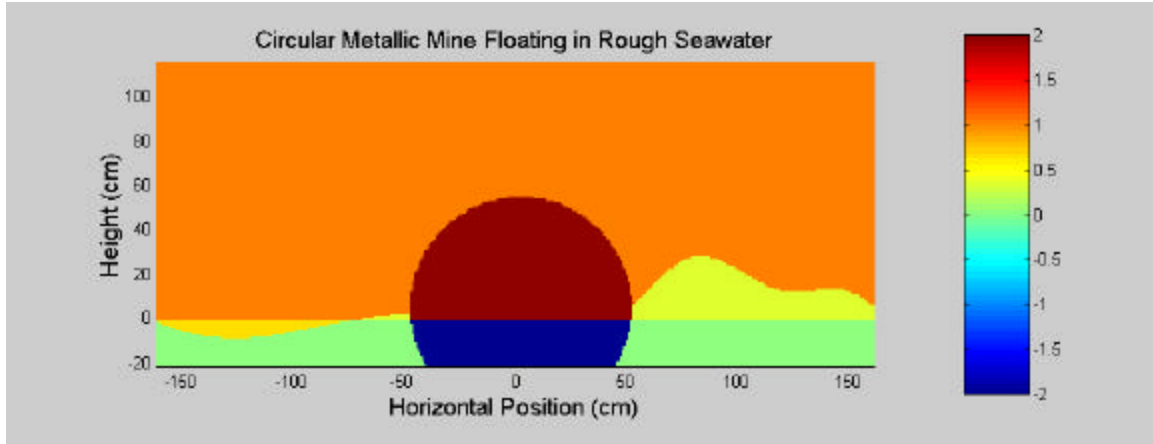


Figure 1 Problem geometry, showing various modeled regions: air (top, orange), nominal water (bottom, green), water peaks (middle right, yellow), water troughs (middle left, salmon), mine below nominal water level (bottom middle, blue), mine above nominal water level (center, red).

Computed Results

The total field for 300 MHz normally incident plane wave excitation (as generated by an airborne radar pointing straight down) for the target geometry of Fig. 1, is shown in Fig. 2. Note the recognizable outline of the circular metallic target, the almost complete exclusion of field from the water, and the change in standing wave structure around the mine. Also note the left/right asymmetry of field due to the differences of the surface heights on the left and right sides of the problem space.

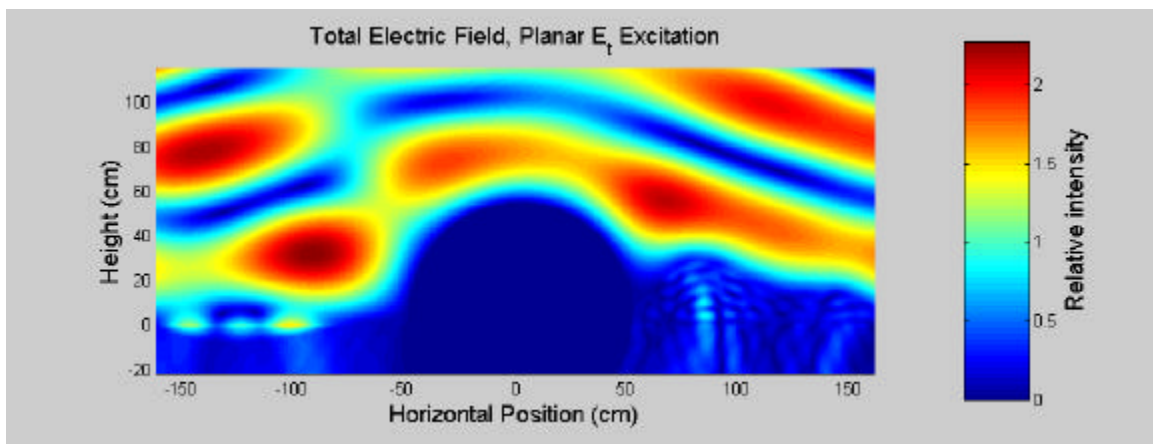


Figure 2 Total electric field at 300 MHz, for horizontally polarized E-field (into plane of figure), for geometry of Figure 1.

Figure 3 shows the scattered field only; e.g., fields of Fig. 2 less the analytic plane wave scattering of the planar water half-space. From a radar detection perspective, it is the field of Fig. 3 that would be measured once the nominal signal is removed. The waves scattered by the mine are significantly greater than those of the rough water surface. It appears clear that this target could be picked out of the background clutter with radar.

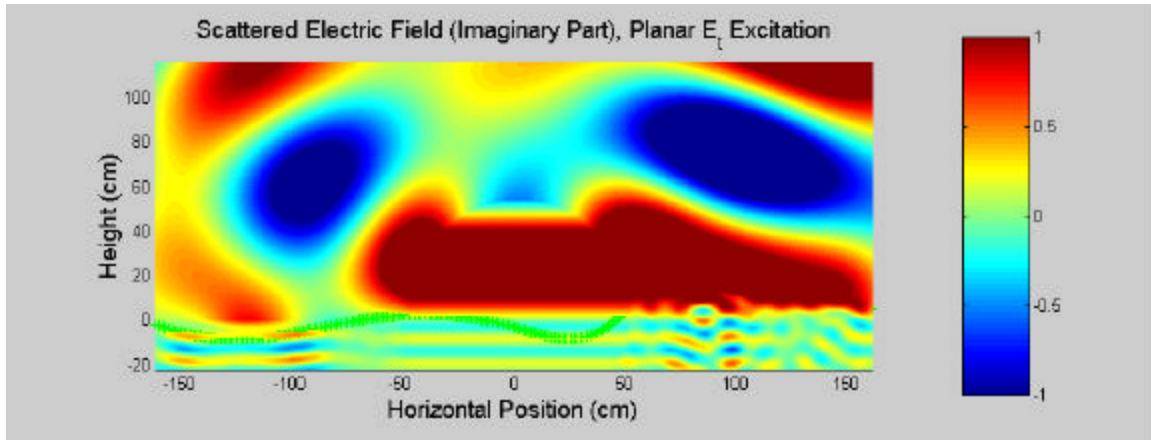


Figure 3 Electric field at 300 MHz due to just the mine and the surface roughness, for horizontally polarized E-field (into plane of figure), for geometry of Figure 1.

Figures 4 and 5 show the total and scattered fields for the same problem geometry at the higher frequency of 500 MHz. Clearly the wave structure is more complicated, with additional nodal structure in the horizontal direction. While the target resolving capability of the radar increases with frequency, the clutter effects increase as well. Choosing the best frequency, or set of frequencies to detect a particular mine target in the presence of a given water surface roughness is an important part of this study.

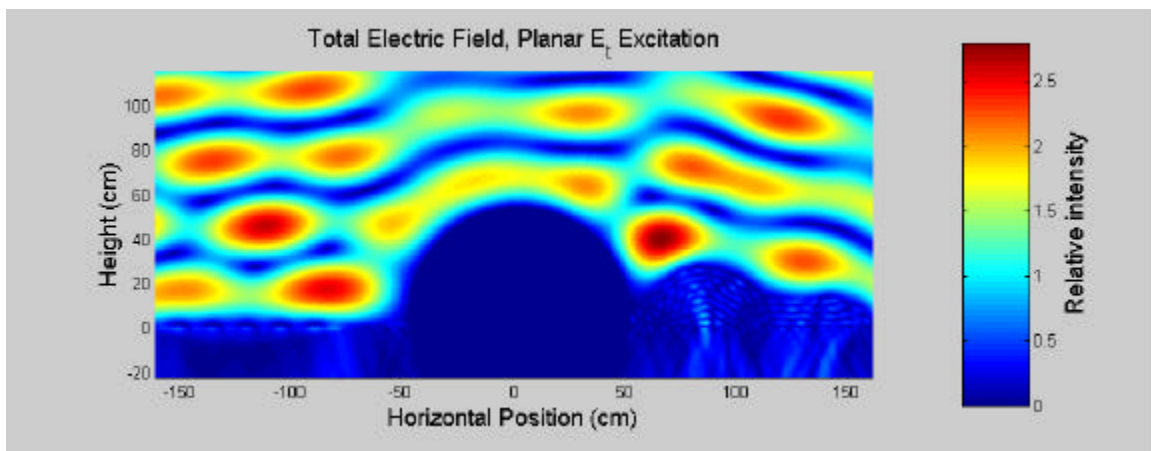


Figure 4 Total electric field at 500 MHz, for horizontally polarized E-field (into plane of figure), for geometry of Figure 1.

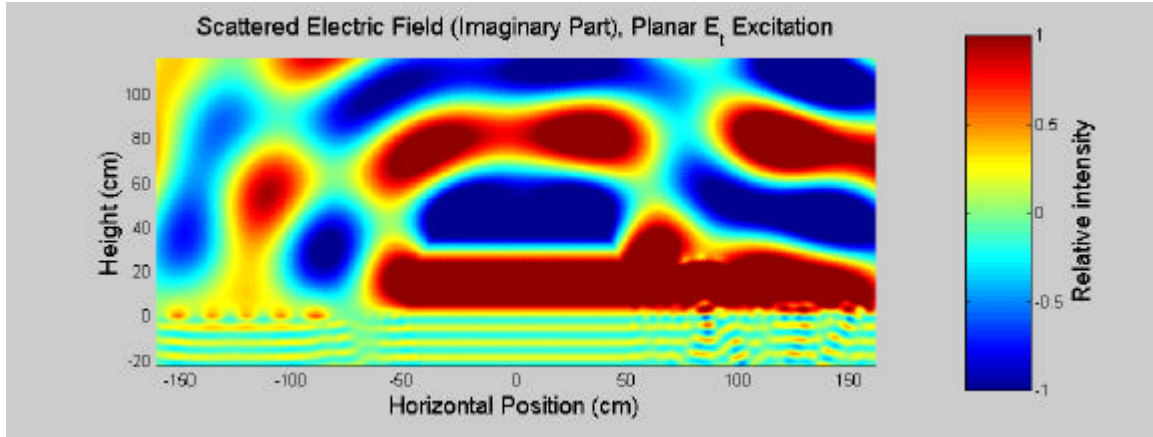


Figure 5 Electric field at 500 MHz due to just the mine and the surface roughness, for horizontally polarized E-field (into plane of figure), for geometry of Figure 1.

To consider a slightly different geometry, we test a deeper mine target, as shown in Figure 6. The water surface is chosen to be the same for comparison purposes. The total and scattered fields for this deeper geometry are shown in Figures 7 and 8. Although this mine is mostly submerged, the scattered fields are still quite characteristic of the large metallic scatterer. Close inspection shows that the scattered waves are scattered most strongly in the vertical direction, as opposed to the more widely scattered waves of the greatly protruding mine of Fig 5. This example shows that even at the higher frequency, only a small amount of the target need be exposed to generate significant scattered field that is distinguishable from a very rough water surface. Using multiple frequencies greatly enhances the potential for target discrimination [7].

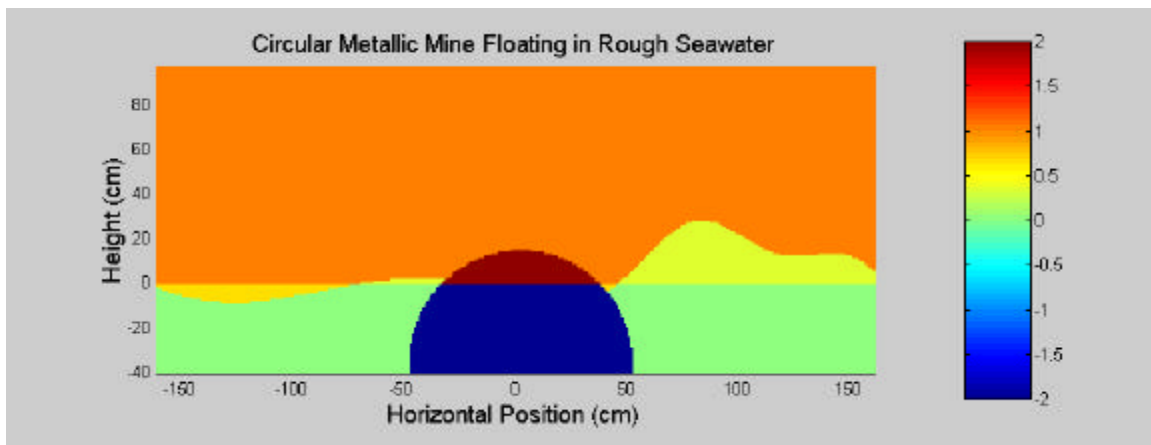


Figure 6 Problem geometry for deeper floating mine with the same water surface profile of Figure 1

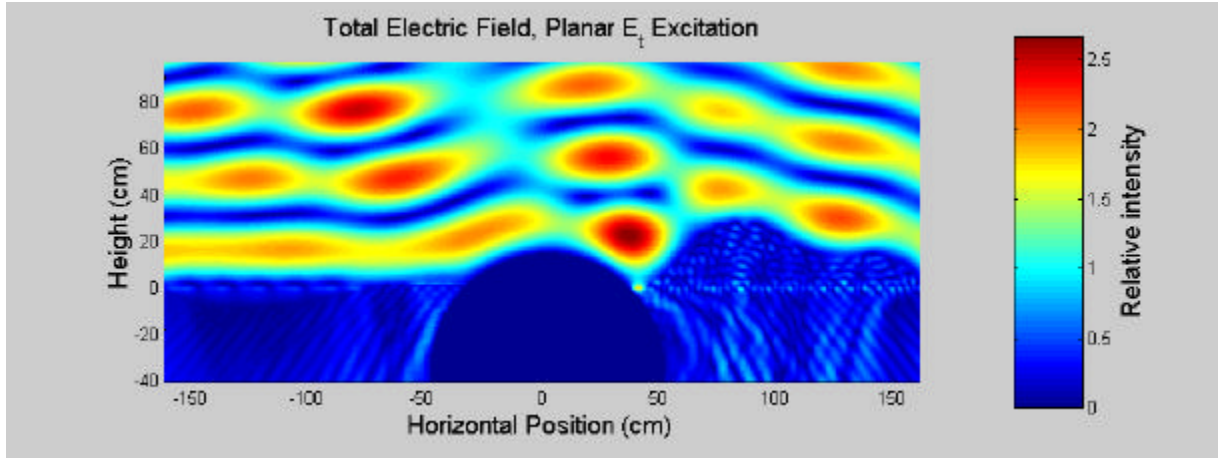


Figure 7 Total electric field at 500 MHz, for horizontally polarized E-field (into plane of figure), for geometry of Figure 6.

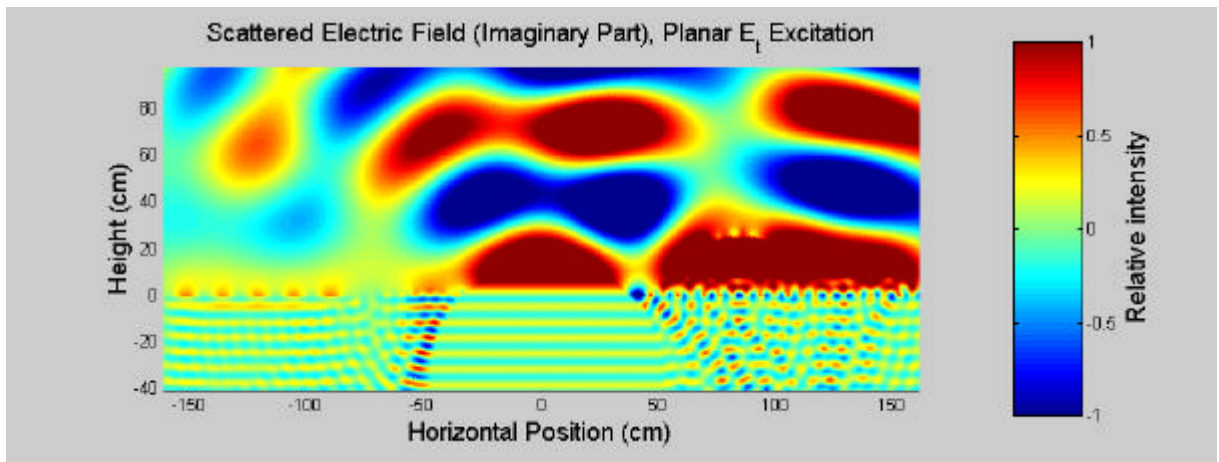


Figure 8 Electric field at 500 MHz due to just the mine and the surface roughness, for horizontally polarized E-field (into plane of figure), for geometry of Figure 6.

Conclusions

Although seawater is relatively impenetrable to RF waves, and causes strong reflections from its random rough surface, computational modeling has shown that it is still possible to detect the stronger and more organized scattering from the top of a floating sea mine. By considering additional information of multiple views of the same metal target with a moving water surface, it is possible to distinguish the target from the background. The scattered wave features of the circular metal target appear to be relatively independent of depth, and thus provide a detectable invariant in the highly cluttered background. There appears to be great potential in the use of radar to find floating sea mines.

Acknowledgements

The author gratefully acknowledges support for this work from the Army Research Office, MURI grant # DAA 0-55-97-0013, 0013 and by CenSSIS, the Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation, award number EEC-9986821.

References

1. Rappaport, C. and Morgenthaler, A., "FDFD Modeling of Plane Wave Interactions with Buried Objects Under Rough Surfaces," **2001 USNC/URSI Radio Sci. Meeting**, Boston, MA, June 2001, pp. 318.
2. Morgenthaler, A. and Rappaport, C., "Scattering from Lossy Dielectric Objects Buried Beneath Randomly Rough Ground: Validating the Semi-Analytic Mode Matching Algorithm with Two-Dimensional FDFD," **IEEE Trans. Geosci. Remote Sens.**, Nov. 2001, pp. 2421-2428.
3. Berenger, J., "A perfectly matched layer for the absorption of electromagnetics waves", **Journal of Computational Physics**, 114, 1, pp. 185-200, 1994.
4. Rappaport, C., " Interpreting and improving the PML absorbing boundary condition using anisotropic lossy mapping of space", **IEEE Trans. on Magnetics**, vol. 32, 3, pp. 968-974, 1996.
5. Rappaport, C., Kilmer, M., and Miller, E., "Accuracy Considerations in Using the PML ABC with FDFD Helmholtz Equation Computation", (Invited paper) **Int'l J. of Numerical Modeling**, vol. 13, no. 471, Sept. 2000, pp. 471—482.
6. Winton, S., and Rappaport, C. "Specifying the PML Conductivities by Considering Numerical Reflection Dependencies," **IEEE Trans. on Antennas and Propagation**, vol. 48, Sept. 2000, pp. 1005—1063.
7. Dumanian, A., Rappaport, C., and Morgenthaler, A., "A Time Domain Signature Investigation for the GPR Detection of Plastic Land Mines Buried in Soils," **2001 USNC/URSI Radio Sci. Meeting**, Boston, MA, June 2001, pp. 248.

Electromagnetic Scattering Interference Between Two Shallow Objects Buried Under 2-D Random Rough Surfaces

Magda El-Shenawee¹ and Carey Rappaport²

¹Department of Electrical Engineering
University of Arkansas
Fayetteville, AR 72701
magda@uark.edu

²Department of Electrical and Computer Engineering
Northeastern University
Boston, MA
rappaport@neu.edu

Abstract

A rigorous electromagnetic model has been developed to analyze the scattering from two dielectric shallow objects buried under the 2-D random rough ground as a means of predicting false alarms due to a clutter object near a target object. The Method of Moments (MoM) accelerated by the Steepest Descent Fast Multipole Method (SDFMM) is used to compute the unknown electric and magnetic surface currents on both the rough ground surface and the two buried objects. The roughness parameters strongly influence the scattered interference of the two objects, especially when their separation distance is several correlation lengths. This could increase the probability of false alarms.

I. INTRODUCTION

In realistic minefields, buried anti-personnel (AP) nonmetallic mines are often closely accompanied by underground clutter-objects. The presence of this object considerably obscures the targets causing a false alarm during the detection process. The separation distance between the AP-mine and the clutter-object plays a primary role on the probability of false alarms. A rigorous electromagnetic model has been developed to analyze the scattering mechanism of two dielectric objects buried beneath a rough ground surface as reported in [1]. Using the $O(N)$ fast

algorithm, the Steepest Descent Fast Multipole Method (SDFMM) [2-3] tremendously accelerates the computations of the N unknown surface currents [1,4]. When the two objects were located close to one another under flat ground, the strong scattering interference generates a false response appearing to be a third buried object [1]. The dependency of the observed scattering interference on the ground roughness parameters is investigated in this work.

II. FORMULATION

The rigorous electromagnetic model derived in [1] is employed in this work where six integral equations are used to obtain the equivalent surface currents on dielectric scatterers shown in Fig. 1 (3-D scattering problem). Four different regions are involved in this scattering problem; air, soil, first object and second object. The unknown electric and magnetic currents on the ground surface, on the target surface, and on the clutter object surface are approximated using the well-known RWG vector basis functions [5]. After some algebraic manipulations, the linear system of equations is obtained as: $\bar{\bar{Z}} \bar{I} = \bar{V}$, where the total impedance matrix $\bar{\bar{Z}}$ has order $2(N_1 + N_2 + N_3) \times 2(N_1 + N_2 + N_3)$. The number of electric and magnetic current unknowns (edges) on the ground, on the target and on the second object are $2N_1$, $2N_2$ and $2N_3$, respectively. The tested tangential incident electric field \bar{E}^{inc} and the tested normalized magnetic field $\mathbf{h}_1 \bar{H}^{inc}$ on the exterior of the ground surface are expressed in \bar{V} . The SDFMM is implemented to significantly accelerate solving the linear system of equations for the unknown current coefficients [1-4].

II. NUMERICAL RESULTS

In this Section, we investigate the scattering interference between the two buried objects as a function of their separation distances and the ground roughness. Several values for the root mean square height σ and the correlation length l_c are considered with emphasis on small roughness

parameters for the AP-mine detection application. The incident wave is assumed to be a Gaussian beam at normal incidence that is carefully tapered to minimize edge effects [6]. The two objects are oblate spheroids with dimensions $a = 0.3\lambda_0$, $b = 0.15\lambda_0$ and at depth $z = -0.4I_o$ measured from their centers. The relative dielectric constant of the ground soil is assumed to be $\epsilon_r = 2.5 - j0.18$ and for both objects is assumed to be $\epsilon_r = 2.9 - j0.072$. The $8I_o \times 8I_o$ ground surface is discretized into 60,000 electric and magnetic surface current unknowns. Each object is discretized into 600 electric and magnetic surface current unknowns. In order to analyze the object signatures, the electric fields scattered from each rough ground are removed by subtraction similar to our work in [4].

In Fig. 2, the object signatures are plotted across the diagonal as shown in Fig.1. When the two objects are separated by $S = 1.4\lambda_0$, the results show three peaks of almost equal magnitudes; the first peak is above the first object, the second peak is above the second object and the third peak is at mid-point between them. This third peak is due to the strong constructive interference between the two objects. This phenomenon could easily cause a false alarm during the detection process. However, when the separation distance increases, the mid-point peak is dissolved into several secondary peaks as observed when $S = 2.1\lambda_0$ and $2.8\lambda_0$. These secondary peaks become insignificant when S is increased to $4.2\lambda_0$. As noticed in this figure, the asymmetry around the mid-point ($x = 4.0\lambda_0$) is clearly caused by the random roughness of the ground. Moreover, the magnitudes of all peaks decrease with increasing separation distance because the objects are further from the beam footprint center (ground center in this work). In Fig. 3, the separation distance is $S = 1.4\lambda_0$, σ is $0.1\lambda_0$ and the range for l_c is $0.4\lambda_0 - 1.0\lambda_0$. The results clearly show the influence of the roughness parameters, however, the observed strong interference (mid-point peak) is slightly affected by the ground roughness in this case. Similar results were obtained

upon varying σ from $0.04\lambda_0$ to $0.1\lambda_0$ with $l_c=0.5\lambda_0$. This study is repeated for $S = 2.1\lambda_0$ as shown in Fig. 4 and for $S = 2.8\lambda_0$ as shown in Figs. 5 and 6. It is interesting to notice that when the separation distance becomes several correlation lengths, as shown in Figs. 4 and 6, stronger scattered interference is observed. This is demonstrated by the increase in the number of secondary peaks from three in the flat ground case to four and five in the rough ground case as shown in Figs. 4 and 6, respectively. This mechanism could increase the possibility of false alarms.

III. CONCLUSIONS

False alarms could easily occur because of the interference mechanism between the target and a second nearby object buried under the ground. The roughness parameters strongly influence the scattering interference mechanism of the two objects, which could increase the probability of false alarms.

ACKNOWLEDGMENTS

This research was sponsored by the Northeastern University's Demining MURI grant # DAA 0-55-97-0013 and in part by the College of Engineering at the University of Arkansas. The SDFMM was originally developed by V. Jandhyala, E. Michielssen and W. Chew at the UIUC.

References

- [1] M. El-Shenawee, "The Steepest Descent Fast Multipole Method for Scattering from Two Penetrable Shallow Objects Buried Under Two-Dimensional Randomly Rough Surface," submitted to the *IEEE Trans. Anten. & Propag.*, in review.
- [2] V. Jandhyala, *Fast Multilevel Algorithms for the Efficient Electromagnetic Analysis of Quasi-Planar Structures*, Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1998.
- [3] V. Jandhyala, E. Michielssen, B. Shanker and W.C. Chew, "A combined steepest descent-fast multipole algorithm for the fast analysis of three-dimensional scattering by rough surfaces," *IEEE Trans. Geosci. & Remote Sensing*, vol. 36, no. 3, pp. 738-748, May 1998.
- [4] M. El-Shenawee, C. Rappaport and M. Silevitch, "Monte Carlo simulations of electromagnetic wave scattering from random rough surface with 3-D penetrable buried object: mine detection application using the SDFMM," *J. Optical Society of America A*, to appear in December 2001
- [5] S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape," *IEEE Trans. on Anten. & Prop.*, vol. AP-30, no. 3, pp.409-418, May 1982.
- [6] P. Tran and A. A. Maradudin, "Scattering of a scalar beam from a two-dimensional randomly rough hard wall: enhanced backscatter," *Phys. Rev. B*, vol. 45, no. 7, pp. 3936-3939, February 1992.

List of Figures

Fig. 1. (a) Cross section along the diagonal direction for two objects buried under the rough ground, (b) Top view.

Fig. 2. Scattered electric near-field of just the two objects at $z = 0.5I_o$ when $S/\lambda_o = 1.4-4.2$ and ground roughness parameters are $\sigma/\lambda_o = 0.1$ and $l_c/\lambda_o = 0.5$. The separation distance S is shown in Fig. 1b.

Fig. 3. Scattered electric near-field of just the two objects at $z = 0.5I_o$ when $S/\lambda_o = 1.4$ and ground roughness parameters are $\sigma/\lambda_o = 0.1$ and $l_c/\lambda_o = 0.4-1.0$.

Fig. 4. Scattered electric near-field of just the two objects at $z = 0.5I_o$ when $S/\lambda_o = 2.1$ and ground roughness parameters are $\sigma/\lambda_o = 0.1$ and $l_c/\lambda_o = 0.4-1.0$.

Fig. 5. Scattered electric near-field of just the two objects at $z = 0.5I_o$ when $S/\lambda_o = 2.8$ and ground roughness parameters are $l_c/\lambda_o = 0.5$ and $\sigma/\lambda_o = 0.04-0.1$

Fig. 6. Scattered electric near-field of just the two objects at $z = 0.5I_o$ when $S/\lambda_o = 2.8$ and ground roughness parameters are $\sigma/\lambda_o = 0.1$ and $l_c/\lambda_o = 0.4-1.0$.

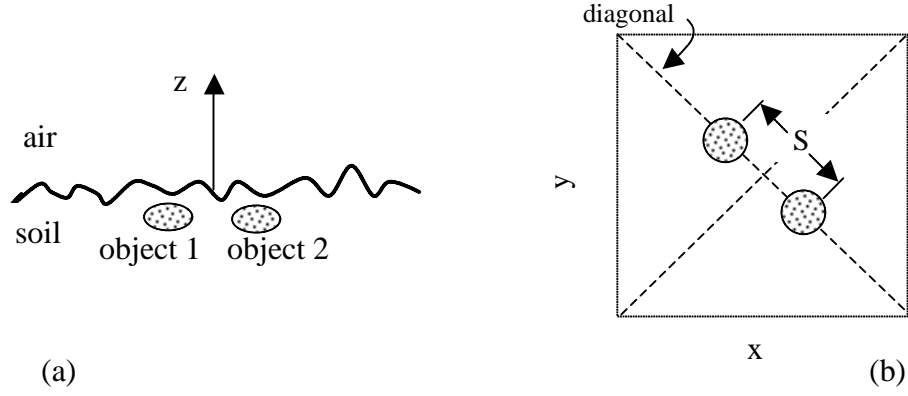


Fig. 1. (a) Cross section along the diagonal direction for two objects buried under rough ground, (b) Top view.

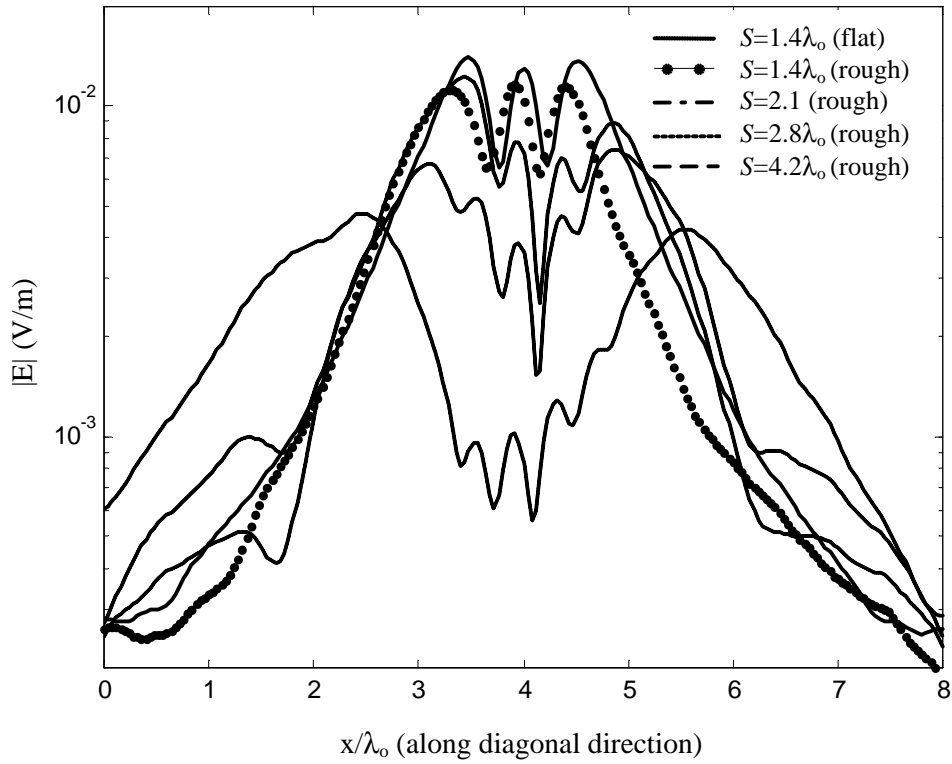


Fig. 2. Scattered electric near-field of just the two objects at $z = 0.5l_o$ when $S/\lambda_o = 1.4-4.2$ and ground roughness parameters are $\sigma/\lambda_o = 0.1$ and $l_c/\lambda_o = 0.5$. The separation distance S is shown in Fig. 1b.

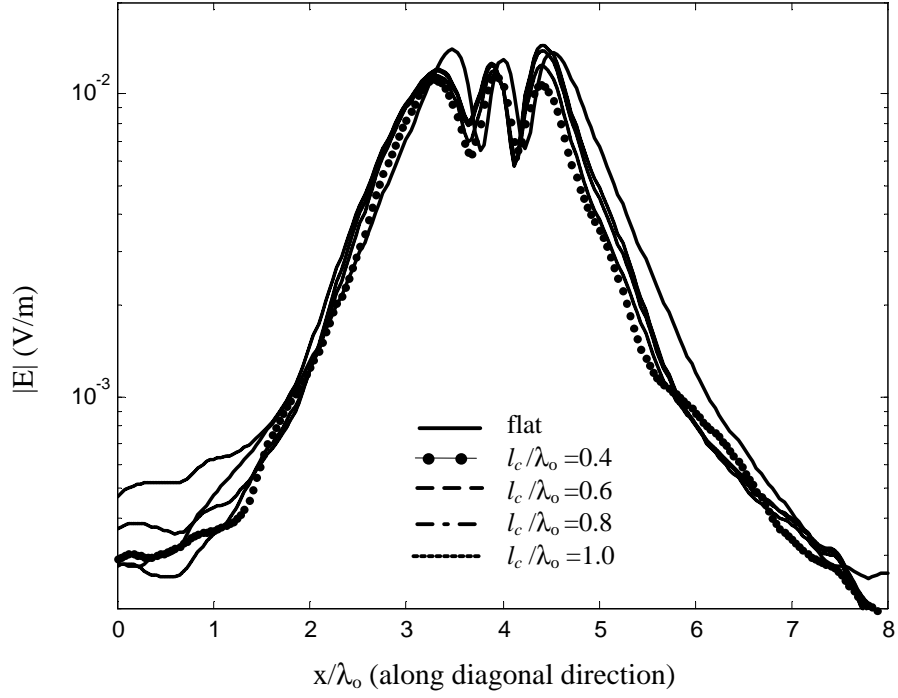


Fig. 3. Scattered electric near-field of just the two objects at $z=0.5I_o$ when $S/\lambda_o=1.4$ and ground roughness parameters are $\sigma/\lambda_o=0.1$ and $l_c/\lambda_o=0.4-1.0$.

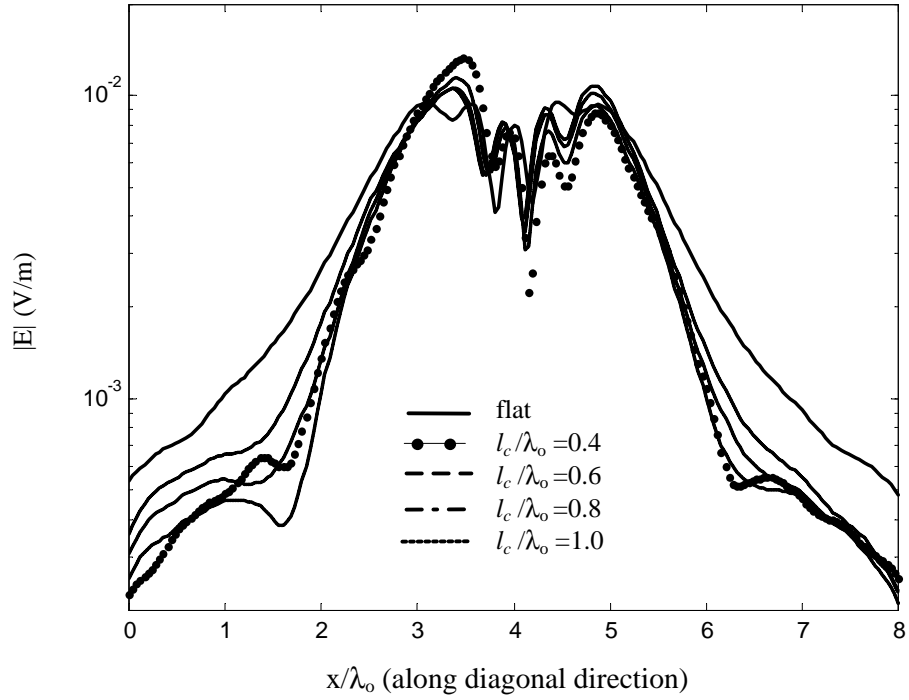


Fig. 4. Scattered electric near-field of just the two objects at $z=0.5I_o$ when $S/\lambda_o=2.1$ and ground roughness parameters are $\sigma/\lambda_o=0.1$ and $l_c/\lambda_o=0.4-1.0$.

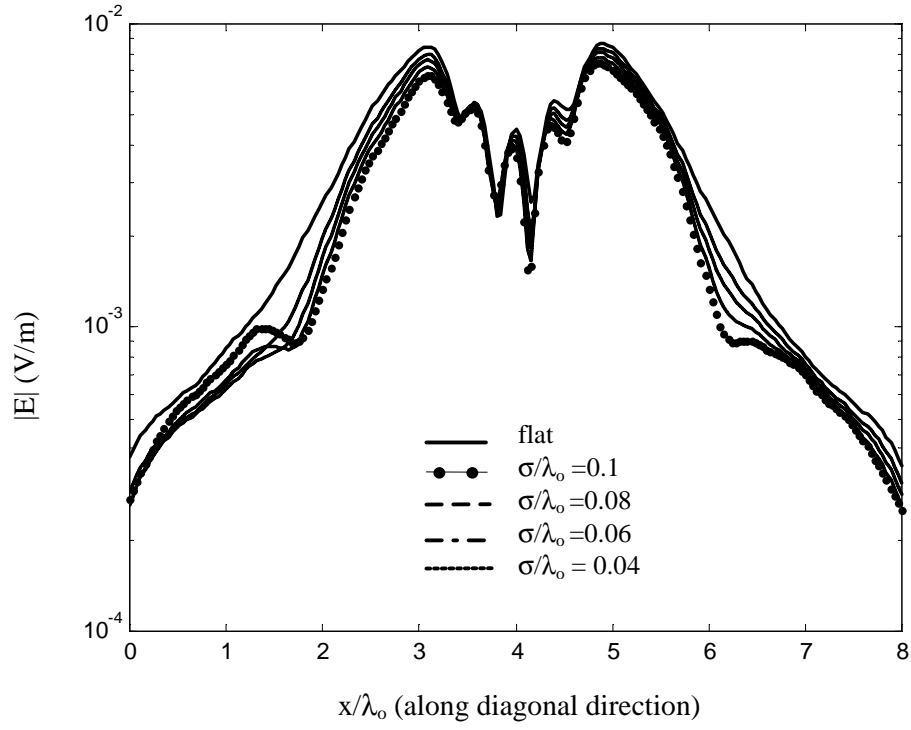


Fig. 5. Scattered electric near-field of just the two objects at $z = 0.5I_o$ when $S/\lambda_o = 2.8$ and ground roughness parameters are $l_c/\lambda_o = 0.5$ and $\sigma/\lambda_o = 0.04-0.1$

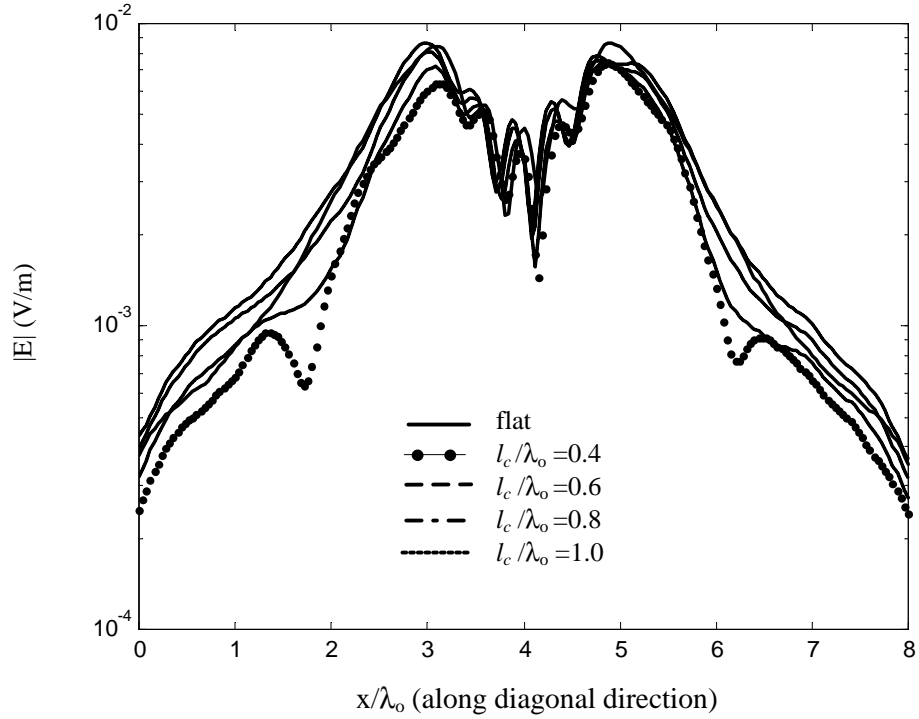


Fig. 6. Scattered electric near-field of just the two objects at $z = 0.5I_o$ when $S/\lambda_o = 2.8$ and ground roughness parameters are $\sigma/\lambda_o = 0.1$ and $l_c/\lambda_o = 0.4-1.0$.

Statistical Method to Detect Subsurface Objects Using Array Ground Penetrating Radar Data

Xiaoyin Xu, Eric L. Miller, Carey M. Rappaport and Gary D. Sower

Abstract

We introduce a combination of high-dimensional analysis of variance (HANOVA) and sequential probability ratio test (SPRT) to detect buried objects from an array ground penetrating radar (GPR) surveying a region of interest in a progressive manner. Using HANOVA, we exploit the transient characteristic of GPR signals in the time domain to extract information about buried objects at fixed positions of the array. Based on the output of the HANOVA, the SPRT is employed to make detection decisions recursively as the array moves downtrack. The method is on-line implementable and of low computational complexity. Our approach is validated using field-data from a landmine detection application.

Index Terms

Analysis of variance (ANOVA), GPR mine detection, array signal processing, sequential detection, transient signal analysis.

I. INTRODUCTION

Ground penetrating radar (GPR) is widely used in detecting subsurface objects such as buried landmines, unexploded ordnance, and utility lines [1]. Compared with other subsurface sensing technologies, GPR has a few advantages. First, it is sensitive to changes in all three electromagnetic characteristics of a media, electric permittivity, electric conductivity, and magnetic permeability. Thus GPR is capable of detecting both metallic and non-metallic

X. Xu, E. L. Miller, and C. M. Rappaport are with the Center for Subsurface Sensing and Imaging Systems, Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA. G. D. Sower is with the EG&G MSI Inc., 2450 Alamo Ave. S.E., Albuquerque, NM 87106. This work was supported by an ARO MURI on Demining under Grant DAAG55-97-1-0013

objects. Second, unlike sensors that can only survey an area directly beneath them, GPR can survey an area in front of it [2, 3]. Therefore a GPR system can be used to detect dangerous objects before the system moves over and past them. This can be important for operations such as buried landmine detection and unexploded ordnance remediation.

A typical GPR transmitter/receiver configuration is shown in Fig. 1(a). The system consists of one transmitter and one receiver. The transmitter emits a short pulse of electromagnetic energy and the receiver collects the echo for a certain time period. The exact type of the transmitter and receiver, shape of the electromagnetic pulse, and system setup depend on the specific application of the GPR [1, 4–6]. To improve performance and efficiency, a GPR array is usually employed to sweep a large area in a relatively short time. Fig. 1(b) shows a typical GPR array moving in the x -direction. At every stop of the array, the GPR array operates in the following sequence: 1) the first transmitter radiates a pulse into the ground and then turns off, 2) the first receiver turns on to collect reflected signal, 3) the first receiver turns off after a short time, usually 10 to 20 ns. The above process repeats from every pair of transmitters and receivers and then the GPR array moves to next position. Based on the echoes, the processing objective is to determine if an object is present in the GPR's field of view.

The inherent near-field nature of the GPR detection problem coupled with the fact that the objects of interest are embedded in an inhomogeneous halfspace with a typically rough interface present some significant challenges in the area of GPR signal processing. Indeed assuming one has detailed knowledge of the air-earth interface as well as the electrical properties of the subsurface, just modeling the received signal using, for example, a three dimensional finite difference time domain code, is a daunting task [4, 7]. The use of such a forward model in any form of on-line processing routine where one might need to account

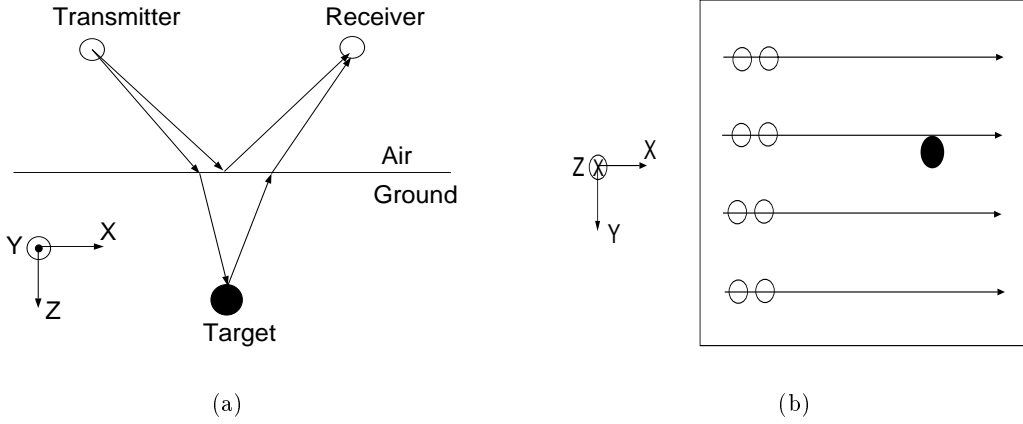


Fig. 1. Setup, a) a single GPR system, b) a GPR array (plane view).

for e.g. unknown ground structure is clearly infeasible at the current time.

Thus, here we consider detection methods which are less computationally demanding with an eye toward approaches that could be used in real-world scenarios. Our interests are in techniques possessing three important characteristics. First, to reflect the manner in which GPR data are acquired and the nature of the GPR mission, the algorithms should be causal in that they need only the data at the current and previous sensor position to determine whether an object is present in the field of view of the sensor. Second, they should be of low complexity. Preferably the number of calculations would grow linearly with the size of the data set. Finally, the processing schemes should be robust to uncertainties in the GPR environment and hence the particular detailed structure of the received signals.

Current signal processing methods with some or all of these characteristics fall into one of three categories. First, pattern matching methods [8] employ techniques such as fuzzy set theory and neural networks. Such methods can be fast but also require extensive training to function well. Second, image-then-detect techniques [9] employ a beamforming or backpropagation approach to build an image of the subsurface which is then post-processed to detect objects. These approaches generally require the data from the full GPR scan to

form an image and are thus not well suited to on-line computations in which information is processed sequentially as the array proceeds down-track. Finally, there has been much work done in statistical signal processing, where one can employ statistical tools to detect objects and examine quantities such as probability of detection and probability of false-alarm [10].

Here we consider a statistical, transient detection approach. By “transient” we mean that the signals of interest are manifest in the GPR data for a small number of sensor positions and for relatively few samples in any received waveform. For example, in Fig. 2 we plot raw observations obtained by one T/R pair from an EG&G GPR system [11], over an M20 metal mine. Each column of this image is a time-series of observations for a given stop of the array. It is seen that the received GPR signal is transient in two ways. First, for each time-series (i.e. for each column of the image) containing an object signal, the signal appears only in a brief window, roughly from samples 300 to 700. The reason is that the object signal always comes after the signal arising from the bounce off of the air-ground interface and attenuates quickly in lossy media. Second, the object signal shows up only at a few down-track positions of the GPR array, specifically locations 15 through 25. In both cases, the appearance of object signal changes the mean value of the data. Our method for object detection then is based on detecting change in this mean first in the cross-track direction and then in the down-track direction.

More specifically our approach consists of two parts. First, at each down-track position of the array, we process the data among all T/R pairs to generate one test statistic. We use high-dimensional analysis of variance (HANOVA) to test whether the data consists of reflected signal from a buried object. The HANOVA is a generalized version of standard analysis of variance (ANOVA), which is a method for testing hypothesis about means of random vectors [12, 13]. Second, a sequential probability ratio test (SPRT) is applied to

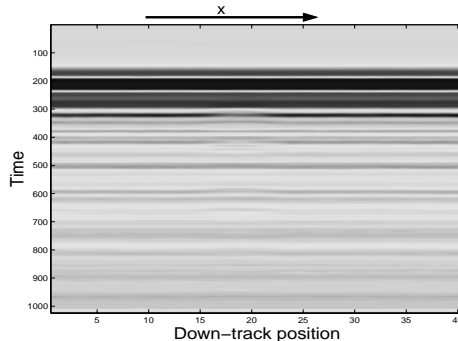


Fig. 2. Observation from one T/R pair, for a metal mine M20 buried at about position 20. Unit in down-track position is about 7.6 cm. Unit in time axes is 0.02 ns.

process the statistic of the HANOVA as the array moves down-track. The SPRT is a recursive statistical hypothesis testing technique that provides early indication of the onset of changes in a time series. The output of the SPRT is compared with a threshold. If it exceeds the threshold, a detection is declared, otherwise, the GPR array moves one more step down-track and new data are collected and processed in the above manner [14].

As explained in greater detail below, our approach does in fact satisfy the three requirements we discussed previously. It is causal and has computational complexity that grows linearly with the size of the data. Moreover, we show through real-data examples that it is robust, requiring little in the way of training and able to successfully address the object detection problem for a number of GPR systems operating in a wide range of environments. We do stress here that the algorithm in this paper is intended *only* to find anomalies beneath the GPR array and *not* to solve the far more challenging classification problem. Thus, from a practical perspective our approach will serve well as an efficient “pre-screener” in a larger automatic target detection algorithm suite. Finally, our method is motivated by landmine detection using GPR, however it can also be used in other detection application, such as laser-induced acoustic subsurface objects detection [15].

The paper is organized as follows. Section II discusses the problem formulation and our

method. Section III gives some examples of using the method in different situations. Field data from different radar configurations and test sites are used to show how the algorithm works. Conclusion and direction of future work are given in Section IV.

II. PROBLEM FORMULATION AND ALGORITHM

To begin, we consider a single GPR T/R pairs as shown in Fig. 1(a). After each transmission, the receiver collects an echo for a certain amount of time. Depending on the presence of an object, there are either two or three components in the echo. One is measurement noise, assumed to be white and Gaussian. Another is background, i.e., “nominal” signal observed in object-free regions. The third component is object signal, reflection from a buried object.

For the GPR array shown in Fig. 1(b), assume we have M GPR Transmitter/Receiver (T/R) pairs surveying an area in N steps, the task is to use present and previous array measurement to detect buried mines as the array moves down-track. At each down-track position, we model the array detection problem in a typical hypotheses testing framework [14],

$$H_0 : \quad \text{there is no object,}$$

$$H_1 : \quad \text{there is an object.}$$

The null hypothesis H_0 means that there is no buried object in the field of view of the GPR array, so the total received signal is comprised of nominal background and measurement noise. By nominal background, we mean any portion of the received waveform not sensor noise and not arising from the interaction of the transmitted pulse with the object. Reflection from the air-ground interface is the dominant component of this part of the signal. The alternative hypothesis H_1 indicates that there is buried object so that the received signal consists of nominal background, measurement noise, and an object signal.

In this paper we assume that the nominal background signal has been removed via a

preprocessing stage. The most used background removal methods include casual methods, such as subtraction of a moving average from the observation [16], and non-causal methods, such as subtraction of an ensemble average from the observation [17, 18]. Causal methods use data from previous and present collection, non-causal methods use data from previous and future collection. In this paper, a moving average (MA) filter is used to eliminate the nominal background.

In practice, the receiver collects time-samples of the reflection and stores it as a vector. For convenience, we use vector notation in our discussion, i.e., $\mathbf{y}(m, n)$ is a column vector representing observation of the m th T/R pair at the n th down-track position. The length of $\mathbf{y}(m, n)$ is K , the number of samples in time. Fig. 3 shows the received signal after the nominal background removal.¹ We then have the hypothesis test

$$\begin{aligned} H_0 : \quad \mathbf{y}(m, n) &= \mathbf{v} \\ H_1 : \quad \mathbf{y}(m, n) &= \mathbf{s}(m, n) + \mathbf{v} \end{aligned} \tag{1}$$

where $m = 1, \dots, M$, $n = 1, \dots, N$ are positions of GPR, $\mathbf{s}(m, n)$ is the assumed signal due to presence of buried object, \mathbf{v} is assumed to be a white Gaussian noise with a zero mean, and covariance matrix $\sigma_v^2 \mathbf{I}$, where \mathbf{I} is the identity matrix of size K and independent of (m, n) .

The statistical assumptions about \mathbf{v} are not strictly accurate in describing the noise in a GPR signal. For example the background removal process will not be perfect leaving a component of correlated “clutter” in the data which may or may not possess Gaussian statistics. Despite the mismatch, the use of the additive white Gaussian noise model is useful for a number of reasons. This model allows us to develop an algorithm for object

¹ For the purpose of illustration, in this section we use field data from a buried metal mine to illustrate clearly the concept under consideration. Examples which demonstrate better the utility of our approach on more challenging problems, including buried plastic mines, are given in Section III.

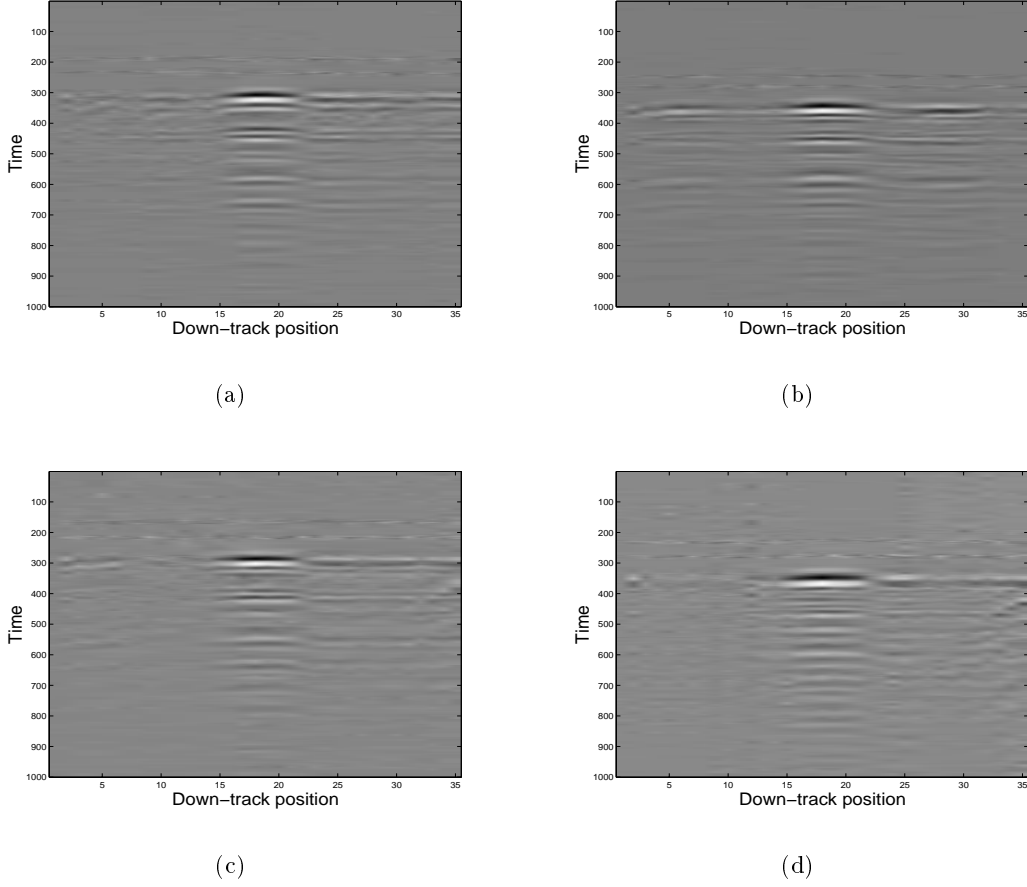


Fig. 3. Signals from four T/R pairs, after background removal, a) pair 1, b) pair 2, c) pair 3, d) pair 4. Unit in time axes is 0.02 ns.

detection which is firmly rooted in Gaussian-based statistical decision theory and which can be generalized in the future for more complex noise processes. Moreover, the complexity of such algorithms is quite low making them well suited for real-world implementation. Finally, test results in Section III from real field data demonstrate that the method is quite effective in detecting objects. Thus, the Gaussian noise model is shown to work in practice. While it may be interesting to explore other, more accurate models for the sensor noise to determine for example what can be gained in terms of performance and what would be lost in terms of computational complexity, such an effort is beyond the scope of the work in this paper.

Based on the previous discussion, after background removal the hypothesis test in (1) may be written as

$$\begin{aligned} H_0 : \quad & \mathbf{y}(m, n) \sim N(\mathbf{0}, \sigma_v^2 \mathbf{I}) \\ H_1 : \quad & \mathbf{y}(m, n) \sim N(\mathbf{s}(m, n), \sigma_v^2 \mathbf{I}), \quad m = 1, \dots, M, \quad n = 1, \dots, N \end{aligned} \quad (2)$$

where the notation $\mathbf{y} \sim N(\bar{\mathbf{x}}, \mathbf{R})$ indicates that \mathbf{y} is distributed as a Gaussian random vector with mean $\bar{\mathbf{x}}$ and covariance matrix \mathbf{R} .

As stated in the Introduction, we take a two-step approach to the processing of $\mathbf{y}(m, n)$. First for each n we use the HANOVA procedure to generate a single test statistic, $Y(n)$, from the data from all T/R pairs. Second, a recursive, sequential detection scheme is employed to process $Y(n)$ as we proceed down track in order to determine where objects are present.

A. Cross-track Processing

We begin by discussing the use of HANOVA to process data in the cross-track direction. HANOVA is a generalized version of analysis of variance (ANOVA). ANOVA is a body of methods to analyze the data with a view to test hypotheses about the effects of one or more factors [19]. To review the basics of ANOVA, we follow the notation established above for the GPR problem and for simplicity assume we have one data vector of size $K \times 1$ from a single T/R pair, $\mathbf{y} \sim N(\mathbf{s}, \sigma^2 \mathbf{I})$ and we wish to test $H_0 : \mathbf{s} = \mathbf{0}$ (i.e., no object) vs. $H_1 : \mathbf{s} \neq \mathbf{0}$ (i.e., an object present)². Standard ANOVA is essentially an “energy detection” scheme [12] where we estimate \mathbf{s} by \mathbf{y} , generate the test statistic $Y = \|\mathbf{y}\|^2$, and compare Y to a threshold, γ . If Y exceeds the threshold, H_1 is chosen, else H_0 is selected. The probability of detection of the standard ANOVA is

$$P_d(H_1|H_1) = Q\left(\frac{\gamma - \frac{\|\mathbf{s}\|^2}{\sigma^2 \sqrt{2K}}}{\sqrt{1 + \frac{2\|\mathbf{s}\|^2}{\sigma^2 K}}}\right) \quad (3)$$

² For notational simplicity, we drop the explicit dependence of all quantities on m and n in this discussion

where γ is the test threshold decided by setting an acceptable probability of false-alarm under H_0 and Q is the complementary cumulative distribution function and is strictly decreasing [20]

$$Q(Y) = \int_{\gamma}^{\infty} P_1(Y) dY. \quad (4)$$

Recently, Fan [12] and Fan and Lin [13], have noted that the performance of ANOVA suffers for problems when the signal of interest is limited to a window of the observation vector. The reason is that a full dimensional test loses its power due to accumulation of stochastic noise. To see why, suppose \mathbf{s} is different from $\mathbf{0}$ only for say the first k_0 samples of the full observation vector. Then on average as $K > k_0$ goes large, $\sum_{k=1}^K [\mathbf{s}]_k^2 / \sigma^2 \sqrt{2K}$ decreases due to the accumulation of zero mean noise samples and the term within the parenthesis of (3) increases, thus reducing P_d . Therefore, for higher probability of detection, we would like to confine the test on a window mostly containing the signal of the observation vector. The window we choose is a box window \mathbf{w} , defined as

$$[\mathbf{w}]_k = \begin{cases} 1, & k = k_1, \dots, k_2 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where $1 \leq k_1 < k_2 \leq K$. The k_1 and k_2 are chosen in a preset manner, as discussed later in this section. Multiplying each element in \mathbf{y} by the corresponding element of \mathbf{w} gives the windowed \mathbf{y}_w

$$[\mathbf{y}_w]_k = [\mathbf{y}]_k \times [\mathbf{w}]_k, \quad k = 1, \dots, K. \quad (6)$$

To demonstrate the utility of HANOVA, we test the time-series shown in Fig. 4(a). We choose to test the vector at its full dimension k_1 in (5) is 1 and $k_2 = 1000$, and two windowed sub-dimensions (each containing fewer and fewer noise components) $k_1 = 100$ and $k_2 = 900$,

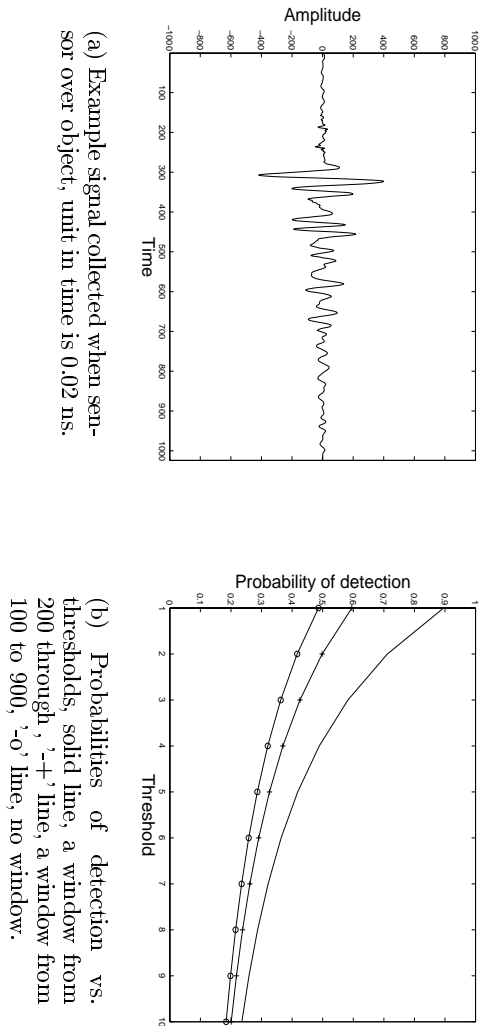


Fig. 4. Motivating example for use of HANOVA rather than ANOVA. (a) displays sample data collected by a GPR when over an object. The transient nature of the relevant portion of the signal is clear. Using this signal, in (b) we display the decreasing detection rates associated with including increasing numbers of “noise” samples in the processing.

and $k_1 = 200$ and $k_2 = 800$. From Fig. 4(b), it is seen that by setting the window properly, higher probability of detection is gained at different levels of detection thresholds, γ in (3). It demonstrates that when signal is not “full-dimensional”, looking for a window of signal-rich sub-dimensions to test will increase the probability of detection.

When the observation is a sequence of high-dimensional vectors whose components are mostly noise, as is the case for our GPR problem, it is desirable to adaptively choose the window to maximize the probability of detection. This kind of method of reducing a full dimensional test to a windowed version is called HANOVA [13]. Fan’s original work was limited to problems in which the first k_0 dimensions are believed to be signal-rich and used in HANOVA, with k_0 found from the data. Here we consider a generalization of Fan’s work to take into account the fact that for the GPR problem the transient signal is significant over a window not generally starting with the first dimension but in the middle of the observation vector. Moreover, this window will vary with (m, n) .

To choose this window we note that (3) indicates that the probability of detection achieves its maximum value when the term inside the parenthesis is minimized. Equally, one wants to maximize the quantity

$$\arg \max_{k_1, k_2} \frac{\sum_{k=k_1}^{k_2} [\mathbf{s}(m, n)]_k^2}{\sigma_v^2 \sqrt{k_2 - k_1 + 1}} - \sqrt{k_2 - k_1 + 1} \quad (7)$$

where $k_1 < k_2$ and $k_1, k_2 \in 1, \dots, K$. The difficulty for us is that in general, the precise structure of \mathbf{s} is not known. Hence, we use the data to form an estimate of \mathbf{s} as follows. Assume we are at the n th stop, then we estimate \mathbf{s} by the mean value of the previous l vectors

$$\hat{\mathbf{y}}(m, n) = \begin{cases} \frac{\sum_{j=1}^n \mathbf{y}(m, j)}{j}, & n = 1, \dots, l \\ \frac{\sum_{j=n-l}^n \mathbf{y}(m, j)}{l}, & n > l \end{cases} \quad (8)$$

where $m = 1, \dots, M$ and the corresponding window $\mathbf{w}(m, n)$ is decided based on $\hat{\mathbf{y}}(m, n)$ as k_1, k_2 are defined by (7). More will be said about choosing a proper l in Appendix A.

Rather than looking for the optimal window by searching over all k_1 - k_2 pairs, we pursue a suboptimal, but more efficient two-stage approach. First, we fix k_1 as 1, incrementally increase k_2 , and stop when (7) is maximized. Thus we determine the end point of the window k_2 . Starting from k_2 , working backward toward the first point, we similarly determine the starting point of the window, k_1 . Both searching steps can be computed in linear complexity, it takes $o(K)$ steps to find the k_2 and $o(k_2)$ steps to find the k_1 ³. In summary the steps for looking for windows at the n th stop of the GPR array are given in Fig. 5.

Having determined the window at the position (m, n) , the next stage of processing is to generate a single detection statistic at stop n . Here we generalize HANOVA to multiple vector observations, via

$$Y(n) = \frac{1}{\sigma_v^2} \sum_{m=1}^M \|\mathbf{y}_w(m, n)\|^2. \quad (10)$$

³ The notation $o(K)$ means that the computational complexity grows slower than or equally fast as K increases.

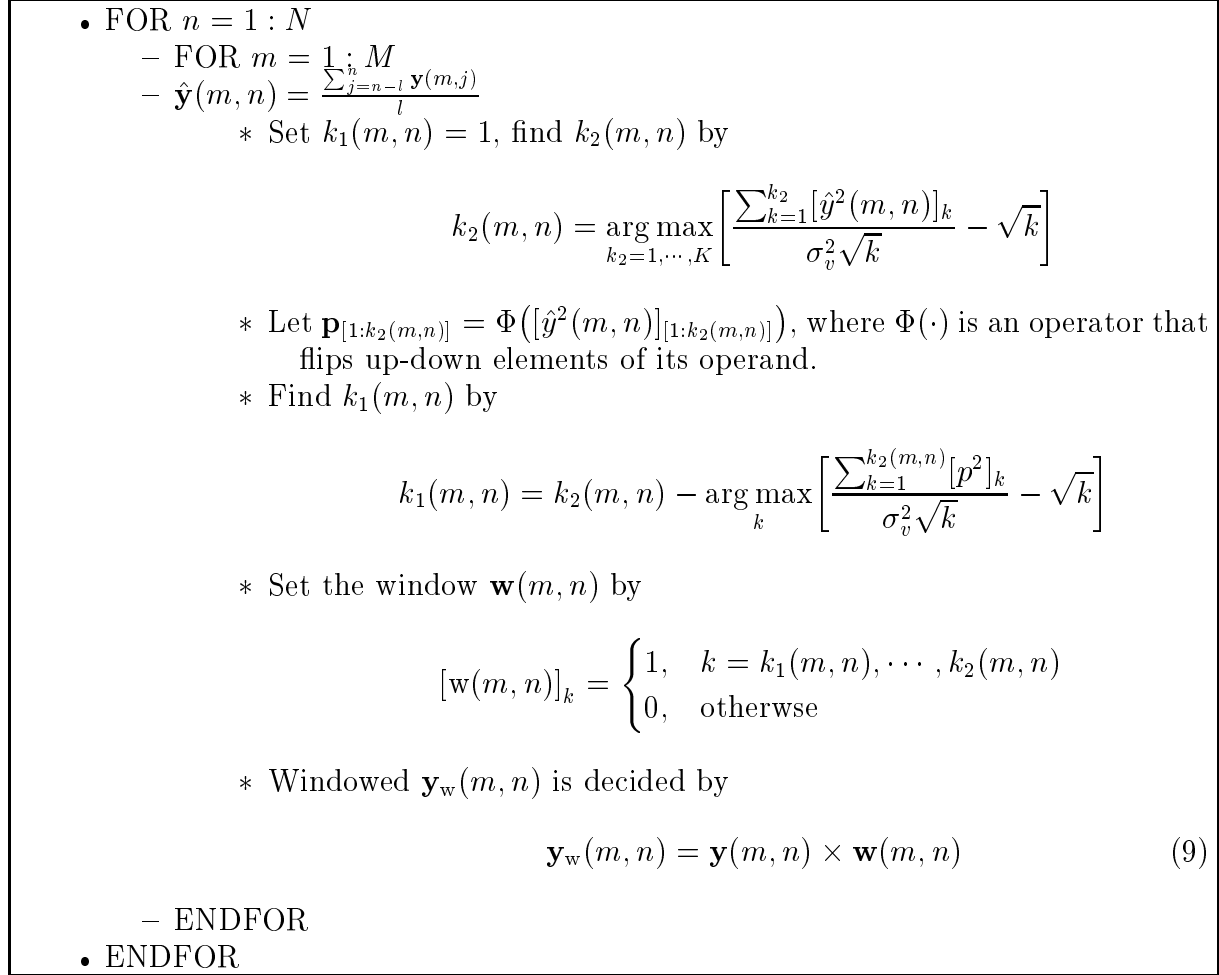


Fig. 5. Steps of deciding window $\mathbf{w}(m, n)$ and $\mathbf{y}_w(m, n)$.

Note $\mathbf{y}_w(m, n)$ can be of different length because of different window applied. Fig. 6(a) shows the result of applying HANOVA to the data in Figure 3. Where the HANOVA output is high, so too is the likelihood of an object being present. Thus in Fig. 6(a), the object is clearly detectable. More examples involving different types of objects will be given in Section III.

B. Down-track Processing

While HANOVA detects statistical significance at one stop of the array, it does not capture the object signal structure seen as the array moves down-track. To improve detection performance, we employ a sequential detection scheme to process $Y(n)$ recursively as n

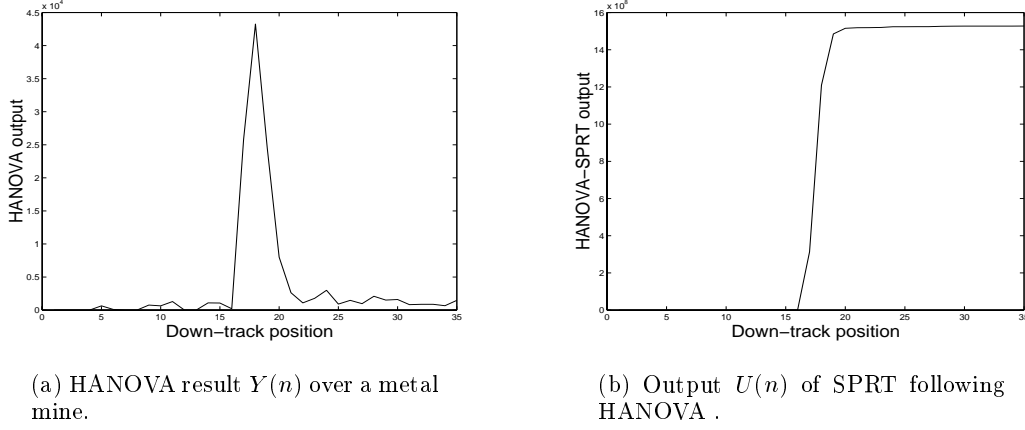


Fig. 6. HANOVA and SPRT processing results for metal mine data shown in Fig. 3

increases in order to identify the transient signal arising from the mine [14,21]. Specification of this sequential probability ratio test (SPRT) begins by noting that under our models $Y(n)$ takes on a χ^2 distribution under both H_0 and H_1 . Standard statistical analysis [12] yields

$$\begin{aligned}
 H_0 : \quad Y(n) &\sim \chi^2_{\sum_{m=1}^M \Delta k(m,n)}(0) \\
 H_1 : \quad Y(n) &\sim \chi^2_{\sum_{m=1}^M \Delta k(m,n)}(\delta^2(n))
 \end{aligned} \tag{11}$$

for $n = 1, \dots, N$ where the notation $x \sim \chi_p^2(\delta^2)$ indicates that the random variable x is distributed according to a χ^2 law of order p and non-centrality parameter δ^2 [20] and $\Delta k(m, n) = k_2(m, n) - k_1(m, n)$ is the length of the (n, m) th window. For the GPR problem it is easy to show that

$$\delta^2(n) = \frac{1}{\sigma_v^2} \sum_{m=1}^M \|\mathbf{s}(m, n) \times \mathbf{w}(m, n)\|^2. \tag{12}$$

For our problem, the length of each window, $\Delta k(m, n)$, is large (on the order of hundreds) and the central limit theorem permits us to approximate the χ^2 distribution using a Gaussian

distribution [20]. We then have

$$\begin{aligned} H_0 : Y(n) &\sim N(\mu_0, \sigma_0^2) \equiv N\left(\sum_{m=1}^M \Delta k(m, n), 2 \sum_{m=1}^M \Delta k(m, n)\right) \\ H_1 : Y(n) &\sim N(\mu_1(n), \sigma_1^2(n)) \equiv N\left(\sum_{m=1}^M \Delta k(m, n) + \delta^2(n), 2 \sum_{m=1}^M \Delta k(m, n) + 4\delta^2(n)\right). \end{aligned} \quad (13)$$

At stop n , the log likelihood ratio for the hypothesis testing problem in (13) is

$$u(n) = \ln \frac{p_n(Y(n))}{p_0(Y(n))}, \quad n = 1, \dots, N \quad (14)$$

where $p_n(Y(n))$ is the PDF of $Y(n)$ evaluated at the n th stop under H_1 and $p_0(Y(n))$ is the PDF of $Y(n)$ evaluated under H_0 . Under H_0 , μ_0 and σ_0^2 are estimated using data from an object-free area. Therefore, for this algorithm, the GPR array must start by collecting data in a calibration region to initialize these variables. Under H_1 , one difficulty with generating $u(n)$ is that $\mu_1(n)$ and $\sigma_1^2(n)$ are typically not known *a priori* since the underlying $\mathbf{s}(m, n)$ are not assumed known. It turns out that we only need to estimate $\mu_1(n)$, and $\sigma_1^2(n)$ can be found from the following relation

$$\begin{aligned} \sigma_1^2(n) &= 2 \sum_{m=1}^M \Delta k(m, n) + 4\delta^2(n) \\ &= 2 \sum_{m=1}^M \Delta k(m, n) + 4 \left[\mu_1(n) - \sum_{m=1}^M \Delta k(m, n) \right] \\ &= 4\mu_1(n) - 2 \sum_{m=1}^M \Delta k(m, n). \end{aligned} \quad (15)$$

At the n th stop, we estimate the mean of $Y(n)$ by its maximum likelihood estimator $\mu_1(n) = Y(n)$.

The *sequential probability ratio test* statistic $U(n)$ is a cumulative sum, changing with the acquisition of each new $u(n)$

$$U(n) = \max(0, U(n-1) + u(n)). \quad (16)$$

- $U(1) = 0$
- FOR $n = 2, \dots, N$
 - $\mu_1(n) = Y(n)$
 - Form $\sigma_1^2(n)$ according to (15)
 - Form $u(n)$ according to (14)
 - $U(n) = \max(0, U(n-1) + u(n))$
 - IF $U(n) > \alpha$, declare object, set $U(n) = 0$, ENDIF
- ENDFOR

Fig. 7. Sequential processing.

Because subsurface object detection is a binary hypothesis testing problem, e.g., we are only interested in knowing whether there is a buried object, the SPRT statistic is bounded from lower bound, zero. When $U(n-1) + u(n)$ is negative, $U(n)$ is reset to zero. For a preset threshold α , the SPRT will make one of two decisions at each n

$$U(n) \geq \alpha \Rightarrow \text{choose } H_1$$

$$U(n) < \alpha \Rightarrow \text{take another observation.}$$

The sequential detection is then essentially a repeated SPRT [22] and summarized in Fig. 7. Fig. 6(b) shows the sequential test statistic when the SPRT is applied to the data in Fig. 6(a). Because the SPRT in (16) has the form of a modified “integrator,” a typical time series for the SPRT statistic takes a step-like form. The larger and sharper the step, the more likely it is that a target is present. At the position where there is an object, the sequential test statistics has a clear upward change again indicating the existence of an object at about position 16.

III. EXAMPLES

In this section we use field data as examples to illustrate the performance of our method. The field data are collected by both single GPR and GPR arrays at different test sites. For each data set, we compare the results from using standard ANOVA, HANOVA, ANOVA

followed by SPRT, and HANOVA followed by SPRT. Comparison indicates that generally HANOVA performs better than ANOVA, and with SPRT, both ANOVA and HANOVA make fewer false-alarms. In other words, HANOVA with SPRT gives the best receiver-operating characteristics, as we shall see later in this section.

At first, we apply our method on data collected by single GPR at different test sites. Some data are taken under relatively favorable condition, while most are from more hostile test sites which involve rough ground surface and other clutter. Fig. 8 compares results of ANOVA and HANOVA on a buried steel object at position 50. For comparison, the outputs of ANOVA and HANOVA are normalized to one. It is observable that while both methods detect the object easily, the HANOVA is better in suppressing noise output where there is no object, e.g., at position 1 through 40 and 60 through 100, Fig. 8(c) and (d). Fig. 9 shows the results from detecting a plastic mine, M19, at position 50. Again, the HANOVA performs better in suppressing noise. At positions 20 through 40, the HANOVA creates a much lower noise level than the ANOVA does. Similarly, the HANOVA produces a cleaner output at the end of the run.

Fig. 10 shows the results of ANOVA and HANOVA in detecting an anti-tank mine, TM62, from a very “noisy” data set. The mine is buried at position 60. Outputs of both HANOVA and ANOVA consist of the correct detection and some false alarms. The HANOVA maintains a better performance than the ANOVA in the sense that, for a given detection threshold, the HANOVA would generally have a smaller number of false-alarms. For the HANOVA, no false alarms will be declared for a threshold greater than 0.5, while for the ANOVA, the threshold must be set above 0.8 to avoid making a wrong decision. Between threshold 0.5 and 0.8, the ANOVA will make two false-alarms while the HANOVA has zero false-alarm.

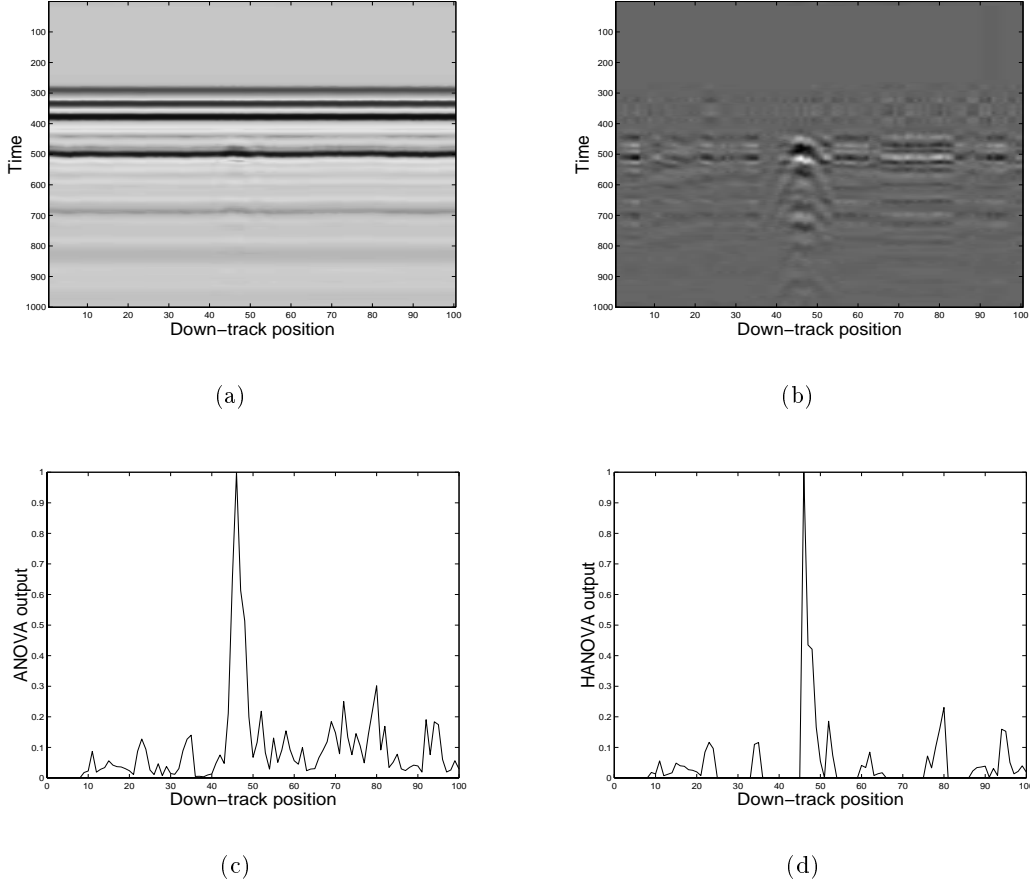


Fig. 8. Results of a single GPR measurement above a steel object around position 50, a) raw observation, b) observation after nominal background removal, c) ANOVA output, d) HANOVA output. In a) and b), each unit in time axes is 0.02 ns.

Next, by comparing the outputs of the SPRT in the above three examples, we see that sequential processing generally smoothes the output and generates fewer false-alarms than by using ANOVA (or HANOVA) only, Fig. 11. In all three examples, SPRT following HANOVA performs better than SPRT following ANOVA, in the sense that the output is more leveled off at object-free area and the jump at the position of the buried object is sharper.

To study the receiver operating characteristic (ROC) of the method, we test our method on multiple runs of different type of targets. Fig. 12(a) shows the ROC curves of ANOVA and HANOVA to detect metallic objects. The objects include metallic mines such as TM15,

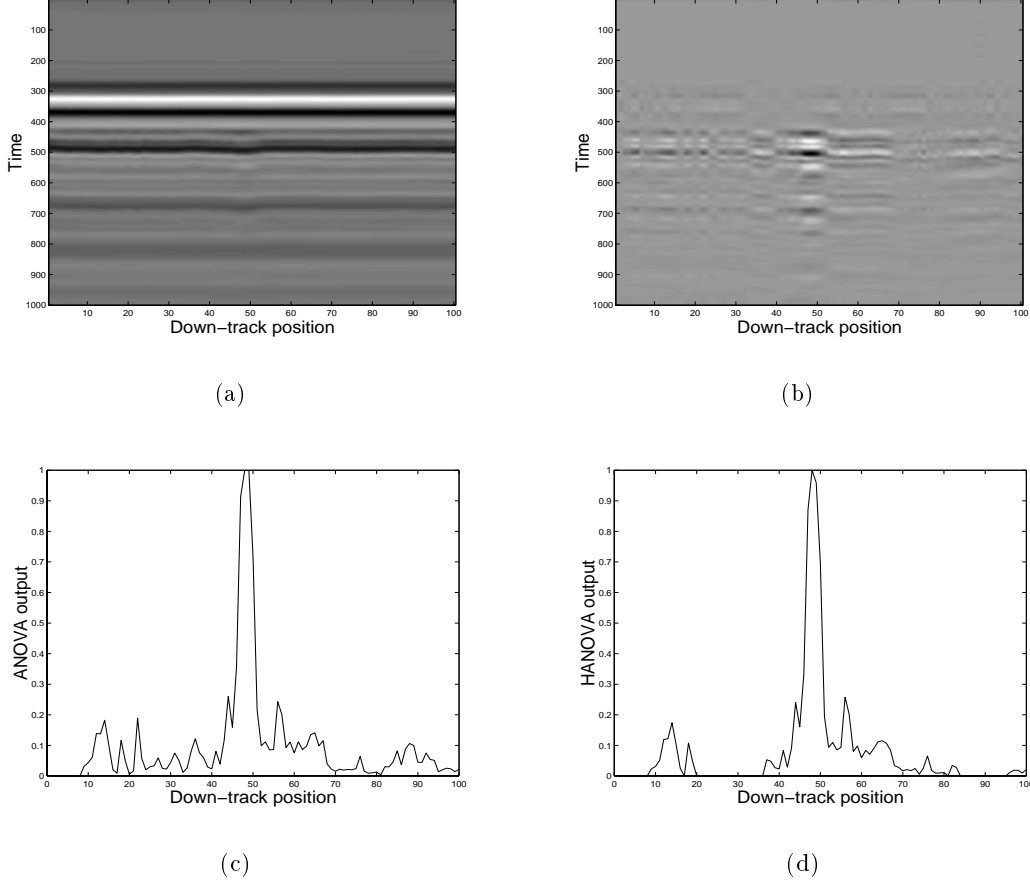


Fig. 9. Comparison between the ANOVA and HANOVA, a) raw data over an M19, anti-tank mine, buried at position 50, b) demeaned data, c) result of the ANOVA over the M19, d) result of the HANOVA.

TM46, and PMN. Fig. 12(b) shows the ROC of ANOVA-SPRT and HANOVA-SPRT. Compared with Fig. 12(a), SPRT improves the performance of both ANOVA and HANOVA. In generating these curves a correct identification of any of the objects was taken to be a “detection” whether or not the object itself was a mine. Indeed, as noted in the Introduction, the algorithm in this paper is intended *only* to detect the presence of objects below the array and not to solve the classification problem. Still, given the “real-world” conditions under which the data were taken, the low false alarm rates here point to the robustness of our approach.

Next, we compare the performance of ANOVA, HANOVA, ANOVA-SPRT, and HANOVA-
April 29, 2001—3 : 44 pm DRAFT

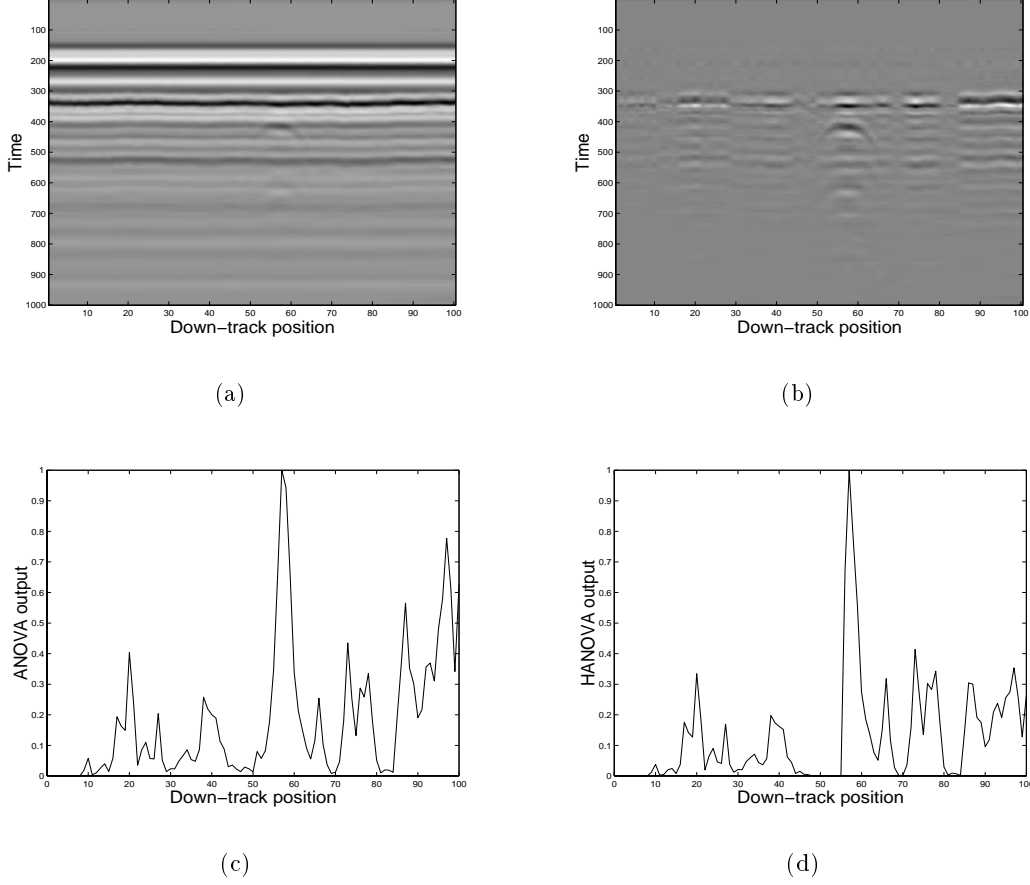


Fig. 10. More comparison between the ANOVA and HANOVA, a) raw data over a TM62, anti-tank mine, buried to the side of the track at position 58, b) demeaned data, c) result of the ANOVA, d) result of the HANOVA.

SPRT in detecting plastic mines. The mines are M19, VS-1.6, T72, and C4A1. Fig. 13 shows the ROC curves of the above four methods. It is seen that both the ANOVA-SPRT and HANOVA-SPRT perform better than the ANOVA and HANOVA, respectively.

As another example, we test our method on a different array radar system at another test site. The setup of the GPR array is shown in Fig. 14. There is one transmitter in this system. In front of the transmitter, four receivers are positioned in a 2×2 pattern. Above the transmitter and the receivers there is a hyperbolic reflection plate, it is set so that the transmitter is at the focal point of the reflection plate. The array moves on a linear track

to collect data. At each step, the transmitter sends a spherical wave to the reflection plate and after reflection, the sphere wave becomes plane wave. The four receivers then collect reflection of this plane wave from the ground. The system has the advantage of generating plane wave and points it forward to reduce ground reflection. Fig. 15 displays collected data from the two front receivers at the Dedham test site of Northeastern University and the corresponding signal after background removal. In an area of 58 m², there are 12 buried landmines of different types, such as M19, PMN, VS-2.2, and so on. Using our method we are able to detect all 12 mines with a few false-alarms, Fig. 16. The results are similar to those obtained by a single GPR. For a detection rate above 90%, the HANOVA has a significantly smaller number of false-alarm.

IV. SUMMARY

In this paper, we have proposed a sequential, high-dimensional ANOVA to process GPR returns. The method is tested on real data and has a relaxed requirement on the physical model used in the processing routine. The method is on-line implementable and has a linear computational load. The method works in two stages: first it looks for statistically significant difference from array observations, second, it applies a sequential detection as new data are obtained. HANOVA is powerful in the sense of maximizing probability of detecting statistically significant difference among sub-dimensions of a full vector of observations. Sequential detection recursively processes the result of the HANOVA and enables real-time processing as new data are collected. We have demonstrated the performance of this technique on samples of field data.

Future research will focus on classification and localization. Classification consists of two steps. First, a feature extraction scheme is applied on reflected signals to generate intermediate result, second, the output from feature extraction is fed into a Bayesian classifier

to make the classification. Localization is based on optimized frequency-wavenumber (F-K) migration. F-K migration is an inversion method that back-propagates wavefield from the ground surface to subsurface and construct an image of subsurface reflectivity. Regular F-K migration is well modeled for seismic signal processing. Though GPR signal is different from seismic signal, F-K migration can still work very well in processing GPR signals. Nonetheless, improvement in terms of resolution and accuracy can be achieved by considering optimization in F-K migration. In our future work, an optimization method will be used to improve the F-K migration.

APPENDIX

I. WINDOW SELECTION IN HANOVA

Ideally, we want to find a window that is sensitive to the presence of a signal and provides little response in the test statistic when there is noise only. But these two requirements are often in conflict with each other. From (8), we can change the order l of the MA process to control the window we use. The smaller is l , the more sensitive the window is to the presence of signal and strong noise. On the other hand, the larger is l , the more robust will the statistic be to noise, which translates into a smaller probability of false-alarm. But a large l reduces sensitivity of the HANOVA to signal. Fig. 17 shows the effect of l on window selection and the corresponding HANOVA results. Three different l are used, i.e., $l = 1, 4, 9$. In the data, there are three mine objects, two metal mines at the position 110 and 170. A weak mine object is at position 25. For comparison, we normalize the HANOVA outputs in each case by its maximum value, which corresponds to the strong metal mine buried at position 110. Fig. 17(a) and (b) show the window chosen by a MA of order 1 and the resulting HANOVA output. The two strong objects can be detected at a threshold of 0.7, the weak object can only be found at a threshold of 0.2. Fig. 17(c) shows the window

chosen by a MA of order 4. The window oscillates much less than the window in Fig. 17(a). From the HANOVA result, Fig. 17(d), we can find all the three objects at a threshold of 0.3. Increasing the order of MA process can make the results worse, Fig. 17(e) and (f). A large window reduces the sensitivity of the HANOVA to signal and actually makes detection more difficult. Now the weak object at position 25 can not be detected at a threshold greater than 0.3. As a guideline, we find that MA processes of order between 3 and 10 yield good windows both in sensitivity to signal and robustness to noise. This selection is affected by the step-size of the array. An array moving at small step-size will allow an MA process of large l in selecting windows, and vice versa.

REFERENCES

- [1] L. Peters Jr., J. J. Daniels, and J. D. Young, “Ground penetrating radar as a subsurface environmental sensing tool”, *Proceedings of the IEEE*, vol. 82, no. 12, pp. 1802–1822, Dec. 1994.
- [2] C.M. Rappaport, S.G. Azevedo, T. Rosenbury, and J. Gough, “Handheld forward-looking focused array mine detection with plane wave excitation”, in *Detection Remediation of Mines and Minelike Targets V*, 2000, vol. SPIE 4038.
- [3] J. Kositsky and P. Milanfar, “A forward-looking high-resolution GPR system”, in *Detection Remediation of Mines and Minelike Targets IV*, 1999, vol. SPIE 3710, pp. 1052–1062.
- [4] J. M. Bourgeois and G. S. Smith, “A fully three-dimensional simulation of a ground-penetrating radar: FDTD theory compared with experiment”, *IEEE Trans. Geosci. Remote Sensing*, vol. 34, no. 1, pp. 36–44, Jan. 1996.
- [5] J. R. Wait, *Geo-electromagnetism*, Academic Press, Inc., New York, 1982.
- [6] G. R. Olhoeft, “Application of ground penetrating radar”, in *GPR’96*, Sendai, Japan, April 29, 2001—3 : 44 pm

- Sept. 1996, pp. 1–3.
- [7] K. Demarest, Z. Huang, and R. Plumb, “An FDTD near- to far-zone transformation for scatterers buried in stratified ground”, *IEEE Trans. Antennas and Propagation*, vol. 44, no. 8, pp. 1150–1157, Aug. 1996.
- [8] P. G. Gader, H. Hichem, B. N. Nelson, G. Vaillette, and J. M. Keller, “New results in fuzzy set based detection of landmines with GPR”, in *Detection and Remediation of Mines and Minelike Targets IV*, 1999, vol. SPIE 3710, pp. 1075–1084.
- [9] M. R. Bradley, T. R. Witten, R. McCummins, M. Crowe, S. Stewart, and M. Duncan, “Mine detection with a multichannel stepped-frequency ground-penetrating radar”, in *Detection and Remediation of Mines and Minelike Targets IV*, 1999, vol. SPIE 3710, pp. 953–960.
- [10] T. Dogaru and L. Carin, “Time-domain sensing of targets buried under a rough air-ground interface”, *IEEE Trans. Antennas and Propagation*, vol. 46, no. 3, pp. 360–372, Mar. 1998.
- [11] “UXO Signature Data”, <http://www.denix.osd.mil/denix/Public/News/UXOCOE/Sigdata/sigdata.html>, 1999.
- [12] J. Fan, “Test of significance based on wavelet thresholding and Neyman’s truncation”, *Journal of American Statistical Association*, vol. 91, pp. 674–688, 1996.
- [13] J. Fan and S. Lin, “Test of significance when data are curves”, *Journal of American Statistical Association*, vol. 93, pp. 1007–1021, 1998.
- [14] C. W. Helstrom, *Elements of Signal Detection and Estimation*, Prentice-Hall, New York, 1995.
- [15] P. Shi and E. L. Miller, “Baseband Wiener filter processing for mine detection from scanned laser induced acoustic data”, in *Detection and Remediation of Mines and Mine-*

- like Targets IV*, 1999, vol. SPIE 3710, pp. 1373–1384.
- [16] X. Xu, E. L. Miller, and C. M. Rappaport, “Combined high-dimensional analysis of variance (HANOVA) and sequential probability ratio test (SPRT) to detect buried mines”, in *Detection and Remediation of Mines and Minelike Targets V*, Apr. 2000, vol. SPIE 4038.
- [17] X. Xu, E. L. Miller, and C. M. Rappaport, “Statistically-based sequential detection of buried mines from array ground penetrating radar data”, in *Detection and Remediation of Mines and Minelike Targets IV*, 1999, vol. SPIE 3710, pp. 1063–1075.
- [18] M. El-Shenawee and C. M. Rappaport, “Quantifying the effects of different rough surface statistics for mine detection using the FDTD technique”, in *Detection Remediation of Mines and Minelike Targets V*, 2000, vol. SPIE 4038, pp. 966–975.
- [19] B. K. Ghosh, *Sequential Tests of Statistical Hypotheses*, Addison-Wesley, Reading, MA, 1970.
- [20] S. C. Port, *Theoretical Probability for Applications*, John Wiley & Sons, New York, 1994.
- [21] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlag, New York, 1985.
- [22] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall, New York, 1993.
- [23] C. Han, P. K. Willett, and D. A. Abraham, “Some methods to evaluate the performance of Page’s test as used to detect transient signals”, *IEEE Trans. Signal Processing*, vol. 47, no. 8, pp. 2112–2127, Aug. 1999.
- [24] E. Bahar, “Full-wave solutions for the depolarization of the scattered radiation fields by rough surfaces of arbitrary slope”, *IEEE Trans. Antennas and Propagation*, vol. AP-29,

pp. 443–454, 1981.

- [25] D. A. Kapp and G. S. Brown, “A new numerical method for rough-surface scattering calculations”, *IEEE Trans. Antennas and Propagation*, vol. 44, no. 5, pp. 711–721, May 1996.

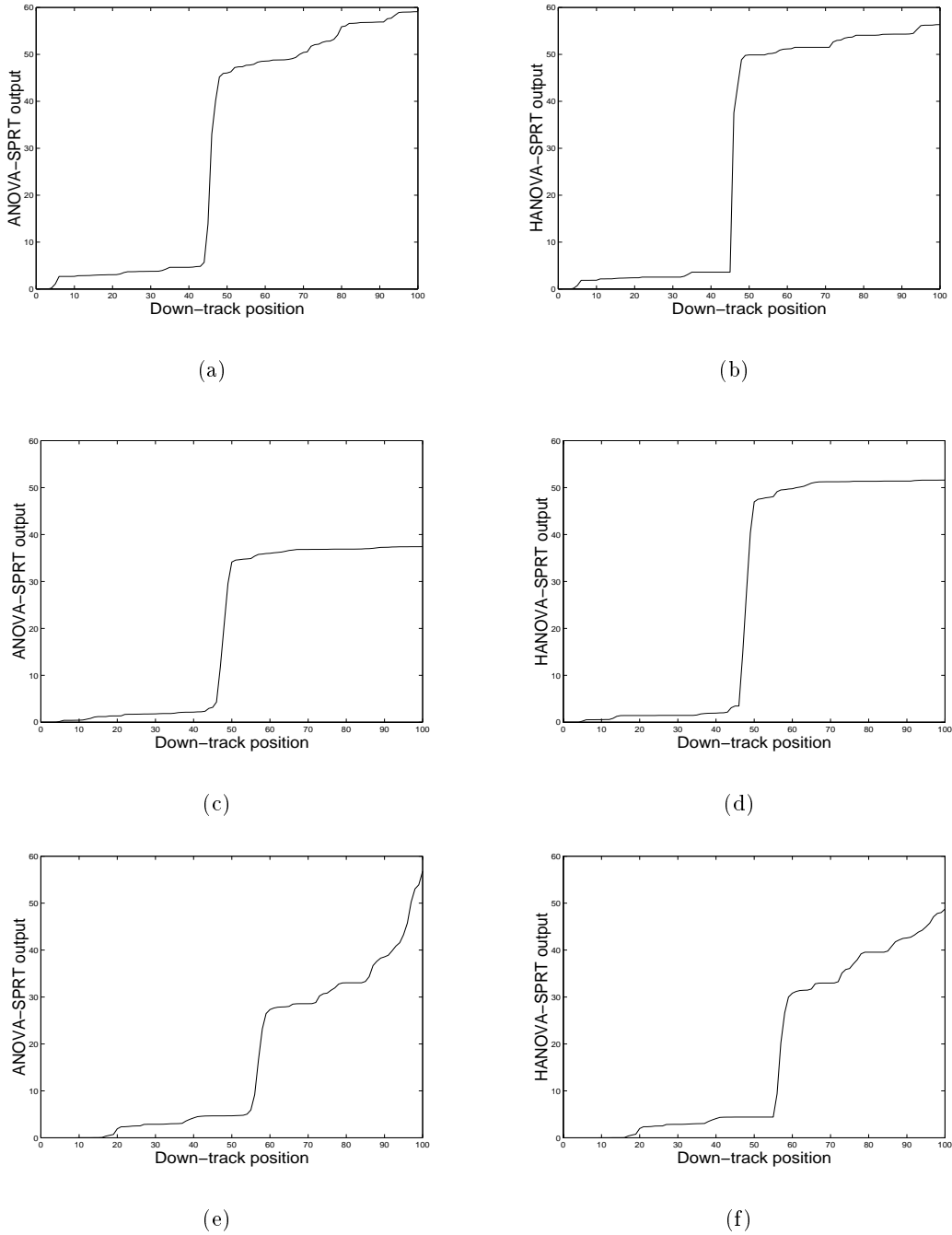


Fig. 11. Results of the SPRT, a buried steel object, a) output of ANOVA-SPRT, b) output of HANOVA-SPRT; a buried M19, c) output of ANOVA-SPRT, d) output of HANOVA-SPRT; a buried TM62, e) output of ANOVA-SPRT, f) output of HANOVA-SPRT.

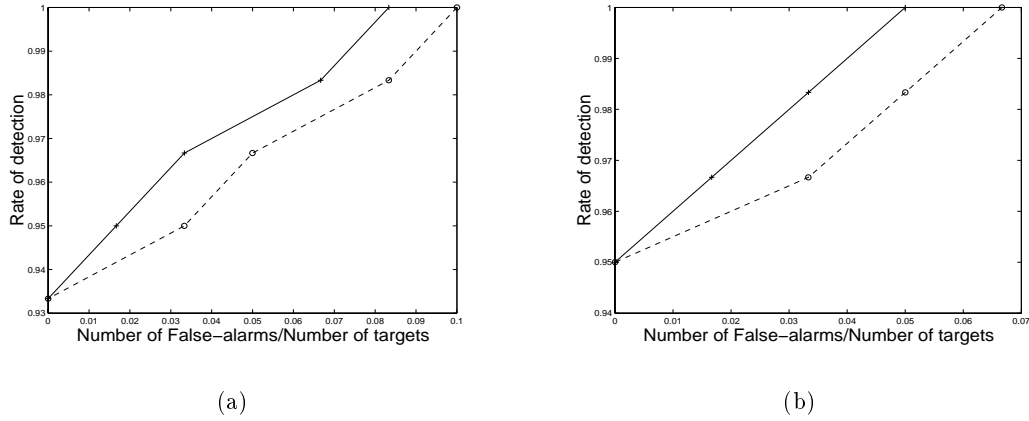


Fig. 12. Rate of detection and rate of false-alarms in detecting metallic objects, solid line is the result of HANOVA, dashed line is the result of ANOVA, a) ANOVA vs. HANOVA, b) ANOVA-SPRT vs. HANOVA-SPRT.

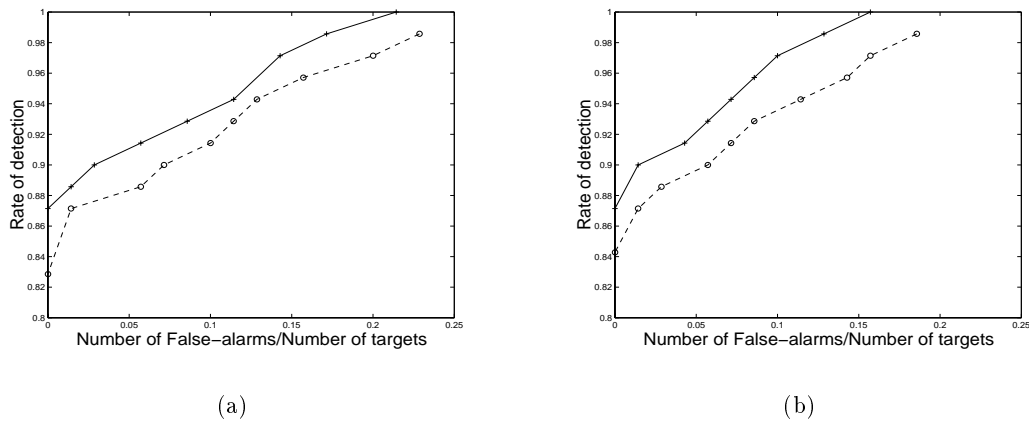


Fig. 13. Rate of detection and rate of false-alarms in detecting plastic mines, solid line is the result of HANOVA, dashed line is the result of ANOVA, a) ANOVA vs. HANOVA, b) ANOVA-SPRT vs. HANOVA-SPRT.

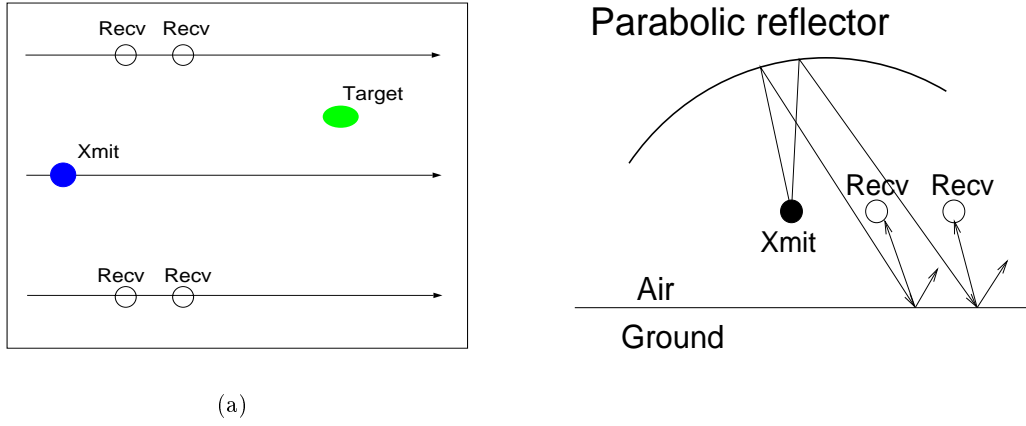


Fig. 14. GPR array used at Dedham test site of Northeastern University, a) plane view, b) side view.

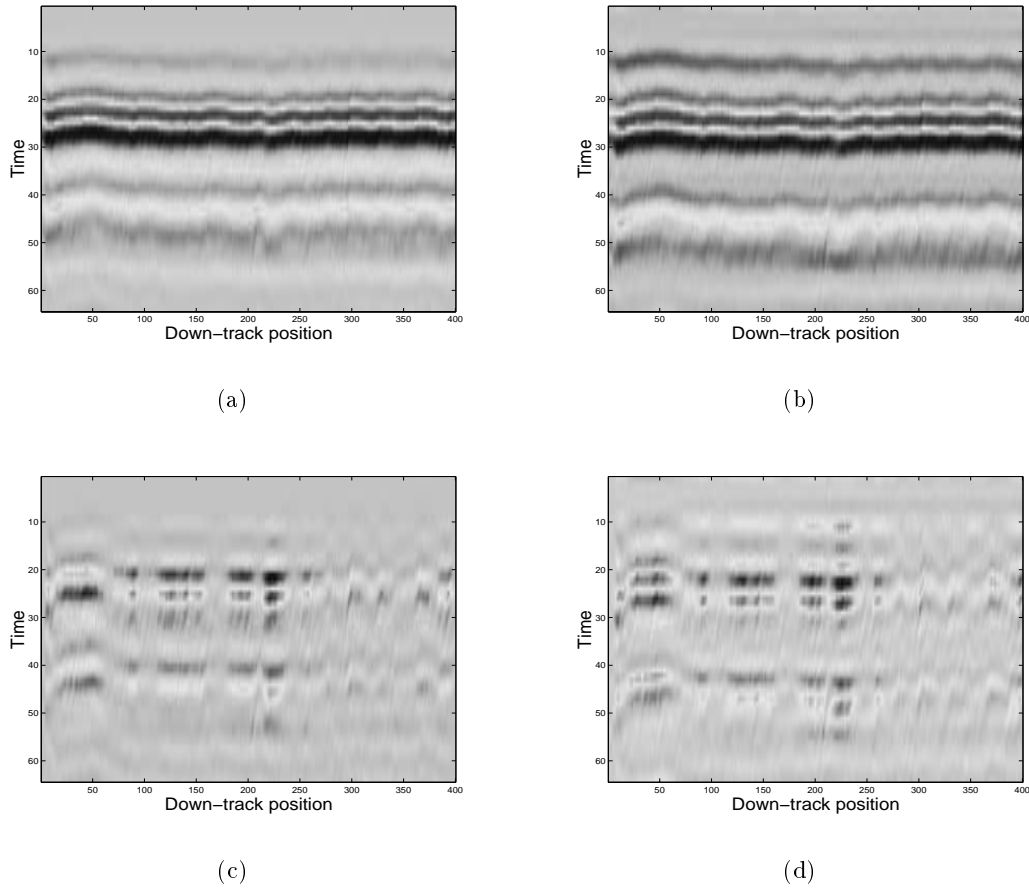


Fig. 15. GPR data from the Dedham test site, (a) from the left front receiver, (b) from the right front receiver, (c) signal (a) after background removal, (d) signal of (b) after background removal. Unit in time axes is 120 ps.

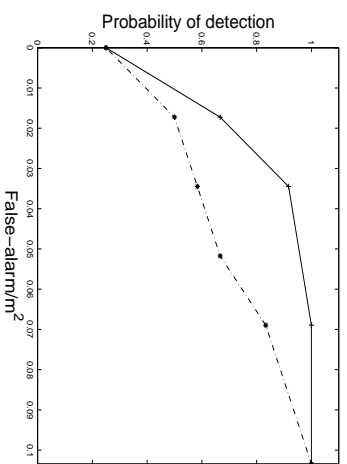


Fig. 16. ROCs of the Dedham test, solid line, HANOVA-SPRT, dash line, ANOVA-SPRT.

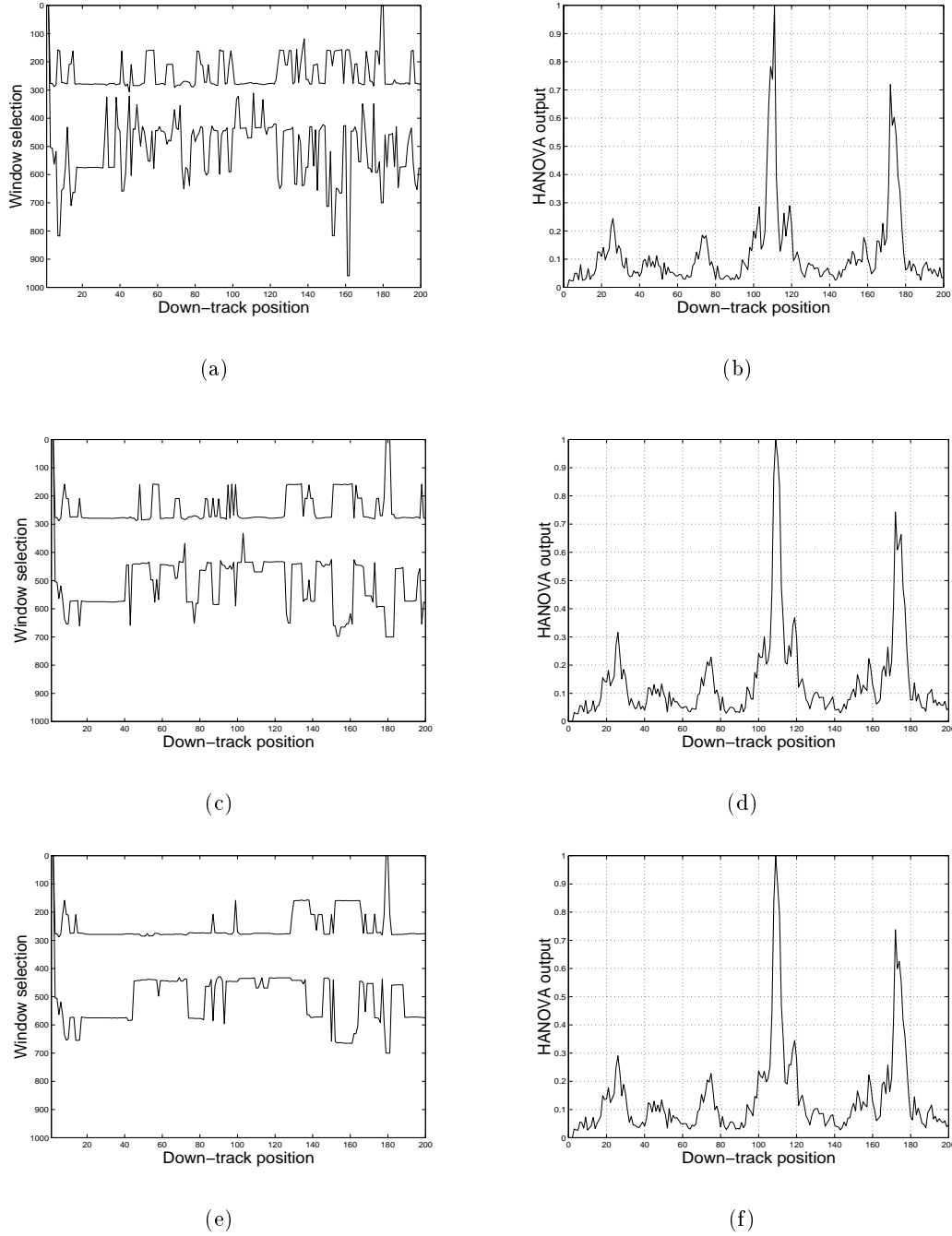


Fig. 17. Choices of window and effect on HANOVA, lower line is $k_1(m, n)$, upper line is $k_2(m, n)$, a) window selected by an order 1 MA process, b) HANOVA result from the window to the left, c) window selected by an order 4 MA process, d) HANOVA result from the window to the left, e) window selected by an order 10 MA process, f) HANOVA test result from the window to the left.